Calculus

Paul Dawkins

Table of Contents

Preface

First, here's a little bit of history on how this material was created (there's a reason for this, I promise). A long time ago (2002 or so) when I decided I wanted to put some mathematics stuff on the web I wanted a format for the source documents that could produce both a pdf version as well as a web version of the material. After some investigation I decided to use MS Word and MathType as the easiest/quickest method for doing that. The result was a pretty ugly HTML (*i.e* web page code) and had the drawback of the mathematics were images which made editing the mathematics painful. However, it was the quickest way or dealing with this stuff.

Fast forward a few years (don't recall how many at this point) and the web had matured enough that it was now much easier to write mathematics in LaTeX (https://en.wikipedia.org/wiki/LaTeX) and have it display on the web (LaTeX was my first choice for writing the source documents). So, I found a tool that could convert the MS Word mathematics in the source documents to LaTeX. It wasn't perfect and I had to write some custom rules to help with the conversion but it was able to do it without "messing" with the mathematics and so I didn't need to worry about any math errors being introduced in the conversion process. The only problem with the tool is that all it could do was convert the mathematics and not the rest of the source document into LaTeX. That meant I just converted the math into LaTeX for the website but didn't convert the source documents.

Now, here's the reason for this history lesson. Fast forward even more years and I decided that I really needed to convert the source documents into LaTeX as that would just make my life easier and I'd be able to enable working links in the pdf as well as a simple way of producing an index for the material. The only issue is that the original tool I'd use to convert the MS Word mathematics had become, shall we say, unreliable and so that was no longer an option and it still has the problem on not converting anything else into proper LaTeX code.

So, the best option that I had available to me is to take the web pages, which already had the mathematics in proper LaTeX format, and convert the rest of the HTML into LaTeX code. I wrote a set of tools to do his and, for the most part, did a pretty decent job. The only problem is that the tools weren't perfect. So, if you run into some "odd" stuff here (things like <sup>, , , <div>, *etc.*) please let me know the section with the code that I missed. I did my best to find all the "orphaned" HTML code but I'm certain I missed some on occasion as I did find my eyes glazing over every once in a while as I went over the converted document.

Now, with that out of the way, let's get into the actual preface of the material.

Here are the notes for my Calculus courses that I teach here at Lamar University. Despite the fact that these are my "class notes", they should be accessible to anyone wanting to learn Calculus or needing a refresher in Calculus. This document contains the majority, if not all, of the topics that are typically taught in full set of Calculus courses (*i.e.* Calculus I, Calculus II and Calculus III). Note that there are also topics that for a variety of reasons (mostly time issues) I am not able to cover in my classes. These topics are included for those that wish to learn those topics and/or for instructors that are using this material for their course and wish to cover the topics. Also, even in sections that I do cover in my classes I may not actually cover all the examples listed here (again for time reasons) and they are provided for those that wish to see another example or two.

I've tried to make these material as self-contained as possible and so all the information needed to get started reading through them is either from an Algebra or Trig class. Note that, outside of the Review chapter, I am assuming that you do recall the Algebra and/or Trig needed at various points and won't, for the most part, be reviewing or covering that as those topics arise. For the most part I will simply assume that you recall those topics or can go back and refresh them as needed. This is not meant to be "difficult" with the reader but simply an acknowledgment that I have to assume you have the prerequisite knowledge at some point so we can focus on the topics we're trying to learn rather than spending all our time refreshing knowledge that you really are supposed to know (or at least be somewhat familiar with) prior to getting into a Calculus course. There is also the reality that if I included discussion/refresher of all the prerequisite material at every step these pages would eventually be so long that it would be hard to focus in on the material that we're actually trying to learn.

Also note that not spending time refreshing prerequisite material will extend into later topics in the material as well. For example, when discussing the integration techniques typically covered in a Calculus II course it is assumed that you know basic integration and don't need more that a cursory, at best, refresher/reminder of basic integration and basic substitutions. Another example of this would be moving into multi-variable Calculus, *i.e.* the material typically taught in a Calculus III course. Once we move into multi-variable Calculus it is assumed that you understand single variable Calculus and can do basic differentiation and integration.

So, if you are jumping into the middle of the material to learn a particular topic and run across something that you don't know there is a good chance that you are missing some knowledge of a prerequisite material and will need to find it in this set of material to cover that prior to learning the topic you wish to learn.

Here are a couple of warnings to my students who may be here to get a copy of what happened on a day that you missed.

1. Because I wanted to make this a fairly complete set of notes for anyone wanting to learn calculus I have included some material that I do not usually have time to cover in class and because this changes from semester to semester it is not noted here. You will need to find one of your fellow class mates to see if there is something in these notes that wasn't covered in class.

2. Because I want these notes to provide some more examples for you to read through, I don't always work the same problems in class as those given in the notes. Likewise, even if I do work some of the problems in here I may work fewer problems in class than are presented here.

3. Sometimes questions in class will lead down paths that are not covered here. I tried to anticipate as many of the questions as possible when writing these up, but the reality is that I can't anticipate all the questions. Sometimes a very good question gets asked in class that leads to insights that I've not included here. You should always talk to someone who was in class on the day you missed and compare these notes to their notes and see what the differences are.

4. This is somewhat related to the previous three items, but is important enough to merit its own item. THESE NOTES ARE NOT A SUBSTITUTE FOR ATTENDING CLASS!! Using these notes as a substitute for class is liable to get you in trouble. As already noted not everything in these notes is covered in class and often material or insights not in these notes is covered in class.

Outline

Here is a listing (and brief description) of the material in this set of notes.

Review - In this chapter we give a brief review of selected topics from Algebra and Trig that are vital to surviving a Calculus course. Included are Functions, Trig Functions, Solving Trig Equations, Exponential/Logarithm Functions and Solving Exponential/Logarithm Equations.

Functions - In this section we will cover function notation/evaluation, determining the domain and range of a function and function composition.

Inverse Functions - In this section we will define an inverse function and the notation used for inverse functions. We will also discuss the process for finding an inverse function.

Trig Functions - In this section we will give a quick review of trig functions. We will cover the basic notation, relationship between the trig functions, the right triangle definition of the trig functions. We will also cover evaluation of trig functions as well as the unit circle (one of the most important ideas from a trig class!) and how it can be used to evaluate trig functions.

Solving Trig Equations - In this section we will discuss how to solve trig equations. The answers to the equations in this section will all be one of the "standard" angles that most students have memorized after a trig class. However, the process used here can be used for any answer regardless of it being one of the standard angles or not.

Solving Trig Equations with Calculators, Part I - In this section we will discuss solving trig equations when the answer will (generally) require the use of a calculator (*i.e.* they aren't one of the standard angles). Note however, the process used here is identical to that for when the answer is one of the standard angles. The only difference is that the answers in here can be a little messy due to the need of a calculator. Included is a brief discussion of inverse trig functions.

Solving Trig Equations with Calculators, Part II - In this section we will continue our discussion of solving trig equations when a calculator is needed to get the answer. The equations in this section tend to be a little trickier than the "normal" trig equation and are not always covered in a trig class.

Exponential Functions - In this section we will discuss exponential functions. We will cover the basic definition of an exponential function, the natural exponential function, *i.e.* $\mathbf{e}^x$, as well as the properties and graphs of exponential functions

Logarithm Functions - In this section we will discuss logarithm functions, evaluation of logarithms and their properties. We will discuss many of the basic manipulations of logarithms that commonly occur in Calculus (and higher) classes. Included is a discussion of the natural (ln(x)) and common logarithm (log(x)) as well as the change of base formula.

Exponential and Logarithm Equations - In this section we will discuss various methods for solving equations that involve exponential functions or logarithm functions.

Common Graphs - In this section we will do a very quick review of many of the most common functions and their graphs that typically show up in a Calculus class.

Limits - In this chapter we introduce the concept of limits. We will discuss the interpretation/meaning of a limit, how to evaluate limits, the definition and evaluation of one-sided limits, evaluation of infinite limits, evaluation of limits at infinity, continuity and the Intermediate Value Theorem. We will also give a brief introduction to a precise definition of the limit and how to use it to evaluate limits.

Tangent Lines and Rates of Change - In this section we will introduce two problems that we will see time and again in this course : Rate of Change of a function and Tangent Lines to functions. Both of these problems will be used to introduce the concept of limits, although we won't formally give the definition or notation until the next section.

The Limit - In this section we will introduce the notation of the limit. We will also take a conceptual look at limits and try to get a grasp on just what they are and what they can tell us. We will be estimating the value of limits in this section to help us understand what they tell us. We will actually start computing limits in a couple of sections.

One-Sided Limits - In this section we will introduce the concept of one-sided limits. We will discuss the differences between one-sided limits and limits as well as how they are related to each other.

Limit Properties - In this section we will discuss the properties of limits that we'll need to use in computing limits (as opposed to estimating them as we've done to this point). We will also compute a couple of basic limits in this section.

Computing Limits - In this section we will looks at several types of limits that require some work before we can use the limit properties to compute them. We will also look at computing limits of piecewise functions and use of the Squeeze Theorem to compute some limits.

Infinite Limits - In this section we will look at limits that have a value of infinity or negative infinity. We'll also take a brief look at vertical asymptotes.

Limits At Infinity, Part I - In this section we will start looking at limits at infinity, *i.e.* limits in which the variable gets very large in either the positive or negative sense. We will concentrate on polynomials and rational expressions in this section. We'll also take a brief look at horizontal asymptotes.

Limits At Infinity, Part II - In this section we will continue covering limits at infinity. We'll be looking at exponentials, logarithms and inverse tangents in this section.

Continuity - In this section we will introduce the concept of continuity and how it relates to limits. We will also see the Intermediate Value Theorem in this section and how it can be used to determine if functions have solutions in a given interval.

The Definition of the Limit - In this section we will give a precise definition of several of the limits covered in this section. We will work several basic examples illustrating how to use this precise definition to compute a limit. We'll also give a precise definition of continuity.

Derivatives - In this chapter we will start looking at the next major topic in a calculus class, derivatives. This chapter is devoted almost exclusively to finding derivatives. We will be looking at one application of them in this chapter. We will be leaving most of the applications of derivatives to the next chapter.

The Definition of the Derivative - In this section we define the derivative, give various notations for the derivative and work a few problems illustrating how to use the definition of the derivative to actually compute the derivative of a function.

Interpretation of the Derivative - In this section we give several of the more important interpretations of the derivative. We discuss the rate of change of a function, the velocity of a moving object and the slope of the tangent line to a graph of a function.

Differentiation Formulas - In this section we give most of the general derivative formulas and properties used when taking the derivative of a function. Examples in this section concentrate mostly on polynomials, roots and more generally variables raised to powers.

Product and Quotient Rule - In this section we will give two of the more important formulas for differentiating functions. We will discuss the Product Rule and the Quotient Rule allowing us to differentiate functions that, up to this point, we were unable to differentiate.

Derivatives of Trig Functions - In this section we will discuss differentiating trig functions. Derivatives of all six trig functions are given and we show the derivation of the derivative of $\sin(x)$ and $\tan(x)$.

Derivatives of Exponential and Logarithm Functions - In this section we derive the formulas for the derivatives of the exponential and logarithm functions.

Derivatives of Inverse Trig Functions - In this section we give the derivatives of all six inverse trig functions. We show the derivation of the formulas for inverse sine, inverse cosine and inverse tangent.

Derivatives of Hyperbolic Functions - In this section we define the hyperbolic functions, give the relationships between them and some of the basic facts involving hyperbolic functions. We also give the derivatives of each of the six hyperbolic functions and show the derivation of the formula for hyperbolic sine.

Chain Rule - In this section we discuss one of the more useful and important differentiation formulas, The Chain Rule. With the chain rule in hand we will be able to differentiate a much wider variety of functions. As you will see throughout the rest of your Calculus courses a great many of derivatives you take will involve the chain rule!

Implicit Differentiation - In this section we will discuss implicit differentiation. Not every function can be explicitly written in terms of the independent variable, e.g. $y = f(x)$ and yet we will still need to know what $f'(x)$ is. Implicit differentiation will allow us to find the derivative in these cases. Knowing implicit differentiation will allow us to do one of the more important applications of derivatives, Related Rates (the next section).

Related Rates - In this section we will discuss the only application of derivatives in this section, Related Rates. In related rates problems we are give the rate of change of one quantity in a problem and asked to determine the rate of one (or more) quantities in the problem. This is often one of the more difficult sections for students. We work quite a few problems in this section so hopefully by the end of this section you will get a decent understanding on how these problems work.

Higher Order Derivatives - In this section we define the concept of higher order derivatives and give a quick application of the second order derivative and show how implicit differentiation works for higher order derivatives.

Logarithmic Differentiation - In this section we will discuss logarithmic differentiation. Logarithmic differentiation gives an alternative method for differentiating products and quotients (sometimes easier than using product and quotient rule). More importantly, however, is the fact that logarithm differentiation allows us to differentiate functions that are in the form of one function raised to another function, *i.e.* there are variables in both the base and exponent of the function.

Derivative Applications - In the previous chapter we focused almost exclusively on the computation of derivatives. In this chapter will focus on applications of derivatives. It is important to always remember that we didn't spend a whole chapter talking about computing derivatives just to be talking about them. There are many very important applications to derivatives.

The two main applications that we'll be looking at in this chapter are using derivatives to determine information about graphs of functions and optimization problems. These will not be the only applications however. We will be revisiting limits and taking a look at an application of derivatives that will allow us to compute limits that we haven't been able to compute previously. We will also see how derivatives can be used to estimate solutions to equations.

Rates of Change - In this section we review the main application/interpretation of derivatives from the previous chapter (*i.e.* rates of change) that we will be using in many of the applications in this chapter.

Critical Points - In this section we give the definition of critical points. Critical points will show up in most of the sections in this chapter, so it will be important to understand them and how to find them. We will work a number of examples illustrating how to find them for a wide variety of functions.

Minimum and Maximum Values - In this section we define absolute (or global) minimum and maximum values of a function and relative (or local) minimum and maximum values of a function. It is important to understand the difference between the two types of minimum/-maximum (collectively called extrema) values for many of the applications in this chapter and so we use a variety of examples to help with this. We also give the Extreme Value Theorem and Fermat's Theorem, both of which are very important in the many of the applications we'll see in this chapter.

Finding Absolute Extrema - In this section we discuss how to find the absolute (or global) minimum and maximum values of a function. In other words, we will be finding the largest and smallest values that a function will have.

The Shape of a Graph, Part I - In this section we will discuss what the first derivative of a function can tell us about the graph of a function. The first derivative will allow us to identify the relative (or local) minimum and maximum values of a function and where a function will be increasing and decreasing. We will also give the First Derivative test which will allow us to classify critical points as relative minimums, relative maximums or neither a minimum or a maximum.

The Shape of a Graph, Part II - In this section we will discuss what the second derivative of a function can tell us about the graph of a function. The second derivative will allow us to determine where the graph of a function is concave up and concave down. The second derivative will also allow us to identify any inflection points (i.e. where concavity changes) that a function may have. We will also give the Second Derivative Test that will give an alternative method for identifying some critical points (but not all) as relative minimums or relative maximums.

The Mean Value Theorem - In this section we will give Rolle's Theorem and the Mean Value Theorem. With the Mean Value Theorem we will prove a couple of very nice facts, one of which will be very useful in the next chapter.

Optimization Problems - In this section we will be determining the absolute minimum and/or maximum of a function that depends on two variables given some constraint, or relationship, that the two variables must always satisfy. We will discuss several methods for determining the absolute minimum or maximum of the function. Examples in this section tend to center around geometric objects such as squares, boxes, cylinders, etc.

More Optimization Problems - In this section we will continue working optimization problems. The examples in this section tend to be a little more involved and will often involve situations that will be more easily described with a sketch as opposed to the 'simple' geometric objects we looked at in the previous section.

L'Hospital's Rule and Indeterminate Forms - In this section we will revisit indeterminate forms and limits and take a look at L'Hospital's Rule. L'Hospital's Rule will allow us to evaluate some limits we were not able to previously.

Linear Approximations - In this section we discuss using the derivative to compute a linear approximation to a function. We can use the linear approximation to a function to approximate values of the function at certain points. While it might not seem like a useful thing to do with when we have the function there really are reasons that one might want to do this. We give two ways this can be useful in the examples.

Differentials - In this section we will compute the differential for a function. We will give an application of differentials in this section. However, one of the more important uses of differentials will come in the next chapter and unfortunately we will not be able to discuss it until then.

Newton's Method - In this section we will discuss Newton's Method. Newton's Method is an application of derivatives that will allow us to approximate solutions to an equation. There are many equations that cannot be solved directly and with this method we can get approximations to the solutions to many of those equations.

Business Applications - In this section we will give a cursory discussion of some basic applications of derivatives to the business field. We will revisit finding the maximum and/or minimum function value and we will define the marginal cost function, the average cost, the revenue function, the marginal revenue function and the marginal profit function. Note that this section is only intended to introduce these concepts and not teach you everything about them.

Integrals In this chapter we will be looking at integrals. Integrals are the third and final major topic that will be covered in this class. As with derivatives this chapter will be devoted almost exclusively to finding and computing integrals. Applications will be given in the following chapter. There are really two types of integrals that we'll be looking at in this chapter : Indefinite Integrals and Definite Integrals. The first half of this chapter is devoted to indefinite integrals and the last half is devoted to definite integrals. As we will see in the last half of the chapter if we don't know indefinite integrals we will not be able to do definite integrals.

Indefinite Integrals - In this section we will start off the chapter with the definition and properties of indefinite integrals. We will not be computing many indefinite integrals in this section. This section is devoted to simply defining what an indefinite integral is and to give many of the properties of the indefinite integral. Actually computing indefinite integrals will start in the next section.

Computing Indefinite Integrals - In this section we will compute some indefinite integrals. The integrals in this section will tend to be those that do not require a lot of manipulation of the function we are integrating in order to actually compute the integral. As we will see starting in the next section many integrals do require some manipulation of the function before we can actually do the integral. We will also take a quick look at an application of indefinite integrals.

Substitution Rule for Indefinite Integrals - In this section we will start using one of the more common and useful integration techniques - The Substitution Rule. With the substitution rule we will be able integrate a wider variety of functions. The integrals in this section will all require some manipulation of the function prior to integrating unlike most of the integrals from the previous section where all we really needed were the basic integration formulas.

More Substitution Rule - In this section we will continue to look at the substitution rule. The problems in this section will tend to be a little more involved than those in the previous section.

Area Problem - In this section we start off with the motivation for definite integrals and give one of the interpretations of definite integrals. We will be approximating the amount of area that lies between a function and the x-axis. As we will see in the next section this problem will lead us to the definition of the definite integral and will be one of the main interpretations of the definite integral that we'll be looking at in this material.

Definition of the Definite Integral - In this section we will formally define the definite integral, give many of its properties and discuss a couple of interpretations of the definite integral. We will also look at the first part of the Fundamental Theorem of Calculus which shows the very close relationship between derivatives and integrals

Computing Definite Integrals - In this section we will take a look at the second part of the Fundamental Theorem of Calculus. This will show us how we compute definite integrals without using (the often very unpleasant) definition. The examples in this section can all be done with a basic knowledge of indefinite integrals and will not require the use of the substitution rule. Included in the examples in this section are computing definite integrals of piecewise and absolute value functions.

Substitution Rule for Definite Integrals - In this section we will revisit the substitution rule as it applies to definite integrals. The only real requirements to being able to do the examples in this section are being able to do the substitution rule for indefinite integrals and understanding how to compute definite integrals in general.

Applications of Integrals In this last chapter of this course we will be taking a look at a couple of Applications of Integrals. There are many other applications, however many of them require integration techniques that are typically taught in Calculus II. We will therefore be focusing on applications that can be done only with knowledge taught in this course.

Because this chapter is focused on the applications of integrals it is assumed in all the examples that you are capable of doing the integrals. There will not be as much detail in the integration process in the examples in this chapter as there was in the examples in the previous chapter.

Average Function Value - In this section we will look at using definite integrals to determine the average value of a function on an interval. We will also give the Mean Value Theorem for Integrals.

Area Between Curves - In this section we'll take a look at one of the main applications of definite integrals in this chapter. We will determine the area of the region bounded by two curves.

Volumes of Solids of Revolution / Method of Rings - In this section, the first of two sections devoted to finding the volume of a solid of revolution, we will look at the method of rings/disks to find the volume of the object we get by rotating a region bounded by two curves (one of which may be the x or y-axis) around a vertical or horizontal axis of rotation.

Volumes of Solids of Revolution / Method of Cylinders - In this section, the second of two sections devoted to finding the volume of a solid of revolution, we will look at the method of cylinders/shells to find the volume of the object we get by rotating a region bounded by two curves (one of which may be the x or y-axis) around a vertical or horizontal axis of rotation.

More Volume Problems - In the previous two sections we looked at solids that could be found by treating them as a solid of revolution. Not all solids can be thought of as solids of revolution and, in fact, not all solids of revolution can be easily dealt with using the methods from the previous two sections. So, in this section we'll take a look at finding the volume of some solids that are either not solids of revolutions or are not easy to do as a solid of revolution.

Work - In this section we will look at is determining the amount of work required to move an object subject to a force over a given distance.

Integration Techniques In this chapter we are going to be looking at various integration techniques. There are a fair number of them and some will be easier than others. The point of the chapter is to teach you these new techniques and so this chapter assumes that you've got a fairly good working knowledge of basic integration as well as substitutions with integrals. In fact, most integrals involving "simple" substitutions will not have any of the substitution work shown. It is going to be assumed that you can verify the substitution portion of the integration yourself.

Also, most of the integrals done in this chapter will be indefinite integrals. It is also assumed that once you can do the indefinite integrals you can also do the definite integrals and so to conserve space we concentrate mostly on indefinite integrals. There is one exception to this and that is the Trig Substitution section and in this case there are some subtleties involved with definite integrals that we're going to have to watch out for. Outside of that however, most sections will have at most one definite integral example and some sections will not have any definite integral examples.

Integration by Parts - In this section we will be looking at Integration by Parts. Of all the techniques we'll be looking at in this class this is the technique that students are most likely to run into down the road in other classes. We also give a derivation of the integration by parts formula.

Integrals Involving Trig Functions - In this section we look at integrals that involve trig functions. In particular we concentrate integrating products of sines and cosines as well as products of secants and tangents. We will also briefly look at how to modify the work for products of these trig functions for some quotients of trig functions.

Trig Substitutions - In this section we will look at integrals (both indefinite and definite) that require the use of a substitutions involving trig functions and how they can be used to simplify certain integrals.

Partial Fractions - In this section we will use partial fractions to rewrite integrands into a form that will allow us to do integrals involving some rational functions.

Integrals Involving Roots - In this section we will take a look at a substitution that can, on occasion, be used with integrals involving roots.

Integrals Involving Quadratics - In this section we are going to look at some integrals that involve quadratics for which the previous techniques won't work right away. In some cases, manipulation of the quadratic needs to be done before we can do the integral. We will see several cases where this is needed in this section.

Integration Strategy - In this section we give a general set of guidelines for determining how to evaluate an integral. The guidelines give here involve a mix of both Calculus I and Calculus II techniques to be as general as possible. Also note that there really isn't one set of guidelines that will always work and so you always need to be flexible in following this set of guidelines.

Improper Integrals - In this section we will look at integrals with infinite intervals of integration and integrals with discontinuous integrands in this section. Collectively, they are called improper integrals and as we will see they may or may not have a finite (*i.e.* not infinite) value. Determining if they have finite values will, in fact, be one of the major topics of this section.

Comparison Test for Improper Integrals - It will not always be possible to evaluate improper integrals and yet we still need to determine if they converge or diverge (*i.e.* if they have a finite value or not). So, in this section we will use the Comparison Test to determine if improper integrals converge or diverge.

Approximating Definite Integrals - In this section we will look at several fairly simple methods of approximating the value of a definite integral. It is not possible to evaluate every definite integral (*i.e.* because it is not possible to do the indefinite integral) and yet we may need to know the value of the definite integral anyway. These methods allow us to at least get an approximate value which may be enough in a lot of cases.

More Applications of Integrals In this section we're going to take a look at some of the Applications of Integrals. It should be noted as well that these applications are presented here, as opposed to Calculus I, simply because many of the integrals that arise from these applications tend to require techniques that we discussed in the previous chapter.

Arc Length - In this section we'll determine the length of a curve over a given interval.

Surface Area - In this section we'll determine the surface area of a solid of revolution, *i.e.* a solid obtained by rotating a region bounded by two curves about a vertical or horizontal axis.

Center of Mass - In this section we will determine the center of mass or centroid of a thin plate where the plate can be described as a region bounded by two curves (one of which may the x or y-axis).

Hydrostatic Pressure and Force - In this section we'll determine the hydrostatic pressure and force on a vertical plate submerged in water. The plates used in the examples can all be described as regions bounded by one or more curves/lines.

Probability - Many quantities can be described with probability density functions. For example, the length of time a person waits in line at a checkout counter or the life span of a light bulb. None of these quantities are fixed values and will depend on a variety of factors. In this section we will look at probability density functions and computing the mean (think average wait in line or average life span of a light blub) of a probability density function.

Parametric Equations and Polar Coordinates In this section we will be looking at parametric equations and polar coordinates. While the two subjects don't appear to have that much in common on the surface we will see that several of the topics in polar coordinates can be done in terms of parametric equations and so in that sense they make a good match in this chapter

We will also be looking at how to do many of the standard calculus topics such as tangents and area in terms of parametric equations and polar coordinates.

Parametric Equations and Curves - In this section we will introduce parametric equations and parametric curves (i.e. graphs of parametric equations). We will graph several sets of parametric equations and discuss how to eliminate the parameter to get an algebraic equation which will often help with the graphing process.

Tangents with Parametric Equations - In this section we will discuss how to find the derivatives $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ for parametric curves. We will also discuss using these derivative formulas to find the tangent line for parametric curves as well as determining where a parametric curve in increasing/decreasing and concave up/concave down.

Area with Parametric Equations - In this section we will discuss how to find the area between a parametric curve and the x-axis using only the parametric equations (rather than eliminating the parameter and using standard Calculus I techniques on the resulting algebraic equation).

Arc Length with Parametric Equations - In this section we will discuss how to find the arc length of a parametric curve using only the parametric equations (rather than eliminating the parameter and using standard Calculus techniques on the resulting algebraic equation).

Surface Area with Parametric Equations - In this section we will discuss how to find the surface area of a solid obtained by rotating a parametric curve about the x or y-axis using only the parametric equations (rather than eliminating the parameter and using standard Calculus techniques on the resulting algebraic equation).

Polar Coordinates - In this section we will introduce polar coordinates an alternative coordinate system to the 'normal' Cartesian/Rectangular coordinate system. We will derive formulas to convert between polar and Cartesian coordinate systems. We will also look at many of the standard polar graphs as well as circles and some equations of lines in terms of polar coordinates.

Tangents with Polar Coordinates - In this section we will discuss how to find the derivative $\frac{dy}{dx}$ for polar curves. We will also discuss using this derivative formula to find the tangent line for polar curves using only polar coordinates (rather than converting to Cartesian coordinates and using standard Calculus techniques).

Area with Polar Coordinates - In this section we will discuss how to the area enclosed by a polar curve. The regions we look at in this section tend (although not always) to be shaped vaguely like a piece of pie or pizza and we are looking for the area of the region from the outer boundary (defined by the polar equation) and the origin/pole. We will also discuss finding the area between two polar curves.

Arc Length with Polar Coordinates - In this section we will discuss how to find the arc length of a polar curve using only polar coordinates (rather than converting to Cartesian coordinates and using standard Calculus techniques).

Surface Area with Polar Coordinates - In this section we will discuss how to find the surface area of a solid obtained by rotating a polar curve about the x or y-axis using only polar coordinates (rather than converting to Cartesian coordinates and using standard Calculus techniques).

Arc Length and Surface Area Revisited - In this section we will summarize all the arc length and surface area formulas we developed over the course of the last two chapters.

Series and Sequences In this chapter we'll be taking a look at sequences and (infinite) series. In fact, this chapter will deal almost exclusively with series. However, we also need to understand some of the basics of sequences in order to properly deal with series. We will therefore, spend a little time on sequences as well.

Series is one of those topics that many students don't find all that useful. To be honest, many students will never see series outside of their calculus class. However, series do play an important role in the field of ordinary differential equations and without series large portions of the field of partial differential equations would not be possible.

In other words, series is an important topic even if you won't ever see any of the applications. Most of the applications are beyond the scope of most Calculus courses and tend to occur in classes that many students don't take. So, as you go through this material keep in mind that these do have applications even if we won't really be covering many of them in this class.

Sequences - In this section we define just what we mean by sequence in a math class and give the basic notation we will use with them. We will focus on the basic terminology, limits of sequences and convergence of sequences in this section. We will also give many of the basic facts and properties we'll need as we work with sequences.

More on Sequences - In this section we will continue examining sequences. We will determine if a sequence in an increasing sequence or a decreasing sequence and hence if it is a monotonic sequence. We will also determine a sequence is bounded below, bounded above and/or bounded.

Series - The Basics - In this section we will formally define an infinite series. We will also give many of the basic facts, properties and ways we can use to manipulate a series. We will also briefly discuss how to determine if an infinite series will converge or diverge (a more in depth discussion of this topic will occur in the next section).

Convergence/Divergence of Series - In this section we will discuss in greater detail the convergence and divergence of infinite series. We will illustrate how partial sums are used to determine if an infinite series converges or diverges. We will also give the Divergence Test for series in this section.

Special Series - In this section we will look at three series that either show up regularly or have some nice properties that we wish to discuss. We will examine Geometric Series, Telescoping Series, and Harmonic Series.

Integral Test - In this section we will discuss using the Integral Test to determine if an infinite series converges or diverges. The Integral Test can be used on a infinite series provided the terms of the series are positive and decreasing. A proof of the Integral Test is also given.

Comparison Test/Limit Comparison Test - In this section we will discuss using the Comparison Test and Limit Comparison Tests to determine if an infinite series converges or diverges. In order to use either test the terms of the infinite series must be positive. Proofs for both tests are also given.

Alternating Series Test - In this section we will discuss using the Alternating Series Test to determine if an infinite series converges or diverges. The Alternating Series Test can be used only if the terms of the series alternate in sign. A proof of the Alternating Series Test is also given.

Absolute Convergence - In this section we will have a brief discussion on absolute convergence and conditionally convergent and how they relate to convergence of infinite series.

Ratio Test - In this section we will discuss using the Ratio Test to determine if an infinite series converges absolutely or diverges. The Ratio Test can be used on any series, but unfortunately will not always yield a conclusive answer as to whether a series will converge absolutely or diverge. A proof of the Ratio Test is also given.

Root Test - In this section we will discuss using the Root Test to determine if an infinite series converges absolutely or diverges. The Root Test can be used on any series, but unfortunately will not always yield a conclusive answer as to whether a series will converge absolutely or diverge. A proof of the Root Test is also given.

Strategy for Series - In this section we give a general set of guidelines for determining which test to use in determining if an infinite series will converge or diverge. Note as well that there really isn't one set of guidelines that will always work and so you always need to be flexible in following this set of guidelines. A summary of all the various tests, as well as conditions that must be met to use them, we discussed in this chapter are also given in this section.

Estimating the Value of a Series - In this section we will discuss how the Integral Test, Comparison Test, Alternating Series Test and the Ratio Test can, on occasion, be used to estimating the value of an infinite series.

Power Series - In this section we will give the definition of the power series as well as the definition of the radius of convergence and interval of convergence for a power series. We will also illustrate how the Ratio Test and Root Test can be used to determine the radius and interval of convergence for a power series.

Power Series and Functions - In this section we discuss how the formula for a convergent Geometric Series can be used to represent some functions as power series. To use the Geometric Series formula, the function must be able to be put into a specific form, which is often impossible. However, use of this formula does quickly illustrate how functions can be represented as a power series. We also discuss differentiation and integration of power series.

Taylor Series - In this section we will discuss how to find the Taylor/Maclaurin Series for a function. This will work for a much wider variety of function than the method discussed in the previous section at the expense of some often unpleasant work. We also derive some well known formulas for Taylor series of $\mathbf{e}^x$, $\cos(x)$ and $\sin(x)$ around $x = 0$.

Applications of Series - In this section we will take a quick look at a couple of applications of series. We will illustrate how we can find a series representation for indefinite integrals that cannot be evaluated by any other method. We will also see how we can use the first few terms of a power series to approximate a function.

Binomial Series - In this section we will give the Binomial Theorem and illustrate how it can be used to quickly expand terms in the form $(a + b)^n$ when n is an integer. In addition, when n is not an integer an extension to the Binomial Theorem can be used to give a power series representation of the term.

Vectors This is a fairly short chapter. We will be taking a brief look at vectors and some of their properties. We will need some of this material in the next chapter and those of you heading on towards Calculus III will use a fair amount of this there as well.

Basic Concepts - In this section we will introduce some common notation for vectors as well as some of the basic concepts about vectors such as the magnitude of a vector and unit vectors. We also illustrate how to find a vector from its starting and end points.

Vector Arithmetic - In this section we will discuss the mathematical and geometric interpretation of the sum and difference of two vectors. We also define and give a geometric interpretation for scalar multiplication. We also give some of the basic properties of vector arithmetic and introduce the common $\vec{i}$, $\vec{j}$, $\vec{k}$ notation for vectors.

Dot Product - In this section we will define the dot product of two vectors. We give some of the basic properties of dot products and define orthogonal vectors and show how to use the dot product to determine if two vectors are orthogonal. We also discuss finding vector projections and direction cosines in this section.

Cross Product - In this section we define the cross product of two vectors and give some of the basic facts and properties of cross products.

Three Dimensional Space In this chapter we will start taking a more detailed look at three dimensional space (3-D space or $\mathbb{R}^3$). This is a very important topic for Calculus III since a good portion of Calculus III is done in three (or higher) dimensional space.

We will be looking at the equations of graphs in 3-D space as well as vector valued functions and how we do calculus with them. We will also be taking a look at a couple of new coordinate systems for 3-D space.

The 3-D Coordinate System - In this section we will introduce the standard three dimensional coordinate system as well as some common notation and concepts needed to work in three dimensions.

Equations of Lines - In this section we will derive the vector form and parametric form for the equation of lines in three dimensional space. We will also give the symmetric equations of lines in three dimensional space. Note as well that while these forms can also be useful for lines in two dimensional space.

Equations of Planes - In this section we will derive the vector and scalar equation of a plane. We also show how to write the equation of a plane from three points that lie in the plane.

Quadric Surfaces - In this section we will be looking at some examples of quadric surfaces. Some examples of quadric surfaces are cones, cylinders, ellipsoids, and elliptic paraboloids.

Functions of Several Variables - In this section we will give a quick review of some important topics about functions of several variables. In particular we will discuss finding the domain of a function of several variables as well as level curves, level surfaces and traces.

Vector Functions - In this section we introduce the concept of vector functions concentrating primarily on curves in three dimensional space. We will however, touch briefly on surfaces as well. We will illustrate how to find the domain of a vector function and how to graph a vector function. We will also show a simple relationship between vector functions and parametric equations that will be very useful at times.

Calculus with Vector Functions - In this section here we discuss how to do basic calculus, i.e. limits, derivatives and integrals, with vector functions.

Tangent, Normal and Binormal Vectors - In this section we will define the tangent, normal and binormal vectors.

Arc Length with Vector Functions - In this section we will extend the arc length formula we used early in the material to include finding the arc length of a vector function. As we will see the new formula really is just an almost natural extension of one we've already seen.

Curvature - In this section we give two formulas for computing the curvature (*i.e.* how fast the function is changing at a given point) of a vector function.

Velocity and Acceleration - In this section we will revisit a standard application of derivatives, the velocity and acceleration of an object whose position function is given by a vector function. For the acceleration we give formulas for both the normal acceleration and the tangential acceleration.

Cylindrical Coordinates - In this section we will define the cylindrical coordinate system, an alternate coordinate system for the three dimensional coordinate system. As we will see cylindrical coordinates are really nothing more than a very natural extension of polar coordinates into a three dimensional setting.

Spherical Coordinates - In this section we will define the spherical coordinate system, yet another alternate coordinate system for the three dimensional coordinate system. This coordinates system is very useful for dealing with spherical objects. We will derive formulas to convert between cylindrical coordinates and spherical coordinates as well as between Cartesian and spherical coordinates (the more useful of the two).

Partial Derivatives In Calculus I and in most of Calculus II we concentrated on functions of one variable. In Calculus III we will extend our knowledge of calculus into functions of two or more variables. Despite the fact that this chapter is about derivatives we will start out the chapter with a section on limits of functions of more than one variable. In the remainder of this chapter we will be looking at differentiating functions of more than one variable. As we will see, while there are differences with derivatives of functions of one variable, if you can do derivatives of functions of one variable you shouldn't have any problems differentiating functions of more than one variable. You'll just need to keep one subtlety in mind as we do the work.

Limits - In the section we'll take a quick look at evaluating limits of functions of several variables. We will also see a fairly quick method that can be used, on occasion, for showing that some limits do not exist.

Partial Derivatives - In this section we will look at the idea of partial derivatives. We will give the formal definition of the partial derivative as well as the standard notations and how to compute them in practice (*i.e.* without the use of the definition). As you will see if you can do derivatives of functions of one variable you won't have much of an issue with partial derivatives. There is only one (very important) subtlety that you need to always keep in mind while computing partial derivatives.

Interpretations of Partial Derivatives - In the section we will take a look at a couple of important interpretations of partial derivatives. First, the always important, rate of change of the function. Although we now have multiple 'directions' in which the function can change (unlike in Calculus I). We will also see that partial derivatives give the slope of tangent lines to the traces of the function.

Higher Order Partial Derivatives - In the section we will take a look at higher order partial derivatives. Unlike Calculus I however, we will have multiple second order derivatives, multiple third order derivatives, *etc.* because we are now working with functions of multiple variables. We will also discuss Clairaut's Theorem to help with some of the work in finding higher order derivatives.

Differentials - In this section we extend the idea of differentials we first saw in Calculus I to functions of several variables.

Chain Rule - In the section we extend the idea of the chain rule to functions of several variables. In particular, we will see that there are multiple variants to the chain rule here all depending on how many variables our function is dependent on and how each of those variables can, in turn, be written in terms of different variables. We will also give a nice method for writing down the chain rule for pretty much any situation you might run into when dealing with functions of multiple variables. In addition, we will derive a very quick way of doing implicit differentiation so we no longer need to go through the process we first did back in Calculus I.

Directional Derivatives - In the section we introduce the concept of directional derivatives. With directional derivatives we can now ask how a function is changing if we allow all the independent variables to change rather than holding all but one constant as we had to do with partial derivatives. In addition, we will define the gradient vector to help with some of the notation and work here. The gradient vector will be very useful in some later sections as well. We will also give a nice fact that will allow us to determine the direction in which a given function is changing the fastest.

Applications of Partial Derivatives In this chapter we will take a look at a several applications of partial derivatives. Most of the applications will be extensions to applications to ordinary derivatives that we saw back in Calculus I. For instance, we will be looking at finding the absolute and relative extrema of a function and we will also be looking at optimization. Both (all three?) of these subjects were major applications back in Calculus I. They will, however, be a little more work here because we now have more than one variable.

Tangent Planes and Linear Approximations - In this section formally define just what a tangent plane to a surface is and how we use partial derivatives to find the equations of tangent planes to surfaces that can be written as $z = f(x, y)$. We will also see how tangent planes can be thought of as a linear approximation to the surface at a given point.

Gradient Vector, Tangent Planes and Normal Lines - In this section discuss how the gradient vector can be used to find tangent planes to a much more general function than in the previous section. We will also define the normal line and discuss how the gradient vector can be used to find the equation of the normal line.

Relative Minimums and Maximums - In this section we will define critical points for functions of two variables and discuss a method for determining if they are relative minimums, relative maximums or saddle points (i.e. neither a relative minimum or relative maximum).

Absolute Minimums and Maximums - In this section we will how to find the absolute extrema of a function of two variables when the independent variables are only allowed to come from a region that is bounded (*i.e.* no part of the region goes out to infinity) and closed (*i.e.* all of the points on the boundary are valid points that can be used in the process).

Lagrange Multipliers - In this section we'll see discuss how to use the method of Lagrange Multipliers to find the absolute minimums and maximums of functions of two or three variables in which the independent variables are subject to one or more constraints. We also give a brief justification for how/why the method works.

Multiple Integrals In Calculus I we moved on to the subject of integrals once we had finished the discussion of derivatives. The same is true in this course. Now that we have finished our discussion of derivatives of functions of more than one variable we need to move on to integrals of functions of two or three variables.

Most of the derivatives topics extended somewhat naturally from their Calculus I counterparts and that will be the same here. However, because we are now involving functions of two or three variables there will be some differences as well. There will be new notation and some new issues that simply don't arise when dealing with functions of a single variable.

Vector Fields - In this section we introduce the concept of a vector field and give several examples of graphing them. We also revisit the gradient that we first saw a few chapters ago.

Line Integrals - Part I - In this section we will start off with a quick review of parameterizing curves. This is a skill that will be required in a great many of the line integrals we evaluate and so needs to be understood. We will then formally define the first kind of line integral we will be looking at : line integrals with respect to arc length..

Line Integrals - Part II - In this section we will continue looking at line integrals and define the second kind of line integral we'll be looking at : line integrals with respect to x, y, and/or z. We also introduce an alternate form of notation for this kind of line integral that will be useful on occasion.

Line Integrals of Vector Fields - In this section we will define the third type of line integrals we'll be looking at : line integrals of vector fields. We will also see that this particular kind of line integral is related to special cases of the line integrals with respect to x, y and z.

Fundamental Theorem for Line Integrals - In this section we will give the fundamental theorem of calculus for line integrals of vector fields. This will illustrate that certain kinds of line integrals can be very quickly computed. We will also give quite a few definitions and facts that will be useful.

Conservative Vector Fields - In this section we will take a more detailed look at conservative vector fields than we've done in previous sections. We will also discuss how to find potential functions for conservative vector fields.

Green's Theorem - In this section we will discuss Green's Theorem as well as an interesting application of Green's Theorem that we can use to find the area of a two dimensional region.

Surface Integrals In the previous chapter we looked at evaluating integrals of functions or vector fields where the points came from a curve in two- or three-dimensional space. We now want to extend this idea and integrate functions and vector fields where the points come from a surface in three-dimensional space. These integrals are called surface integrals.

Curl and Divergence - In this section we will introduce the concepts of the curl and the divergence of a vector field. We will also give two vector forms of Green's Theorem and show how the curl can be used to identify if a three dimensional vector field is conservative field or not.

Parametric Surfaces - In this section we will take a look at the basics of representing a surface with parametric equations. We will also see how the parameterization of a surface can be used to find a normal vector for the surface (which will be very useful in a couple of sections) and how the parameterization can be used to find the surface area of a surface.

Surface Integrals - In this section we introduce the idea of a surface integral. With surface integrals we will be integrating over the surface of a solid. In other words, the variables will always be on the surface of the solid and will never come from inside the solid itself. Also, in this section we will be working with the first kind of surface integrals we'll be looking at in this chapter : surface integrals of functions.

Surface Integrals of Vector Fields - In this section we will introduce the concept of an oriented surface and look at the second kind of surface integral we'll be looking at : surface integrals of vector fields.

Stokes' Theorem - In this section we will discuss Stokes' Theorem.

Divergence Theorem - In this section we will discuss the Divergence Theorem.

Extras In this appendix is material that didn't fit into other sections for a variety of reasons. Also, in order to not obscure the mechanics of actually working problems, most of the proofs of various facts and formulas are in this chapter as opposed to being in the section with the fact/formula.

Proof of Various Limit Properties - In this section we prove several of the limit properties and facts that were given in various sections of the Limits chapter.

Proof of Various Derivative Facts/Formulas/Properties - In this section we prove several of the rules/formulas/properties of derivatives that we saw in Derivatives Chapter.

Proof of Trig Limits - In this section we give proofs for the two limits that are needed to find the derivative of the sine and cosine functions using the definition of the derivative.

Proofs of Derivative Applications Facts/Formulas - In this section we prove many of the facts that we saw in the Applications of Derivatives chapter.

Proof of Various Integral Facts/Formulas/Properties - In this section we prove some of the facts and formulas from the Integral Chapter as well as a couple from the Applications of Integrals chapter.

Area and Volume Formulas - In this section we derive the formulas for finding area between two curves and finding the volume of a solid of revolution.

Types of Infinity - In this section we have a discussion on the types of infinity and how these affect certain limits. Note that there is a lot of theory going on 'behind the scenes' so to speak that we are not going to cover in this section. This section is intended only to give you a feel for what is going on here. To get a fuller understanding of some of the ideas in this section you will need to take some upper level mathematics courses.

Summation Notation - In this section we give a quick review of summation notation. Summation notation is heavily used when defining the definite integral and when we first talk about determining the area between a curve and the x-axis.

Constant of Integration - In this section we have a discussion on a couple of subtleties involving constants of integration that many students don't think about when doing indefinite integrals. Not understanding these subtleties can lead to confusion on occasion when students get different answers to the same integral. We include two examples of this kind of situation.

1 Review

Technically a student coming into a Calculus class is supposed to know both Algebra and Trigonometry. Unfortunately, the reality is often much different. Most students enter a Calculus class woefully unprepared for both the algebra and the trig that is in a Calculus class. This is very unfortunate since good algebra skills are absolutely vital to successfully completing any Calculus course and if your Calculus course includes trig (as this one does) good trig skills are also important in many sections.

The above statement is not meant to denigrate your favorite Algebra or Trig instructor. It is simply an acknowledgment of the fact that many of these courses, especially Algebra courses, are aimed at a more general audience and so do not always put the time into topics that are vital to a Calculus course and/or the level of difficulty is kept lower than might be best for students heading on towards Calculus.

Far too often the biggest impediment to students being successful in a Calculus course is they do not have sufficient skills in the underlying algebra and trig that will be in many of the calculus problems we'll be looking at. These students end up struggling with the algebra and trig in the problems rather than working to understand the calculus topics which in turn negatively impacts their grade in a Calculus course. The intent of this chapter, therefore, is to do a very cursory review of some algebra and trig skills that are vital to a calculus course that many students just didn't learn as well as they should have from their Algebra and Trig courses.

This chapter does not include all the algebra and trig skills that are needed to be successful in a Calculus course. It only includes those topics that most students are particularly deficient in. For instance, factoring is also vital to completing a standard calculus class but is not included here as it is assumed that if you are taking a Calculus course then you do know how to factor. Likewise, it is assumed that if you are taking a Calculus course then you know how to solve linear and quadratic equations so those topics are not covered here either. For a more in depth review of Algebra topics you should check out the full set of Algebra notes at http://tutorial.math.lamar.edu.

Note that even though these topics are very important to a Calculus class we rarely cover all of them in the actual class itself. We simply don't have the time to do that. We will cover certain portions of this chapter in class, but for the most part we leave it to the students to read this chapter on their own to make sure they are ready for these topics as they arise in class.

1.1 Functions

In this section we're going to make sure that you're familiar with functions and function notation. Both will appear in almost every section in a Calculus class so you will need to be able to deal with them.

First, what exactly is a function? The simplest definition is an equation will be a function if, for any x in the domain of the equation (the domain is all the x's that can be plugged into the equation), the equation will yield exactly one value of y when we evaluate the equation at a specific x.

This is usually easier to understand with an example.

Example 1

Determine if each of the following are functions.

(a) $y = x^2 + 1$

(b) $y^2 = x + 1$

Solution

(a) $y = x^2 + 1$

This first one is a function. Given an x, there is only one way to square it and then add 1 to the result. So, no matter what value of x you put into the equation, there is only one possible value of y when we evaluate the equation at that value of x.

(b) $y^2 = x + 1$

The only difference between this equation and the first is that we moved the exponent off the x and onto the y. This small change is all that is required, in this case, to change the equation from a function to something that isn't a function.

To see that this isn't a function is fairly simple. Choose a value of x, say $x = 3$ and plug this into the equation.

$$y^2 = 3 + 1 = 4$$

Now, there are two possible values of y that we could use here. We could use $y = 2$ or $y = -2$. Since there are two possible values of y that we get from a single x this equation isn't a function.

Note that this only needs to be the case for a single value of x to make an equation not be a function. For instance, we could have used $x = -1$ and in this case, we would get a single y ($y = 0$). However, because of what happens at $x = 3$ this equation will not be a function.

Next, we need to take a quick look at function notation. Function notation is nothing more than a fancy way of writing the y in a function that will allow us to simplify notation and some of our work a little.

Let's take a look at the following function.

$$y = 2x^2 - 5x + 3$$

Using function notation, we can write this as any of the following.

$$\begin{aligned} f(x) &= 2x^2 - 5x + 3 & g(x) &= 2x^2 - 5x + 3 \\ h(x) &= 2x^2 - 5x + 3 & R(x) &= 2x^2 - 5x + 3 \\ w(x) &= 2x^2 - 5x + 3 & y(x) &= 2x^2 - 5x + 3 \\ & & \vdots & \end{aligned}$$

Recall that this is NOT a letter times x, this is just a fancy way of writing y.

So, why is this useful? Well let's take the function above and let's get the value of the function at $x = -3$. Using function notation we represent the value of the function at $x = -3$ as $f(-3)$. Function notation gives us a nice compact way of representing function values.

Now, how do we actually evaluate the function? That's really simple. Everywhere we see an x on the right side we will substitute whatever is in the parenthesis on the left side. For our function this gives,

$$\begin{aligned} f(-3) &= 2(-3)^2 - 5(-3) + 3 \\ &= 2(9) + 15 + 3 \\ &= 36 \end{aligned}$$

Let's take a look at some more function evaluation.

Example 2

Given $f(x) = -x^2 + 6x - 11$ find each of the following.

(a) $f(2)$

(b) $f(-10)$

(c) $f(t)$

(d) $f(t-3)$

(e) $f(x-3)$

(f) $f(4x-1)$

Solution

(a) $f(2)$

$$f(2) = -(2)^2 + 6(2) - 11 = -3$$

(b) $f(-10)$

$$f(-10) = -(-10)^2 + 6(-10) - 11 = -100 - 60 - 11 = -171$$

Be careful when squaring negative numbers!

(c) $f(t)$

$$f(t) = -t^2 + 6t - 11$$

Remember that we substitute for the x's WHATEVER is in the parenthesis on the left. Often this will be something other than a number. So, in this case we put t's in for all the x's on the left.

(d) $f(t-3)$

$$f(t-3) = -(t-3)^2 + 6(t-3) - 11 = -t^2 + 12t - 38$$

Often instead of evaluating functions at numbers or single letters we will have some fairly complex evaluations so make sure that you can do these kinds of evaluations.

(e) $f(x-3)$

$$f(x-3) = -(x-3)^2 + 6(x-3) - 11 = -x^2 + 12x - 38$$

The only difference between this one and the previous one is that we changed the t to an x. Other than that, there is absolutely no difference between the two! Don't get excited if an x appears inside the parenthesis on the left.

(f) $f(4x-1)$

$$f(4x-1) = -(4x-1)^2 + 6(4x-1) - 11 = -16x^2 + 32x - 18$$

This one is not much different from the previous part. All we did was change the equation that we were plugging into the function.

All throughout a calculus course we will be finding roots of functions. A root of a function is nothing more than a number for which the function is zero. In other words, finding the roots of a function, $g(x)$, is equivalent to solving

$$g(x) = 0$$

Example 3

Determine all the roots of $f(t) = 9t^3 - 18t^2 + 6t$

Solution

So, we will need to solve,

$$9t^3 - 18t^2 + 6t = 0$$

First, we should factor the equation as much as possible. Doing this gives,

$$3t\left(3t^2 - 6t + 2\right) = 0$$

Next recall that if a product of two things are zero then one (or both) of them had to be zero. This means that,

$$3t = 0 \qquad \text{OR} \qquad 3t^2 - 6t + 2 = 0$$

From the first it's clear that one of the roots must then be $t = 0$. To get the remaining roots we will need to use the quadratic formula on the second equation. Doing this gives,

$$\begin{aligned} t &= \frac{-(-6) \pm \sqrt{(-6)^2 - 4(3)(2)}}{2(3)} \\ &= \frac{6 \pm \sqrt{12}}{6} \\ &= \frac{6 \pm \sqrt{(4)(3)}}{6} \\ &= \frac{6 \pm 2\sqrt{3}}{6} \\ &= \frac{3 \pm \sqrt{3}}{3} \\ &= 1 \pm \frac{1}{3}\sqrt{3} = 1 \pm \frac{1}{\sqrt{3}} \end{aligned}$$

In order to remind you how to simplify radicals we gave several forms of the answer.

To complete the problem, here is a complete list of all the roots of this function.

$$t = 0, \ t = \frac{3 + \sqrt{3}}{3}, \ t = \frac{3 - \sqrt{3}}{3}$$

Note we didn't use the final form for the roots from the quadratic. This is usually where we'll stop with the simplification for these kinds of roots. Also note that, for the sake of the practice, we broke up the compact form for the two roots of the quadratic. You will need to be able to do this so make sure that you can.

This example had a couple of points other than finding roots of functions.

The first was to remind you of the quadratic formula. This won't be the last time that you'll need it in this class.

The second was to get you used to seeing "messy" answers. In fact, the answers in the above example are not really all that messy. However, most students come out of an Algebra class very used to seeing only integers and the occasional "nice" fraction as answers.

So, here is fair warning. In this class I often will intentionally make the answers look "messy" just to get you out of the habit of always expecting "nice" answers. In "real life" (whatever that is) the answer is rarely a simple integer such as two. In most problems the answer will be a decimal that came about from a messy fraction and/or an answer that involved radicals.

One of the more important ideas about functions is that of the **domain** and **range** of a function. In simplest terms the domain of a function is the set of all values that can be plugged into a function and have the function exist and have a real number for a value. So, for the domain we need to avoid division by zero, square roots of negative numbers, logarithms of zero and logarithms of negative numbers (if not familiar with logarithms we'll take a look at them a little later), *etc*. The range of a function is simply the set of all possible values that a function can take.

Let's find the domain and range of a few functions.

Example 4

Find the domain and range of each of the following functions.

(a) $f(x) = 5x - 3$

(b) $g(t) = \sqrt{4 - 7t}$

(c) $h(x) = -2x^2 + 12x + 5$

(d) $f(z) = |z - 6| - 3$

(e) $g(x) = 8$

Solution

(a) $f(x) = 5x - 3$

We know that this is a line and that it's not a horizontal line (because the slope is 5 and not zero...). This means that this function can take on any value and so the range is all real numbers. Using "mathematical" notation this is,

$$\text{Range : } \quad (-\infty, \infty)$$

This is more generally a polynomial and we know that we can plug any value into a polynomial and so the domain in this case is also all real numbers or,

$$\text{Domain : } -\infty < x < \infty \quad \text{or} \quad (-\infty, \infty)$$

(b) $g(t) = \sqrt{4 - 7t}$

This is a square root and we know that square roots are always positive or zero. We know then that the range will be,

$$\text{Range : } [0, \infty)$$

For the domain we have a little bit of work to do, but not much. We need to make sure that we don't take square roots of any negative numbers, so we need to require that,

$$\begin{aligned} 4 - 7t &\geq 0 \\ 4 &\geq 7t \\ \frac{4}{7} &\geq t \qquad \Rightarrow \qquad t \leq \frac{4}{7} \end{aligned}$$

The domain is then,

$$\text{Domain : } t \leq \frac{4}{7} \quad \text{or} \quad \left(-\infty, \frac{4}{7}\right]$$

(c) $h(x) = -2x^2 + 12x + 5$

Here we have a quadratic, which is a polynomial, so we again know that the domain is all real numbers or,

$$\text{Domain : } -\infty < x < \infty \quad \text{or} \quad (-\infty, \infty)$$

In this case the range requires a little bit of work. From an Algebra class we know that the graph of this will be a parabola that opens down (because the coefficient of the x^2 is negative) and so the vertex will be the highest point on the graph. If we know the vertex we can then get the range. The vertex is then,

$$x = -\frac{12}{2(-2)} = 3 \qquad y = h(3) = -2(3)^2 + 12(3) + 5 = 23 \qquad \Rightarrow \qquad (3, 23)$$

So, as discussed, we know that this will be the highest point on the graph or the largest value of the function and the parabola will take all values less than this, so the range is then,

$$\text{Range : } (-\infty, 23]$$

(d) $f(z) = |z - 6| - 3$

This function contains an absolute value and we know that absolute value will be either positive or zero. In this case the absolute value will be zero if $z = 6$ and so the absolute value portion of this function will always be greater than or equal to zero. We are subtracting 3 from the absolute value portion and so we then know that the range will be,

$$\text{Range}: \quad [-3, \infty)$$

We can plug any value into an absolute value and so the domain is once again all real numbers or,

$$\text{Domain}: \quad -\infty < z < \infty \qquad \text{or} \qquad (-\infty, \infty)$$

(e) $g(x) = 8$

This function may seem a little tricky at first but is actually the easiest one in this set of examples. This is a constant function and so any value of x that we plug into the function will yield a value of 8. This means that the range is a single value or,

$$\text{Range}: \quad 8$$

The domain is all real numbers,

$$\text{Domain}: \quad -\infty < x < \infty \qquad \text{or} \qquad (-\infty, \infty)$$

In general, determining the range of a function can be somewhat difficult. As long as we restrict ourselves down to "simple" functions, some of which we looked at in the previous example, finding the range is not too bad, but for most functions it can be a difficult process.

Because of the difficulty in finding the range for a lot of functions we had to keep those in the previous set somewhat simple, which also meant that we couldn't really look at some of the more complicated domain examples that are liable to be important in a Calculus course. So, let's take a look at another set of functions only this time we'll just look for the domain.

Example 5

Find the domain of each of the following functions.

(a) $f(x) = \dfrac{x - 4}{x^2 - 2x - 15}$

(b) $g(t) = \sqrt{6 + t - t^2}$

(c) $h(x) = \dfrac{x}{\sqrt{x^2 - 9}}$

Solution

(a) $f(x) = \dfrac{x-4}{x^2-2x-15}$

Okay, with this problem we need to avoid division by zero, so we need to determine where the denominator is zero which means solving,

$$x^2 - 2x - 15 = (x-5)(x+3) = 0 \qquad \Rightarrow \qquad x = -3, \; x = 5$$

So, these are the only values of x that we need to avoid and so the domain is,

$$\text{Domain : } \quad \text{All real numbers except } x = -3 \;\; \& \; x = 5$$

(b) $g(t) = \sqrt{6 + t - t^2}$

In this case we need to avoid square roots of negative numbers and so need to require that,

$$6 + t - t^2 \ge 0 \qquad \Rightarrow \qquad t^2 - t - 6 \le 0$$

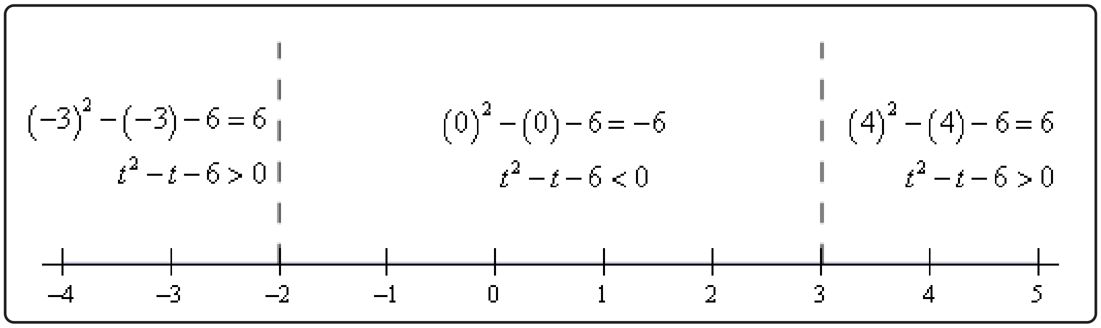
Note that we multiplied the whole inequality by -1 (and remembered to switch the direction of the inequality) to make this easier to deal with. You'll need to be able to solve inequalities like this more than a few times in a Calculus course so let's make sure you can solve these.

The first thing that we need to do is determine where the function is zero and that's not too difficult in this case.

$$t^2 - t - 6 = (t-3)(t+2) = 0$$

So, the function will be zero at $t = -2$ and $t = 3$. Recall that these points will be the only place where the function *may* change sign. It's not required to change sign at these points, but these will be the only points where the function can change sign. This means that all we need to do is break up a number line into the three regions that avoid these two points and test the sign of the function at a single point in each of the regions. If the function is positive at a single point in the region it will be positive at all points in that region because it doesn't contain the any of the points where the function may change sign. We'll have a similar situation if the function is negative for the test point.

So, here is a number line showing these computations.



From this we can see that the only region in which the quadratic (in its modified form) will be negative is in the middle region. Recalling that we got to the modified region by multiplying the quadratic by a -1 this means that the quadratic under the root will only be positive in the middle region and so the domain for this function is then,

$$\text{Domain : } -2 \le t \le 3 \quad \text{or} \quad [-2, 3]$$

(c) $h(x) = \dfrac{x}{\sqrt{x^2 - 9}}$

In this case we have a mixture of the two previous parts. We have to worry about division by zero and square roots of negative numbers. We can cover both issues by requiring that,

$$x^2 - 9 > 0$$

Note that we need the inequality here to be strictly greater than zero to avoid the division by zero issues. We can either solve this by the method from the previous example or, in this case, it is easy enough to solve by inspection. The domain is this case is,

$$\text{Domain : } x < -3 \ \& \ x > 3 \quad \text{or} \quad (-\infty, -3) \ \& \ (3, \infty)$$

The next topic that we need to discuss here is that of **function composition**. The composition of $f(x)$ and $g(x)$ is

$$(f \circ g)(x) = f(g(x))$$

In other words, compositions are evaluated by plugging the second function listed into the first function listed. Note as well that order is important here. Interchanging the order will more often than not result in a different answer.

Example 6

Given $f(x) = 3x^2 - x + 10$ and $g(x) = 1 - 20x$ find each of the following.

(a) $(f \circ g)(5)$

(b) $(f \circ g)(x)$

(c) $(g \circ f)(x)$

(d) $(g \circ g)(x)$

Solution

(a) $(f \circ g)(5)$

In this case we've got a number instead of an x but it works in exactly the same way.

$$\begin{aligned}(f \circ g)(5) &= f(g(5)) \\ &= f(-99) = 29512\end{aligned}$$

(b) $(f \circ g)(x)$

$$\begin{aligned}(f \circ g)(x) &= f(g(x)) \\ &= f(1 - 20x) \\ &= 3(1 - 20x)^2 - (1 - 20x) + 10 \\ &= 3\left(1 - 40x + 400x^2\right) - 1 + 20x + 10 \\ &= 1200x^2 - 100x + 12\end{aligned}$$

Compare this answer to the next part and notice that answers are NOT the same. The order in which the functions are listed is important!

(c) $(g \circ f)(x)$

$$\begin{aligned}(g \circ f)(x) &= g(f(x)) \\ &= g\left(3x^2 - x + 10\right) \\ &= 1 - 20\left(3x^2 - x + 10\right) \\ &= -60x^2 + 20x - 199\end{aligned}$$

And just to make the point one more time. This answer is different from the previous part. Order is important in composition.

(d) $(g \circ g)(x)$

In this case do not get excited about the fact that it's the same function. Composition still works the same way.

$$\begin{aligned}(g \circ g)(x) &= g(g(x)) \\ &= g(1 - 20x) \\ &= 1 - 20(1 - 20x) \\ &= 400x - 19\end{aligned}$$

Let's work one more example that will lead us into the next section.

Example 7

Given $f(x) = 3x - 2$ and $g(x) = \frac{1}{3}x + \frac{2}{3}$ find each of the following.

(a) $(f \circ g)(x)$

(b) $(g \circ f)(x)$

Solution

(a) $(f \circ g)(x)$

$$\begin{aligned}(f \circ g)(x) &= f(g(x)) \\ &= f\left(\frac{1}{3}x + \frac{2}{3}\right) \\ &= 3\left(\frac{1}{3}x + \frac{2}{3}\right) - 2 \\ &= x + 2 - 2 \\ &= x\end{aligned}$$

(b) $(g \circ f)(x)$

$$\begin{aligned}(g \circ f)(x) &= g(f(x)) \\ &= g(3x - 2) \\ &= \frac{1}{3}(3x - 2) + \frac{2}{3} \\ &= x - \frac{2}{3} + \frac{2}{3} \\ &= x\end{aligned}$$

In this case the two compositions were the same and in fact the answer was very simple.

$$(f \circ g)(x) = (g \circ f)(x) = x$$

This will usually not happen. However, when the two compositions are both x there is a very nice relationship between the two functions. We will take a look at that relationship in the next section.

1.2 Inverse Functions

In the last example from the previous section we looked at the two functions $f(x) = 3x - 2$ and $g(x) = \frac{x}{3} + \frac{2}{3}$ and saw that

$$(f \circ g)(x) = (g \circ f)(x) = x$$

and as noted in that section this means that there is a nice relationship between these two functions. Let's see just what that relationship is. Consider the following evaluations.

$$f(-1) = 3(-1) - 2 = -5 \qquad \Rightarrow \qquad g(-5) = \frac{-5}{3} + \frac{2}{3} = \frac{-3}{3} = -1$$

$$g(2) = \frac{2}{3} + \frac{2}{3} = \frac{4}{3} \qquad \Rightarrow \qquad f\left(\frac{4}{3}\right) = 3\left(\frac{4}{3}\right) - 2 = 4 - 2 = 2$$

In the first case we plugged $x = -1$ into $f(x)$ and got a value of -5. We then turned around and plugged $x = -5$ into $g(x)$ and got a value of -1, the number that we started off with.

In the second case we did something similar. Here we plugged $x = 2$ into $g(x)$ and got a value of $\frac{4}{3}$, we turned around and plugged this into $f(x)$ and got a value of 2, which is again the number that we started with.

Note that we really are doing some function composition here. The first case is really,

$$(g \circ f)(-1) = g[f(-1)] = g[-5] = -1$$

and the second case is really,

$$(f \circ g)(2) = f[g(2)] = f\left[\frac{4}{3}\right] = 2$$

Note as well that these both agree with the formula for the compositions that we found in the previous section. We get back out of the function evaluation the number that we originally plugged into the composition.

So, just what is going on here? In some way we can think of these two functions as undoing what the other did to a number. In the first case we plugged $x = -1$ into $f(x)$ and then plugged the result from this function evaluation back into $g(x)$ and in some way $g(x)$ undid what $f(x)$ had done to $x = -1$ and gave us back the original x that we started with.

Function pairs that exhibit this behavior are called **inverse functions**. Before formally defining inverse functions and the notation that we're going to use for them we need to get a definition out of the way.

A function is called **one-to-one** if no two values of x produce the same y. Mathematically this is the same as saying,

$$f(x_1) \neq f(x_2) \qquad \text{whenever} \qquad x_1 \neq x_2$$

So, a function is one-to-one if whenever we plug different values into the function we get different function values.

Sometimes it is easier to understand this definition if we see a function that isn't one-to-one. Let's take a look at a function that isn't one-to-one. The function $f(x) = x^2$ is not one-to-one because both $f(-2) = 4$ and $f(2) = 4$. In other words, there are two different values of x that produce the same value of y. Note that we can turn $f(x) = x^2$ into a one-to-one function if we restrict ourselves to $0 \le x < \infty$. This can sometimes be done with functions.

Showing that a function is one-to-one is often tedious and/or difficult. For the most part we are going to assume that the functions that we're going to be dealing with in this course are either one-to-one or we have restricted the domain of the function to get it to be a one-to-one function.

Now, let's formally define just what inverse functions are. Given two one-to-one functions $f(x)$ and $g(x)$ if

$$(f \circ g)(x) = x \qquad \text{AND} \qquad (g \circ f)(x) = x$$

then we say that $f(x)$ and $g(x)$ are **inverses** of each other. More specifically we will say that $g(x)$ is the **inverse** of $f(x)$ and denote it by

$$g(x) = f^{-1}(x)$$

Likewise, we could also say that $f(x)$ is the **inverse** of $g(x)$ and denote it by

$$f(x) = g^{-1}(x)$$

The notation that we use really depends upon the problem. In most cases either is acceptable.

For the two functions that we started off this section with we could write either of the following two sets of notation.

$$f(x) = 3x - 2 \qquad\qquad f^{-1}(x) = \frac{x}{3} + \frac{2}{3}$$

$$g(x) = \frac{x}{3} + \frac{2}{3} \qquad\qquad g^{-1}(x) = 3x - 2$$

Now, be careful with the notation for inverses. The "-1" is NOT an exponent despite the fact that it sure does look like one! When dealing with inverse functions we've got to remember that

$$f^{-1}(x) \neq \frac{1}{f(x)}$$

This is one of the more common mistakes that students make when first studying inverse functions.

The process for finding the inverse of a function is a fairly simple one although there are a couple of steps that can on occasion be somewhat messy. Here is the process

Finding the Inverse of a Function

Given the function $f(x)$ we want to find the inverse function, $f^{-1}(x)$.

1. First, replace $f(x)$ with y. This is done to make the rest of the process easier.
2. Replace every x with a y and replace every y with an x.
3. Solve the equation from Step 2 for y. This is the step where mistakes are most often made so be careful with this step.
4. Replace y with $f^{-1}(x)$. In other words, we've managed to find the inverse at this point!
5. Verify your work by checking that $\left(f \circ f^{-1}\right)(x) = x$ and $\left(f^{-1} \circ f\right)(x) = x$ are both true. This work can sometimes be messy making it easy to make mistakes so again be careful.

That's the process. Most of the steps are not all that bad but as mentioned in the process there are a couple of steps that we really need to be careful with since it is easy to make mistakes in those steps.

In the verification step we technically really do need to check that both $\left(f \circ f^{-1}\right)(x) = x$ and $\left(f^{-1} \circ f\right)(x) = x$ are true. For all the functions that we are going to be looking at in this course if one is true then the other will also be true. However, there are functions (they are beyond the scope of this course however) for which it is possible for only one of these to be true. This is brought up because in all the problems here we will be just checking one of them. We just need to always remember that technically we should check both.

Let's work some examples.

Example 1

Given $f(x) = 3x - 2$ find $f^{-1}(x)$.

Solution

Now, we already know what the inverse to this function is as we've already done some work with it. However, it would be nice to actually start with this since we know what we should get. This will work as a nice verification of the process.

So, let's get started. We'll first replace $f(x)$ with y.

$$y = 3x - 2$$

Next, replace all x's with y and all y's with x.

$$x = 3y - 2$$

Now, solve for y.

$$\begin{aligned} x + 2 &= 3y \\ \frac{1}{3}\left(x+2\right) &= y \\ \frac{x}{3} + \frac{2}{3} &= y \end{aligned}$$

Finally replace y with $f^{-1}\left(x\right)$.

$$f^{-1}\left(x\right) = \frac{x}{3} + \frac{2}{3}$$

Now, we need to verify the results. We already took care of this in the previous section, however, we really should follow the process so we'll do that here. It doesn't matter which of the two that we check we just need to check one of them. This time we'll check that $\left(f \circ f^{-1}\right)\left(x\right) = x$ is true.

$$\begin{aligned} \left(f \circ f^{-1}\right)\left(x\right) &= f\left[f^{-1}\left(x\right)\right] \\ &= f\left[\frac{x}{3} + \frac{2}{3}\right] \\ &= 3\left(\frac{x}{3} + \frac{2}{3}\right) - 2 \\ &= x + 2 - 2 \\ &= x \end{aligned}$$

Example 2

Given $g\left(x\right) = \sqrt{x-3}$ find $g^{-1}\left(x\right)$.

Solution

The fact that we're using $g\left(x\right)$ instead of $f\left(x\right)$ doesn't change how the process works. Here are the first few steps.

$$y = \sqrt{x-3} \qquad \Rightarrow \qquad x = \sqrt{y-3}$$

Now, to solve for y we will need to first square both sides and then proceed as normal.

$$\begin{aligned} x &= \sqrt{y-3} \\ x^2 &= y - 3 \\ x^2 + 3 &= y \end{aligned}$$

This inverse is then,

$$g^{-1}(x) = x^2 + 3$$

Finally let's verify and this time we'll use the other one just so we can say that we've gotten both down somewhere in an example.

$$\begin{aligned}\left(g^{-1} \circ g\right)(x) &= g^{-1}\left[g(x)\right] \\ &= g^{-1}\left(\sqrt{x-3}\right) \\ &= \left(\sqrt{x-3}\right)^2 + 3 \\ &= x - 3 + 3 \\ &= x\end{aligned}$$

So, we did the work correctly and we do indeed have the inverse.

The next example can be a little messy so be careful with the work here.

Example 3

Given $h(x) = \dfrac{x+4}{2x-5}$ find $h^{-1}(x)$.

Solution

The first couple of steps are pretty much the same as the previous examples so here they are,

$$y = \frac{x+4}{2x-5} \qquad \Rightarrow \qquad x = \frac{y+4}{2y-5}$$

Now, be careful with the solution step. With this kind of problem it is very easy to make a mistake here.

$$\begin{aligned}x(2y-5) &= y+4 \\ 2xy - 5x &= y + 4 \\ 2xy - y &= 4 + 5x \\ (2x-1)y &= 4 + 5x \\ y &= \frac{4+5x}{2x-1}\end{aligned}$$

So, if we've done all of our work correctly the inverse should be,

$$h^{-1}(x) = \frac{4+5x}{2x-1}$$

Finally, we'll need to do the verification. This is also a fairly messy process and it doesn't really matter which one we work with.

$$\begin{aligned}\left(h \circ h^{-1}\right)(x) &= h\left[h^{-1}(x)\right] \\ &= h\left[\frac{4+5x}{2x-1}\right] \\ &= \frac{\frac{4+5x}{2x-1}+4}{2\left(\frac{4+5x}{2x-1}\right)-5}\end{aligned}$$

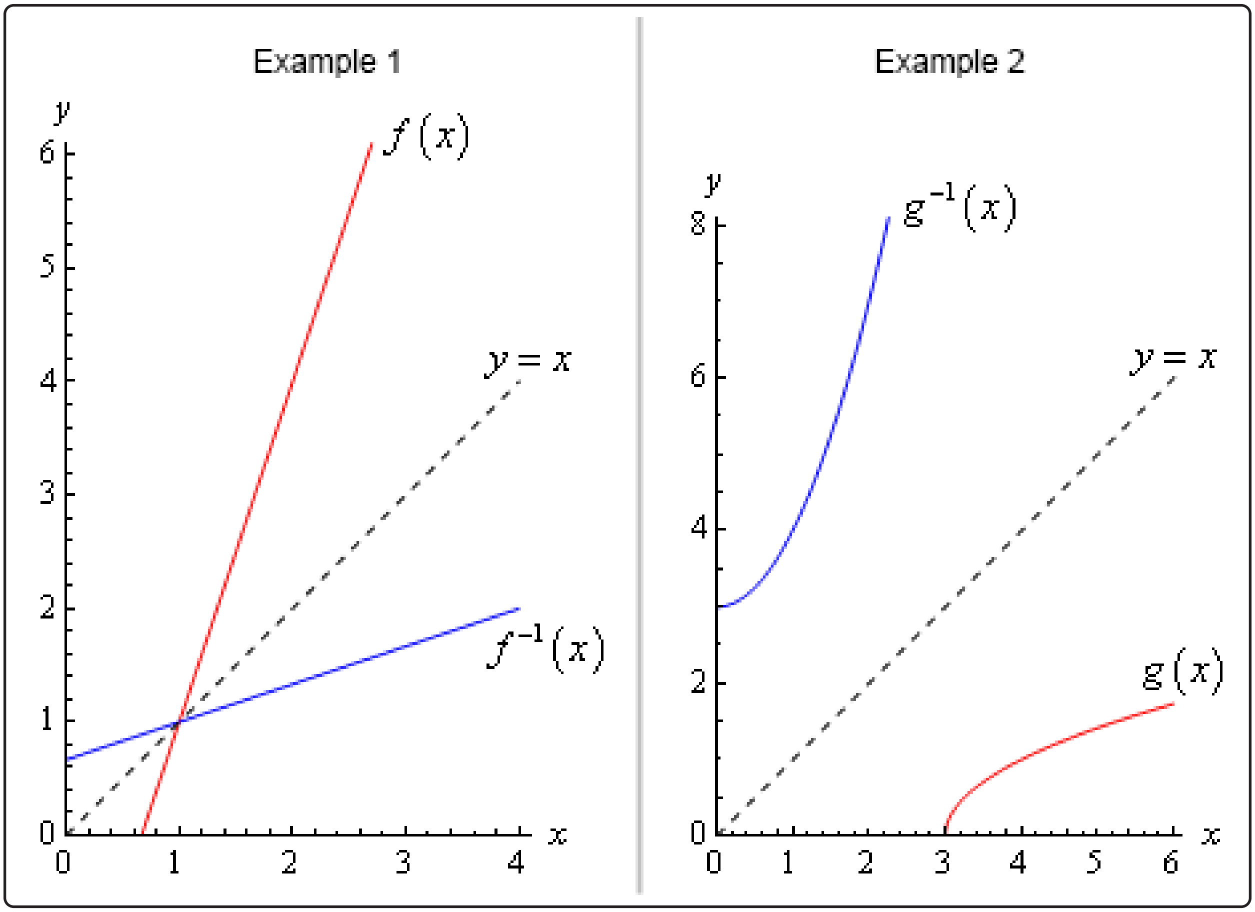
Okay, this is a mess. Let's simplify things up a little bit by multiplying the numerator and denominator by $2x-1$.

$$\begin{aligned}\left(h \circ h^{-1}\right)(x) &= \frac{2x-1}{2x-1}\,\frac{\frac{4+5x}{2x-1}+4}{2\left(\frac{4+5x}{2x-1}\right)-5} \\ &= \frac{(2x-1)\left(\frac{4+5x}{2x-1}+4\right)}{(2x-1)\left(2\left(\frac{4+5x}{2x-1}\right)-5\right)} \\ &= \frac{4+5x+4(2x-1)}{2(4+5x)-5(2x-1)} \\ &= \frac{4+5x+8x-4}{8+10x-10x+5} \\ &= \frac{13x}{13} = x\end{aligned}$$

Wow. That was a lot of work, but it all worked out in the end. We did all of our work correctly and we do in fact have the inverse.

There is one final topic that we need to address quickly before we leave this section. There is an interesting relationship between the graph of a function and the graph of its inverse.

Here is the graph of the function and inverse from the first two examples.



In both cases we can see that the graph of the inverse is a reflection of the actual function about the line $y = x$. This will always be the case with the graphs of a function and its inverse.

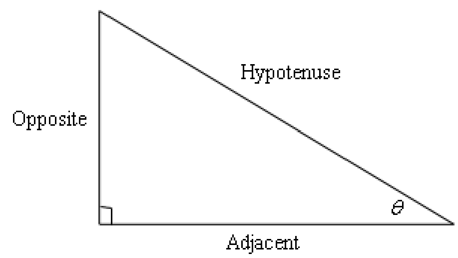
1.3 Trig Functions

The intent of this section is to remind you of some of the more important (from a Calculus standpoint...) topics from a trig class. One of the most important (but not the first) of these topics will be how to use the unit circle. We will leave the most important topic to the next section.

First let's start with the six trig functions and how they relate to each other.

$$\begin{aligned}
&\cos(x) && \sin(x)\\
&\tan(x) = \frac{\sin(x)}{\cos(x)} && \cot(x) = \frac{\cos(x)}{\sin(x)} = \frac{1}{\tan(x)}\\
&\sec(x) = \frac{1}{\cos(x)} && \csc(x) = \frac{1}{\sin(x)}
\end{aligned}$$

Recall as well that all the trig functions can be defined in terms of a right triangle.



From this right triangle we get the following definitions of the six trig functions.

$$\begin{aligned}
&\cos\theta = \frac{\text{adjacent}}{\text{hypotenuse}} && \sin\theta = \frac{\text{opposite}}{\text{hypotenuse}}\\
&\tan\theta = \frac{\text{opposite}}{\text{adjacent}} && \cot\theta = \frac{\text{adjacent}}{\text{opposite}}\\
&\sec\theta = \frac{\text{hypotenuse}}{\text{adjacent}} && \csc\theta = \frac{\text{hypotenuse}}{\text{opposite}}
\end{aligned}$$

Remembering both the relationship between all six of the trig functions and their right triangle definitions will be useful in this course on occasion.

Next, we need to touch on radians. In most trig classes instructors tend to concentrate on doing everything in terms of degrees (probably because it's easier to visualize degrees). The same is true in many science classes. However, in a calculus course almost everything is done in radians. The following table gives some of the basic angles in both degrees and radians.

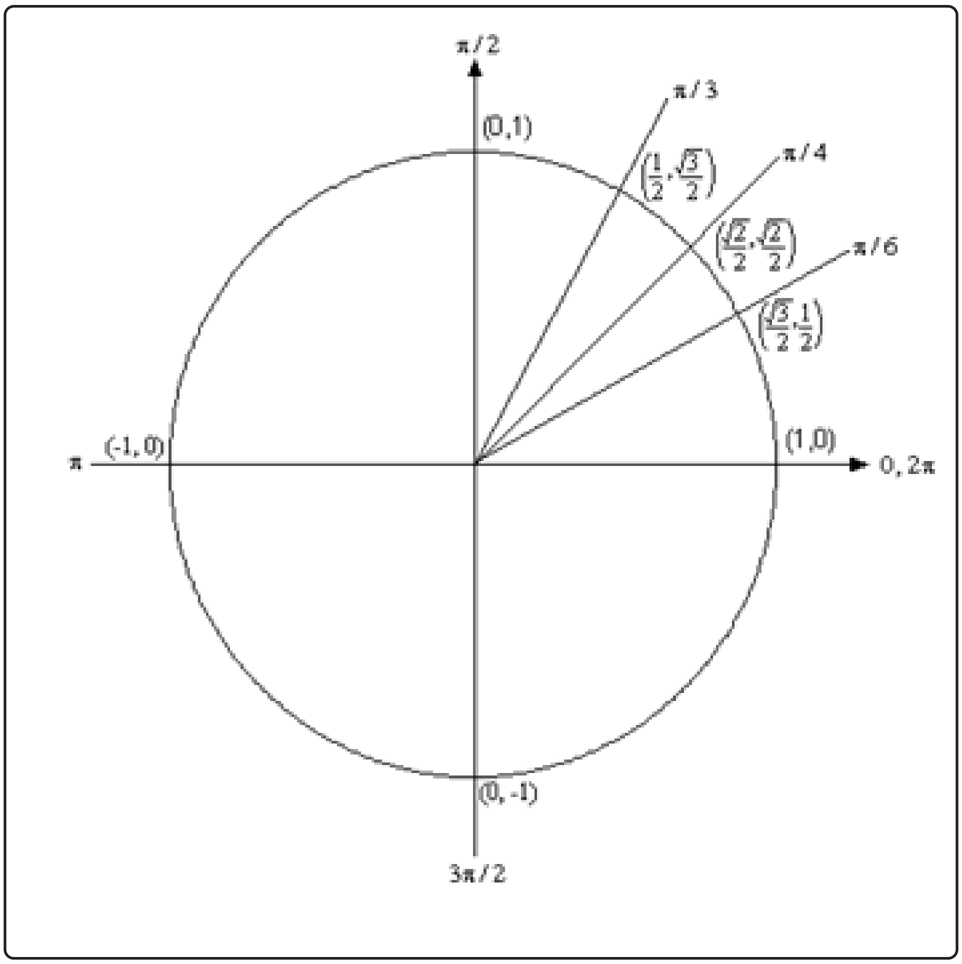
Degree	0	30	45	60	90	180	270	360
Radians	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π

Know this table! We may not see these specific angles all that much when we get into the Calculus portion of these notes, but knowing these can help us to visualize each angle. Now, one more time just make sure this is clear.

Be forewarned, everything in most calculus classes will be done in radians!

Let's next take a look at one of the most overlooked ideas from a trig class. The unit circle is one of the more useful tools to come out of a trig class. Unfortunately, most people don't learn it as well as they should in their trig class.

Below is unit circle with just the first quadrant filled in with the "standard" angles. The way the unit circle works is to draw a line from the center of the circle outwards corresponding to a given angle. Then look at the coordinates of the point where the line and the circle intersect. The first coordinate, *i.e.* the x-coordinate, is the cosine of that angle and the second coordinate, *i.e.* the y-coordinate, is the sine of that angle. We've put some of the angles along with the coordinates of their intersections on the unit circle.



So, from the unit circle above we can see that $\cos\left(\frac{\pi}{6}\right) = \frac{\sqrt{3}}{2}$ and $\sin\left(\frac{\pi}{6}\right) = \frac{1}{2}$.

Also, remember how the signs of angles work. If you rotate in a counter clockwise direction the angle is positive and if you rotate in a clockwise direction the angle is negative.

Recall as well that one complete revolution is 2π, so the positive x-axis can correspond to either an angle of 0 or 2π (or 4π, or 6π, or -2π, or -4π, *etc.* depending on the direction of rotation). Likewise, the angle $\frac{\pi}{6}$ (to pick an angle completely at random) can also be any of the following angles:

$\dfrac{\pi}{6} + 2\pi = \dfrac{13\pi}{6}$ (start at $\dfrac{\pi}{6}$ then rotate once around counter clockwise)

$\dfrac{\pi}{6} + 4\pi = \dfrac{25\pi}{6}$ (start at $\dfrac{\pi}{6}$ then rotate around twice counter clockwise)

$\dfrac{\pi}{6} - 2\pi = -\dfrac{11\pi}{6}$ (start at $\dfrac{\pi}{6}$ then rotate once around clockwise)

$\dfrac{\pi}{6} - 4\pi = -\dfrac{23\pi}{6}$ (start at $\dfrac{\pi}{6}$ then rotate around twice clockwise)

etc.

In fact, $\frac{\pi}{6}$ can be any of the following angles $\frac{\pi}{6} + 2\pi\, n\,, \quad n = 0, \pm 1, \pm 2, \pm 3, \ldots$ In this case n is the number of complete revolutions you make around the unit circle starting at $\frac{\pi}{6}$. Positive values of n correspond to counter clockwise rotations and negative values of n correspond to clockwise rotations.

So, why did we only put in the first quadrant? The answer is simple. If you know the first quadrant then you can get all the other quadrants from the first with a small application of geometry. You'll see how this is done in the following set of examples.

Example 1

Evaluate each of the following.

(a) $\sin\left(\frac{2\pi}{3}\right)$ and $\sin\left(-\frac{2\pi}{3}\right)$

(b) $\cos\left(\frac{7\pi}{6}\right)$ and $\cos\left(-\frac{7\pi}{6}\right)$

(c) $\tan\left(-\frac{\pi}{4}\right)$ and $\tan\left(\frac{7\pi}{4}\right)$

(d) $\sec\left(\frac{25\pi}{6}\right)$

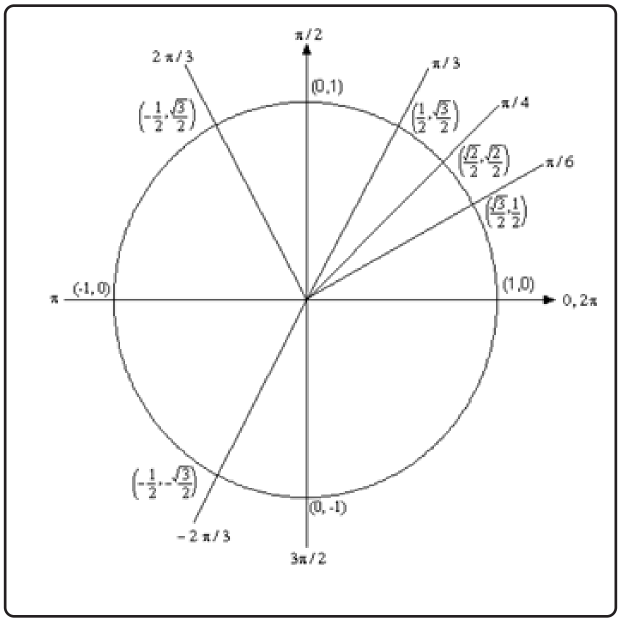
Solution

(a) $\sin\left(\frac{2\pi}{3}\right)$ and $\sin\left(-\frac{2\pi}{3}\right)$

The first evaluation in this part uses the angle $\frac{2\pi}{3}$. That's not on our unit circle above, however notice that $\frac{2\pi}{3} = \pi - \frac{\pi}{3}$. So $\frac{2\pi}{3}$ is found by rotating up $\frac{\pi}{3}$ from the negative x-axis. This means that the line for $\frac{2\pi}{3}$ will be a mirror image of the line for $\frac{\pi}{3}$ only in the second quadrant. The coordinates for $\frac{2\pi}{3}$ will be the coordinates for $\frac{\pi}{3}$ except the x coordinate will be negative.

Likewise, for $-\frac{2\pi}{3}$ we can notice that $-\frac{2\pi}{3} = -\pi + \frac{\pi}{3}$, so this angle can be found by rotating down $\frac{\pi}{3}$ from the negative x-axis. This means that the line for $-\frac{2\pi}{3}$ will be a mirror image of the line for $\frac{\pi}{3}$ only in the third quadrant and the coordinates will be the same as the coordinates for $\frac{\pi}{3}$ except both will be negative.

Both of these angles, along with the coordinates of the intersection points, are shown on the following unit circle.



From this unit circle we can see that $\sin\left(\frac{2\pi}{3}\right) = \frac{\sqrt{3}}{2}$ and $\sin\left(-\frac{2\pi}{3}\right) = -\frac{\sqrt{3}}{2}$.

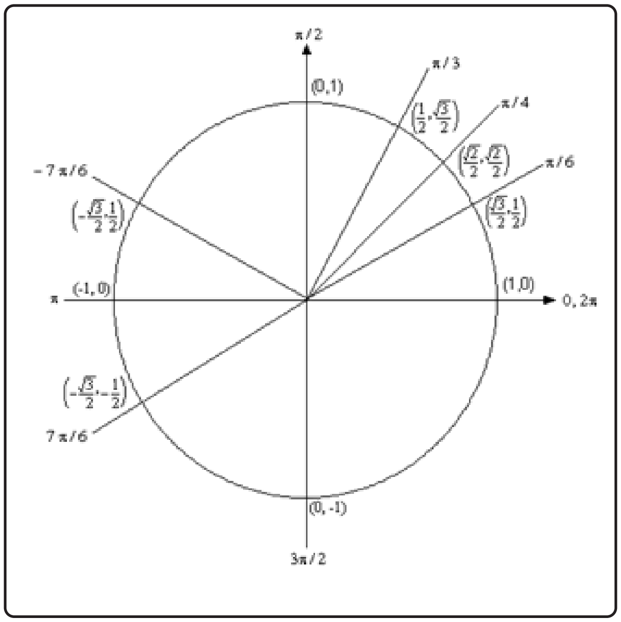
This leads to a nice fact about the sine function. The sine function is called an **odd** function and so for ANY angle we have

$$\sin(-\theta) = -\sin(\theta)$$

(b) $\cos\left(\frac{7\pi}{6}\right)$ and $\cos\left(-\frac{7\pi}{6}\right)$

For this example, notice that $\frac{7\pi}{6} = \pi + \frac{\pi}{6}$ so this means we would rotate down $\frac{\pi}{6}$ from the negative x-axis to get to this angle. Also $-\frac{7\pi}{6} = -\pi - \frac{\pi}{6}$ so this means we would rotate up $\frac{\pi}{6}$ from the negative x-axis to get to this angle. So, as with the last part, both of these angles will be mirror images of $\frac{\pi}{6}$ in the third and second quadrants respectively and we can use this to determine the coordinates for both of these new angles.

Both of these angles are shown on the following unit circle along with the coordinates for the intersection points.

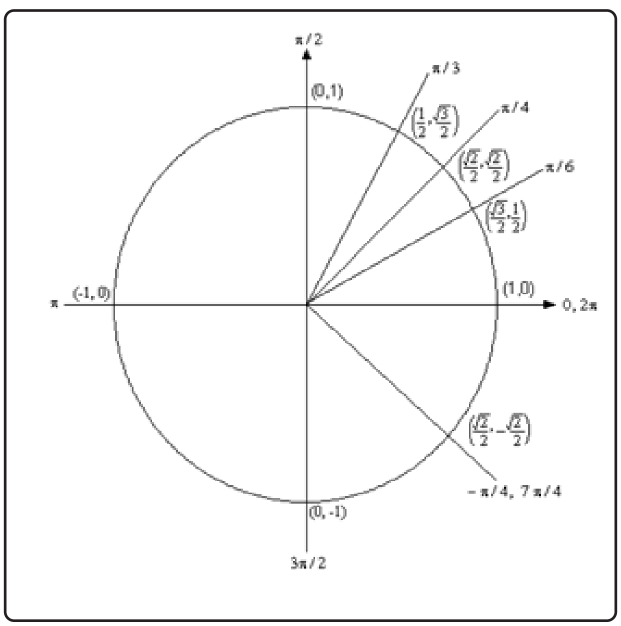


From this unit circle we can see that cos $\left(\frac{7\pi}{6}\right) = -\frac{\sqrt{3}}{2}$ and cos $\left(-\frac{7\pi}{6}\right) = -\frac{\sqrt{3}}{2}$. In this case the cosine function is called an **even** function and so for ANY angle we have

$$\cos(-\theta) = \cos(\theta)$$

(c) tan $\left(-\frac{\pi}{4}\right)$ and tan $\left(\frac{7\pi}{4}\right)$

Here we should note that $\frac{7\pi}{4} = 2\pi - \frac{\pi}{4}$ so $\frac{7\pi}{4}$ and $-\frac{\pi}{4}$ are in fact the same angle! Also note that this angle will be the mirror image of $\frac{\pi}{4}$ in the fourth quadrant. The unit circle for this angle is



Now, if we remember that $\tan(x) = \frac{\sin(x)}{\cos(x)}$ we can use the unit circle to find the values of the tangent function. So,

$$\tan\left(\frac{7\pi}{4}\right) = \tan\left(-\frac{\pi}{4}\right) = \frac{\sin(-\pi/4)}{\cos(-\pi/4)} = \frac{-\sqrt{2}/2}{\sqrt{2}/2} = -1$$

On a side note, notice that $\tan\left(\frac{\pi}{4}\right) = 1$ and we can see that the tangent function is also called an **odd** function and so for ANY angle we will have

$$\tan(-\theta) = -\tan(\theta)$$

(d) $\sec\left(\frac{25\pi}{6}\right)$

Here we need to notice that $\frac{25\pi}{6} = 4\pi + \frac{\pi}{6}$. In other words, we've started at $\frac{\pi}{6}$ and rotated around twice to end back up at the same point on the unit circle. This means that

$$\sec\left(\frac{25\pi}{6}\right) = \sec\left(4\pi + \frac{\pi}{6}\right) = \sec\left(\frac{\pi}{6}\right)$$

Now, let's also not get excited about the secant here. Just recall that

$$\sec(x) = \frac{1}{\cos(x)}$$

and so all we need to do here is evaluate a cosine! Therefore,

$$\sec\left(\frac{25\pi}{6}\right) = \sec\left(\frac{\pi}{6}\right) = \frac{1}{\cos\left(\frac{\pi}{6}\right)} = \frac{1}{\sqrt{3}/2} = \frac{2}{\sqrt{3}}$$

So, in the last example we saw how the unit circle can be used to determine the value of the trig functions at any of the "common" angles. It's important to notice that all of these examples used the fact that if you know the first quadrant of the unit circle and can relate all the other angles to "mirror images" of one of the first quadrant angles you don't really need to know whole unit circle. If you'd like to see a complete unit circle I've got one on my Trig Cheat Sheet that is available at https://tutorial.math.lamar.edu.

Another important idea from the last example is that when it comes to evaluating trig functions all that you really need to know is how to evaluate sine and cosine. The other four trig functions are defined in terms of these two so if you know how to evaluate sine and cosine you can also evaluate the remaining four trig functions.

We've not covered many of the topics from a trig class in this section, but we did cover some of the more important ones from a calculus standpoint. There are many important trig formulas that you will use occasionally in a calculus class. Most notably are the half-angle and double-angle formulas. If you need reminded of what these are, you might want to download my Trig Cheat Sheet as most of the important facts and formulas from a trig class are listed there.

1.4 Solving Trig Equations

In this section we will take a look at solving trig equations. This is something that you will be asked to do on a fairly regular basis in many classes.

Let's just jump into the examples and see how to solve trig equations.

Example 1

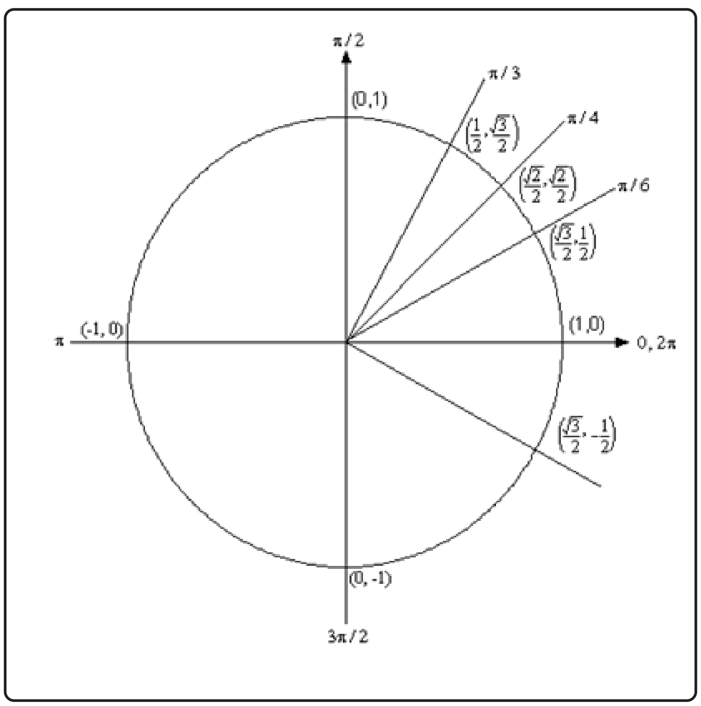
Solve $2\cos(t) = \sqrt{3}$.

Solution

There's really not a whole lot to do in solving this kind of trig equation. We first need to get the trig function on one side by itself. To do this all we need to do is divide both sides by 2.

$$\begin{aligned} 2\cos(t) &= \sqrt{3} \\ \cos(t) &= \frac{\sqrt{3}}{2} \end{aligned}$$

We are looking for all the values of t for which cosine will have the value of $\frac{\sqrt{3}}{2}$. So, let's take a look at the following unit circle.



From quick inspection we can see that $t = \frac{\pi}{6}$ is a solution. However, as we have shown on the unit circle there is another angle which will also be a solution. We need to determine what this angle is. When we look for these angles we typically want *positive* angles that lie between 0 and 2π. This angle will not be the only possibility of course, but we typically look for angles that meet these conditions.

To find this angle for this problem all we need to do is use a little geometry. The angle in the first quadrant makes an angle of $\frac{\pi}{6}$ with the positive x-axis, then so must the angle in the fourth quadrant. So, we have two options. We could use $-\frac{\pi}{6}$, but again, it's more common to use positive angles. To get a positive angle all we need to do is use the fact that the angle is $\frac{\pi}{6}$ with the positive x-axis (as noted above) and a positive angle will be $t = 2\pi - \frac{\pi}{6} = \frac{11\pi}{6}$.

One way to remember how to get the positive form of the second angle is to think of making one full revolution from the positive x-axis (*i.e.* 2π) and then backing off (*i.e.* subtracting) $\frac{\pi}{6}$.

We aren't done with this problem. As the discussion about finding the second angle has shown there are many ways to write any given angle on the unit circle. Sometimes it will be $-\frac{\pi}{6}$ that we want for the solution and sometimes we will want both (or neither) of the listed angles. Therefore, since there isn't anything in this problem (contrast this with the next problem) to tell us which is the correct solution we will need to list ALL possible solutions.

This is very easy to do. Recall from the previous section and you'll see there that we used

$$\frac{\pi}{6} + 2\pi\, n\,, \quad n = 0, \pm 1, \pm 2, \pm 3, \ldots$$

to represent all the possible angles that can end at the same location on the unit circle, *i.e.* angles that end at $\frac{\pi}{6}$. Remember that all this says is that we start at $\frac{\pi}{6}$ then rotate around in the counter-clockwise direction (n is positive) or clockwise direction (n is negative) for n complete rotations. The same thing can be done for the second solution.

So, all together the complete solution to this problem is

$$\begin{aligned} &\frac{\pi}{6} + 2\pi\, n\,, \quad n = 0, \pm 1, \pm 2, \pm 3, \ldots \\ &\frac{11\pi}{6} + 2\pi\, n\,, \quad n = 0, \pm 1, \pm 2, \pm 3, \ldots \end{aligned}$$

As a final thought, notice that we can get $-\frac{\pi}{6}$ by using $n = -1$ in the second solution.

Now, in a calculus class this is not a typical trig equation that we'll be asked to solve. A more typical example is the next one.

Example 2

Solve $2\cos(t) = \sqrt{3}$ on $[-2\pi, 2\pi]$.

Solution

In a calculus class we are often more interested in only the solutions to a trig equation that fall in a certain interval. The first step in this kind of problem is to find all possible solutions. We did this in the previous example.

$$\begin{aligned} &\frac{\pi}{6} + 2\pi\, n\,, \quad n = 0, \pm 1, \pm 2, \pm 3, \ldots \\ &\frac{11\pi}{6} + 2\pi\, n\,, \quad n = 0, \pm 1, \pm 2, \pm 3, \ldots \end{aligned}$$

Now, to find the solutions in the interval all we need to do is start picking values of n, plugging them in and getting the solutions that will fall into the interval that we've been given.

$n = 0$.

$$\begin{aligned} \frac{\pi}{6} + 2\pi\,(0) &= \frac{\pi}{6} < 2\pi \\ \frac{11\pi}{6} + 2\pi\,(0) &= \frac{11\pi}{6} < 2\pi \end{aligned}$$

Now, notice that if we take any positive value of n we will be adding on positive multiples of 2π onto a positive quantity and this will take us past the upper bound of our interval so we don't need to take any positive value of n.

However, just because we aren't going to take any positive value of n doesn't mean that we shouldn't also look at negative values of n.

$n = -1$.

$$\begin{aligned} \frac{\pi}{6} + 2\pi\,(-1) &= -\frac{11\pi}{6} > -2\pi \\ \frac{11\pi}{6} + 2\pi\,(-1) &= -\frac{\pi}{6} > -2\pi \end{aligned}$$

These are both greater than -2π and so are solutions, but if we subtract another 2π off (*i.e* use $n = -2$) we will once again be outside of the interval so we've found all the possible solutions that lie inside the interval $[-2\pi, 2\pi]$.

So, the solutions are : $\frac{\pi}{6}$, $\frac{11\pi}{6}$, $-\frac{\pi}{6}$, $-\frac{11\pi}{6}$.

So, let's see if you've got all this down.

Example 3

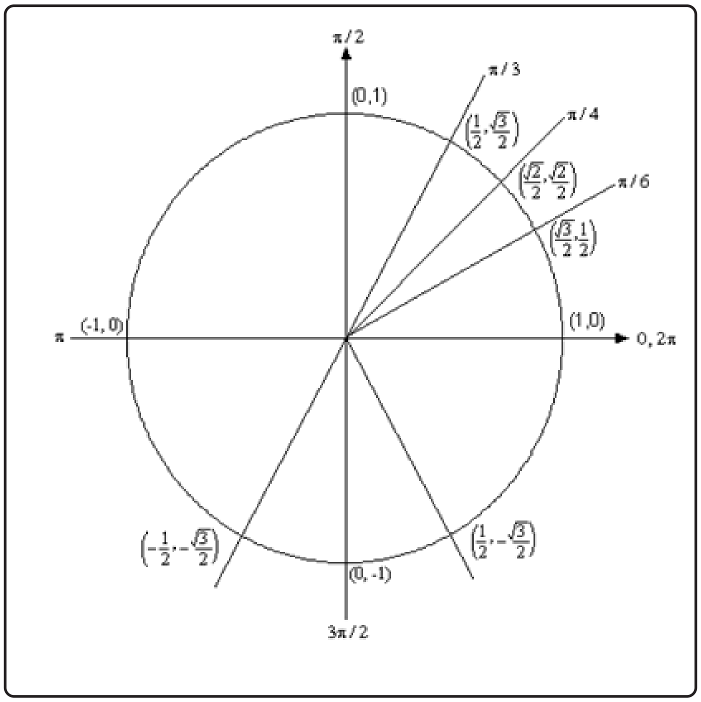
Solve $2\sin(5x) = -\sqrt{3}$ on $[-\pi, 2\pi]$.

Solution

This problem is very similar to the other problems in this section with a very important difference. We'll start this problem in exactly the same way as we did in the first example. So, first get the sine on one side by itself.

$$\begin{aligned} 2\sin(5x) &= -\sqrt{3} \\ \sin(5x) &= \frac{-\sqrt{3}}{2} \end{aligned}$$

We are looking for angles that will give $-\frac{\sqrt{3}}{2}$ out of the sine function. Let's again go to our trusty unit circle.



Now, there are no angles in the first quadrant for which sine has a value of $-\frac{\sqrt{3}}{2}$. However, there are two angles in the lower half of the unit circle for which sine will have a value of $-\frac{\sqrt{3}}{2}$. So, what are these angles?

Notice that $\sin\left(\frac{\pi}{3}\right) = \frac{\sqrt{3}}{2}$. Given this we now know that the angle in the third quadrant will be $\frac{\pi}{3}$ below the ***negative*** x-axis or $\pi + \frac{\pi}{3} = \frac{4\pi}{3}$. An easy way to remember this is to notice that we'll rotate half a revolution from the positive x-axis to get to the negative x-axis then add on $\frac{\pi}{3}$ to reach the angle we are looking for.

Likewise, the angle in the fourth quadrant will $\frac{\pi}{3}$ below the ***positive*** x-axis. So, we could use $-\frac{\pi}{3}$ or $2\pi - \frac{\pi}{3} = \frac{5\pi}{3}$. Remember that we're typically looking for positive angles between 0 and 2π so we'll use the positive angle. An easy way to remember how to the positive angle here is to rotate one full revolution from the positive x-axis (*i.e.* 2π) and then backing off (*i.e.* subtracting) $\frac{\pi}{3}$.

Now we come to the very important difference between this problem and the previous problems in this section. The solution is **NOT**

$$x = \frac{4\pi}{3} + 2\pi n, \quad n = 0, \pm 1, \pm 2, \ldots$$
$$x = \frac{5\pi}{3} + 2\pi n, \quad n = 0, \pm 1, \pm 2, \ldots$$

This is not the set of solutions because we are NOT looking for values of x for which $\sin(x) = -\frac{\sqrt{3}}{2}$, but instead we are looking for values of x for which $\sin(5x) = -\frac{\sqrt{3}}{2}$. Note the difference in the arguments of the sine function! One is x and the other is $5x$. This makes all the difference in the world in finding the solution! Therefore, the set of solutions is

$$5x = \frac{4\pi}{3} + 2\pi n, \quad n = 0, \pm 1, \pm 2, \ldots$$
$$5x = \frac{5\pi}{3} + 2\pi n, \quad n = 0, \pm 1, \pm 2, \ldots$$

Well, actually, that's not quite the solution. We are looking for values of x so divide everything by 5 to get.

$$x = \frac{4\pi}{15} + \frac{2\pi n}{5}, \quad n = 0, \pm 1, \pm 2, \ldots$$
$$x = \frac{\pi}{3} + \frac{2\pi n}{5}, \quad n = 0, \pm 1, \pm 2, \ldots$$

Notice that we also divided the $2\pi n$ by 5 as well! This is important! If we don't do that you **WILL** miss solutions. For instance, take $n = 1$.

$$x = \frac{4\pi}{15} + \frac{2\pi}{5} = \frac{10\pi}{15} = \frac{2\pi}{3} \qquad \Rightarrow \qquad \sin\left(5\left(\frac{2\pi}{3}\right)\right) = \sin\left(\frac{10\pi}{3}\right) = -\frac{\sqrt{3}}{2}$$
$$x = \frac{\pi}{3} + \frac{2\pi}{5} = \frac{11\pi}{15} \qquad \Rightarrow \qquad \sin\left(5\left(\frac{11\pi}{15}\right)\right) = \sin\left(\frac{11\pi}{3}\right) = -\frac{\sqrt{3}}{2}$$

We'll leave it to you to verify our work showing they are solutions. However, it makes the point. If you didn't divide the $2\pi n$ by 5 you would have missed these solutions!

Okay, now that we've gotten all possible solutions it's time to find the solutions on the given interval. We'll do this as we did in the previous problem. Pick values of n and get the solutions.

$n = 0$.

$$x = \frac{4\pi}{15} + \frac{2\pi(0)}{5} = \frac{4\pi}{15} < 2\pi$$
$$x = \frac{\pi}{3} + \frac{2\pi(0)}{5} = \frac{\pi}{3} < 2\pi$$

$n = 1$.

$$x = \frac{4\pi}{15} + \frac{2\pi(1)}{5} = \frac{2\pi}{3} < 2\pi$$
$$x = \frac{\pi}{3} + \frac{2\pi(1)}{5} = \frac{11\pi}{15} < 2\pi$$

$n = 2$.

$$x = \frac{4\pi}{15} + \frac{2\pi(2)}{5} = \frac{16\pi}{15} < 2\pi$$
$$x = \frac{\pi}{3} + \frac{2\pi(2)}{5} = \frac{17\pi}{15} < 2\pi$$

$n = 3$.

$$x = \frac{4\pi}{15} + \frac{2\pi(3)}{5} = \frac{22\pi}{15} < 2\pi$$
$$x = \frac{\pi}{3} + \frac{2\pi(3)}{5} = \frac{23\pi}{15} < 2\pi$$

$n = 4$.

$$x = \frac{4\pi}{15} + \frac{2\pi(4)}{5} = \frac{28\pi}{15} < 2\pi$$
$$x = \frac{\pi}{3} + \frac{2\pi(4)}{5} = \frac{29\pi}{15} < 2\pi$$

$n = 5$.

$$x = \frac{4\pi}{15} + \frac{2\pi(5)}{5} = \frac{34\pi}{15} > 2\pi$$
$$x = \frac{\pi}{3} + \frac{2\pi(5)}{5} = \frac{35\pi}{15} > 2\pi$$

Okay, so we finally got past the right endpoint of our interval so we don't need any more positive n. Now let's take a look at the negative n and see what we've got.

$n = -1$.

$$x = \frac{4\pi}{15} + \frac{2\pi(-1)}{5} = -\frac{2\pi}{15} > -\pi$$
$$x = \frac{\pi}{3} + \frac{2\pi(-1)}{5} = -\frac{\pi}{15} > -\pi$$

$n = -2$.

$$x = \frac{4\pi}{15} + \frac{2\pi(-2)}{5} = -\frac{8\pi}{15} > -\pi$$
$$x = \frac{\pi}{3} + \frac{2\pi(-2)}{5} = -\frac{7\pi}{15} > -\pi$$

$n = -3$.

$$x = \frac{4\pi}{15} + \frac{2\pi(-3)}{5} = -\frac{14\pi}{15} > -\pi$$
$$x = \frac{\pi}{3} + \frac{2\pi(-3)}{5} = -\frac{13\pi}{15} > -\pi$$

$n = -4$.

$$x = \frac{4\pi}{15} + \frac{2\pi(-4)}{5} = -\frac{4\pi}{3} < -\pi$$
$$x = \frac{\pi}{3} + \frac{2\pi(-4)}{5} = -\frac{19\pi}{15} < -\pi$$

And we're now past the left endpoint of the interval. Sometimes, there will be many solutions as there were in this example. Putting all of this together gives the following set of solutions that lie in the given interval.

$$\frac{4\pi}{15}, \frac{\pi}{3}, \frac{2\pi}{3}, \frac{11\pi}{15}, \frac{16\pi}{15}, \frac{17\pi}{15}, \frac{22\pi}{15}, \frac{23\pi}{15}, \frac{28\pi}{15}, \frac{29\pi}{15}$$
$$-\frac{\pi}{15}, -\frac{2\pi}{15}, -\frac{7\pi}{15}, -\frac{8\pi}{15}, -\frac{13\pi}{15}, -\frac{14\pi}{15}$$

Let's work another example.

Example 4

Solve $\sin(2x) = -\cos(2x)$ on $\left[-\frac{3\pi}{2}, \frac{3\pi}{2}\right]$.

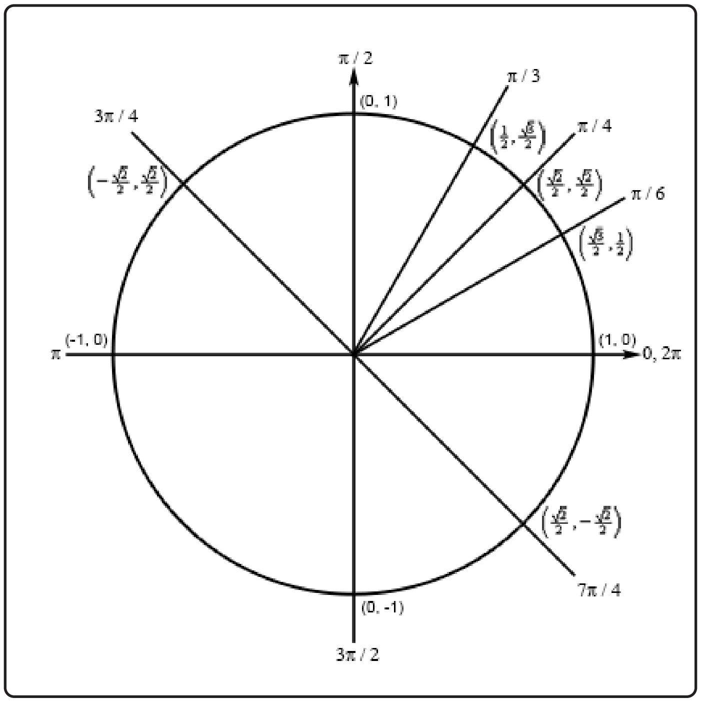
Solution

This problem is a little different from the previous ones. First, we need to do some rearranging and simplification.

$$\sin(2x) = -\cos(2x)$$
$$\frac{\sin(2x)}{\cos(2x)} = -1$$
$$\tan(2x) = -1$$

So, solving $\sin(2x) = -\cos(2x)$ is the same as solving $\tan(2x) = -1$. Hopefully, you'll recall that the smallest positive angle where tangent it -1 is $\frac{3\pi}{4}$ and this angle is in the 2^{nd} quadrant.

There is also a second angle for which tangent will be -1 and we can use the unit circle to illustrate this second angle. Let's take a look at the following unit circle.



As shown in this unit circle if we add π to our first angle we get $\frac{3\pi}{4} + \pi = \frac{7\pi}{4}$ and we get an angle that is in the fourth quadrant and has the same coordinates except for opposite signs. This means that tangent will also have a value of -1 here and so is a second angle.

This will always be true when solving tangent equations. Once we have one angle that will solve the equation a second angle will always be π plus the first angle.

All possible angles are then,

$$\begin{aligned} 2x &= \frac{3\pi}{4} + 2\pi n, \quad n = 0, \pm 1, \pm 2, \ldots \\ 2x &= \frac{7\pi}{4} + 2\pi n, \quad n = 0, \pm 1, \pm 2, \ldots \end{aligned}$$

Or, upon dividing by the 2 we get all possible solutions.

$$\begin{aligned} x &= \frac{3\pi}{8} + \pi n, \quad n = 0, \pm 1, \pm 2, \ldots \\ x &= \frac{7\pi}{8} + \pi n, \quad n = 0, \pm 1, \pm 2, \ldots \end{aligned}$$

Now, let's determine the solutions that lie in the given interval.

$n = 0$.

$$\begin{aligned} x &= \frac{3\pi}{8} + \pi(0) = \frac{3\pi}{8} < \frac{3\pi}{2} \\ x &= \frac{7\pi}{8} + \pi(0) = \frac{7\pi}{8} < \frac{3\pi}{2} \end{aligned}$$

$n = 1$.

$$x = \frac{3\pi}{8} + \pi(1) = \frac{11\pi}{8} < \frac{3\pi}{2}$$
$$x = \frac{7\pi}{8} + \pi(1) = \frac{15\pi}{8} > \frac{3\pi}{2}$$

Unlike the previous example only one of these will be in the interval. This will happen occasionally so don't always expect both answers from a particular n to work. Also, we should now check $n = 2$ for the first to see if it will be in or out of the interval. I'll leave it to you to check that it's out of the interval.

Now, let's check the negative n.

$n = -1$.

$$x = \frac{3\pi}{8} + \pi(-1) = -\frac{5\pi}{8} > -\frac{3\pi}{2}$$
$$x = \frac{7\pi}{8} + \pi(-1) = -\frac{\pi}{8} > -\frac{3\pi}{2}$$

$n = -2$.

$$x = \frac{3\pi}{8} + \pi(-2) = -\frac{13\pi}{8} < -\frac{3\pi}{2}$$
$$x = \frac{7\pi}{8} + \pi(-2) = -\frac{9\pi}{8} > -\frac{3\pi}{2}$$

Again, only one will work here. I'll leave it to you to verify that $n = -3$ will give two answers that are both out of the interval.

The complete list of solutions is then,

$$-\frac{9\pi}{8}, -\frac{5\pi}{8}, -\frac{\pi}{8}, \frac{3\pi}{8}, \frac{7\pi}{8}, \frac{11\pi}{8}$$

Before moving on we need to address one issue about the previous example. The solution method used there is not the "standard" solution method. Because the second angle is just π plus the first and if we added π onto the second angle we'd be back at the line representing the first angle the more standard solution method is to just add πn onto the first angle to get,

$$2x = \frac{3\pi}{4} + \pi n, \quad n = 0, \pm 1, \pm 2, \ldots$$

Then dividing by 2 to get the full set of solutions,

$$x = \frac{3\pi}{8} + \frac{\pi n}{2}, \quad n = 0, \pm 1, \pm 2, \ldots$$

This set of solutions is identical to the set of solutions we got in the example (we'll leave it to you to plug in some n's and verify that). So, why did we not use the method in the previous example? Simple. The method in the previous example more closely mirrors the solution method for cosine and sine (*i.e.* they both, generally, give two sets of angles) and so for students that aren't comfortable with solving trig equations this gives a "consistent" solution method.

Let's work one more example so that we can make a point that needs to be understood when solving some trig equations.

Example 5

Solve $\cos(3x) = 2$.

Solution

This example is designed to remind you of certain properties about sine and cosine. Recall that $-1 \le \cos(\theta) \le 1$ and $-1 \le \sin(\theta) \le 1$. Therefore, since cosine will never be greater that 1 it definitely can't be 2. So **THERE ARE NO SOLUTIONS** to this equation!

It is important to remember that not all trig equations will have solutions.

In this section we solved some simple trig equations. There are more complicated trig equations that we can solve so don't leave this section with the feeling that there is nothing harder out there in the world to solve. In fact, we'll see at least one of the more complicated problems in the next section. Also, every one of these problems came down to solutions involving one of the "common" or "standard" angles. Most trig equations won't come down to one of those and will in fact need a calculator to solve. The next section is devoted to this kind of problem.

1.5 Solving Trig Equations with Calculators, Part I

In the previous section we started solving trig equations. The only problem with the equations we solved in there is that they pretty much all had solutions that came from a handful of "standard" angles and of course there are many equations out there that simply don't. So, in this section we are going to take a look at some more trig equations, the majority of which will require the use of a calculator to solve (a couple won't need a calculator).

The fact that we are using calculators in this section does not however mean that the problems in the previous section aren't important. It is going to be assumed in this section that the basic ideas of solving trig equations are known and that we don't need to go back over them here. In particular, it is assumed that you can use a unit circle to help you find all answers to the equation (although the process here is a little different as we'll see) and it is assumed that you can find answers in a given interval. If you are unfamiliar with these ideas you should first go to the previous section and go over those problems.

Before proceeding with the problems we need to go over how our calculators work so that we can get the correct answers. Calculators are great tools but if you don't know how they work and how to interpret their answers you can get in serious trouble.

First, as already pointed out in previous sections, everything we are going to be doing here will be in radians so make sure that your calculator is set to radians before attempting the problems in this section. Also, we are going to use 4 decimal places of accuracy in the work here. You can use more if you want, but in this class, we'll always use at least 4 decimal places of accuracy.

Next, and somewhat more importantly, we need to understand how calculators give answers to inverse trig functions. We didn't cover inverse trig functions in this review, but they are just inverse functions and we have talked a little bit about inverse functions in a review section. The only real difference is that we are now using trig functions. We'll only be looking at three of them and they are

$$\begin{array}{lcl} \text{Inverse Cosine} & : & \cos^{-1}(x) = \arccos(x) \\ \text{Inverse Sine} & : & \sin^{-1}(x) = \arcsin(x) \\ \text{Inverse Tangent} & : & \tan^{-1}(x) = \arctan(x) \end{array}$$

As shown there are two different notations that are commonly used. In these notes we'll be using the first form since it is a little more compact. Most calculators these days will have buttons on them for these three so make sure that yours does as well.

We now need to deal with how calculators give answers to these. Let's suppose, for example, that we wanted our calculator to compute $\cos^{-1}\left(\frac{3}{4}\right)$. First, remember that what the calculator is computing is the angle, let's say x, that we would plug into cosine to get a value of $\frac{3}{4}$, or

$$x = \cos^{-1}\left(\frac{3}{4}\right) \qquad \Rightarrow \qquad \cos(x) = \frac{3}{4}$$

So, in other words, when we are using our calculator to compute an inverse trig function we are really solving a simple trig equation.

Having our calculator compute $\cos^{-1}\left(\frac{3}{4}\right)$ and hence solve $\cos(x) = \frac{3}{4}$ gives,

$$x = \cos^{-1}\left(\frac{3}{4}\right) = 0.7227$$

From the previous section we know that there should in fact be an infinite number of answers to this including a second angle that is in the interval $[0, 2\pi]$. However, our calculator only gave us a single answer. How to determine what the other angles are will be covered in the following examples so we won't go into detail here about that. We did need to point out however, that the calculators will only give a single answer and that we're going to have more work to do than just plugging a number into a calculator.

Since we know that there are supposed to be an infinite number of solutions to $\cos(x) = \frac{3}{4}$ the next question we should ask then is just how did the calculator decide to return the answer that it did? Why this one and not one of the others? Will it give the same answer every time?

There are rules that determine just what answer the calculator gives when computing inverse trig functions. All calculators will give answers in the following ranges.

$$0 \le \cos^{-1}(x) \le \pi \qquad -\frac{\pi}{2} \le \sin^{-1}(x) \le \frac{\pi}{2} \qquad -\frac{\pi}{2} < \tan^{-1}(x) < \frac{\pi}{2}$$

If you think back to the unit circle and recall that we think of cosine as the horizontal axis then we can see that we'll cover all possible values of cosine in the upper half of the circle and this is exactly the range given above for the inverse cosine. Likewise, since we think of sine as the vertical axis in the unit circle we can see that we'll cover all possible values of sine in the right half of the unit circle and that is the range given above.

For the tangent range look back to the graph of the tangent function itself and we'll see that one branch of the tangent is covered in the range given above and so that is the range we'll use for inverse tangent. Note as well that we don't include the endpoints in the range for inverse tangent since tangent does not exist there.

So, if we can remember these rules we will be able to determine the remaining angle in $[0, 2\pi]$ that also works for each solution.

As a final quick topic let's note that it will, on occasion, be useful to remember the decimal representations of some basic angles. So here they are,

$$\frac{\pi}{2} = 1.5708 \qquad \pi = 3.1416 \qquad \frac{3\pi}{2} = 4.7124 \qquad 2\pi = 6.2832$$

Using these we can quickly see that $\cos^{-1}\left(\frac{3}{4}\right)$ must be in the first quadrant since 0.7227 is between 0 and 1.5708. This will be of great help when we go to determine the remaining angles

So, once again, we can't stress enough that calculators are great tools that can be of tremendous help to us, but if you don't understand how they work you will often get the answers to problems wrong.

So, with all that out of the way let's take a look at our first problem.

Example 1

Solve $4\cos(t) = 3$ on $[-8, 10]$.

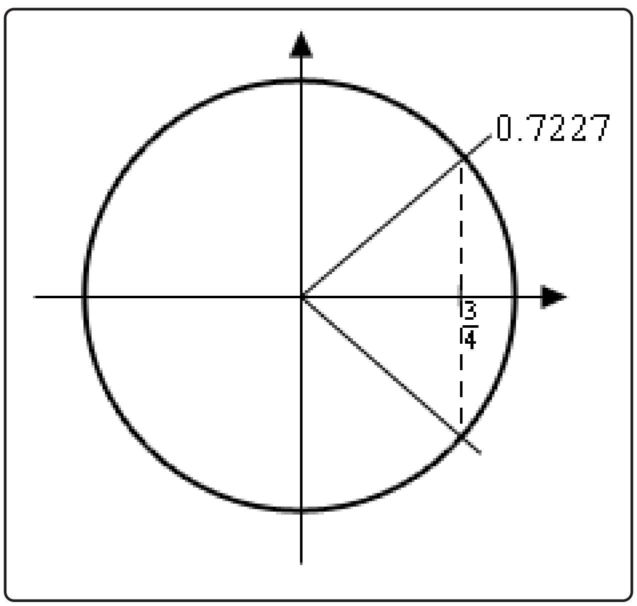
Solution

Okay, the first step here is identical to the problems in the previous section. We first need to isolate the cosine on one side by itself and then use our calculator to get the first answer.

$$\cos(t) = \frac{3}{4} \qquad \Rightarrow \qquad t = \cos^{-1}\left(\frac{3}{4}\right) = 0.7227$$

So, this is the one we were using above in the opening discussion of this section. At the time we mentioned that there were infinite number of answers and that we'd be seeing how to find them later. Well that time is now.

First, let's take a quick look at a unit circle for this example.



The angle that we've found is shown on the circle as well as the other angle that we know should also be an answer. Finding this angle here is just as easy as in the previous section. Since the line segment in the first quadrant forms an angle of 0.7227 radians with the positive x-axis then so does the line segment in the fourth quadrant. This means that we can use either -0.7227 as the second angle or $2\pi - 0.7227 = 5.5605$. Which you use depends on which you prefer. We'll pretty much always use the positive angle to avoid the possibility that we'll lose the minus sign.

So, all possible solutions, ignoring the interval for a second, are then,

$$\begin{aligned} t &= 0.7227 + 2\pi n \\ t &= 5.5605 + 2\pi n \end{aligned} \qquad n = 0, \pm 1, \pm 2, \ldots$$

Now, all we need to do is plug in values of n to determine the angle that are actually in the interval. Here's the work for that.

$$\begin{array}{llll} n = -2 & : & t = \cancel{-11.8437} & \text{and} \quad -7.0059 \\ n = -1 & : & t = -5.5605 & \text{and} \quad -0.7227 \\ n = 0 & : & t = 0.7227 & \text{and} \quad 5.5605 \\ n = 1 & : & t = 7.0059 & \text{and} \quad \cancel{11.8437} \end{array}$$

So, the solutions to this equation, in the given interval, are,

$$t = -7.0059, \ -5.5605, \ -0.7227, \ 0.7227, \ 5.5605, \ 7.0059$$

Note that we had a choice of angles to use for the second angle in the previous example. The choice of angles there will also affect the value(s) of n that we'll need to use to get all the solutions. In the end, regardless of the angle chosen, we'll get the same list of solutions, but the value(s) of n that give the solutions will be different depending on our choice.

Also, in the above example we put in a little more explanation than we'll show in the remaining examples in this section to remind you how these work.

Example 2

Solve $-10\cos(3t) = 7$ on $[-2, 5]$.

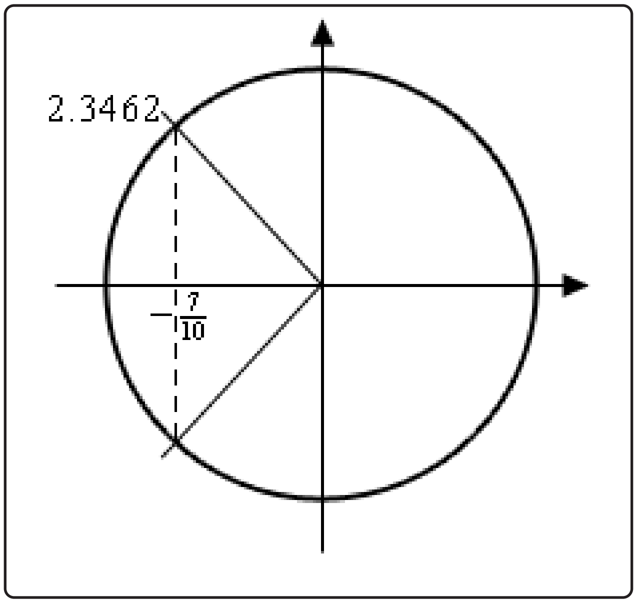
Solution

Okay, let's first get the inverse cosine portion of this problem taken care of.

$$\cos(3t) = -\frac{7}{10} \qquad \Rightarrow \qquad 3t = \cos^{-1}\left(-\frac{7}{10}\right) = 2.3462$$

Don't forget that we still need the "3"!

Now, let's look at a quick unit circle for this problem. As we can see the angle 2.3462 radians is in the second quadrant and the other angle that we need is in the third quadrant. We can find this second angle in exactly the same way we did in the previous example. We can use either -2.3462 or we can use $2\pi - 2.3462 = 3.9370$. As with the previous example we'll use the positive choice, but that is purely a matter of preference. You could use the negative if you wanted to.



So, let's now finish out the problem. First, let's acknowledge that the values of $3t$ that we need are,

$$\begin{aligned} 3t &= 2.3462 + 2\pi n \\ 3t &= 3.9370 + 2\pi n \end{aligned} \qquad n = 0, \pm 1, \pm 2, \ldots$$

Now, we need to properly deal with the 3, so divide that out to get all the solutions to the trig equation.

$$\begin{aligned} t &= 0.7821 + \frac{2\pi n}{3} \\ t &= 1.3123 + \frac{2\pi n}{3} \end{aligned} \qquad n = 0, \pm 1, \pm 2, \ldots$$

Finally, we need to get the values in the given interval.

$$\begin{array}{lcll} n = -2 & : & t = \cancel{-3.4067} & \text{and} \quad \cancel{-2.8765} \\ n = -1 & : & t = -1.3123 & \text{and} \quad -0.7821 \\ n = 0 & : & t = 0.7821 & \text{and} \quad 1.3123 \\ n = 1 & : & t = 2.8765 & \text{and} \quad 3.4067 \\ n = 2 & : & t = 4.9709 & \text{and} \quad \cancel{5.5011} \end{array}$$

The solutions to this equation, in the given interval are then,

$$t = -1.3123, \ -0.7821, \ 0.7821, \ 1.3123, \ 2.8765, \ 3.4067, \ 4.9709$$

We've done a couple of basic problems with cosines, now let's take a look at how solving equations with sines work.

Example 3

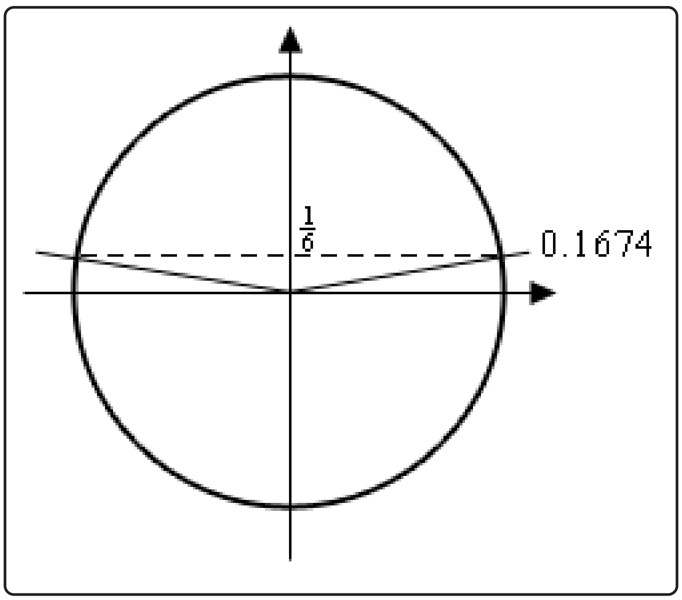
Solve $6\sin\left(\frac{x}{2}\right) = 1$ on $[-20, 30]$

Solution

Let's first get the calculator work out of the way since that isn't where the difference comes into play.

$$\sin\left(\frac{x}{2}\right) = \frac{1}{6} \qquad \Rightarrow \qquad \frac{x}{2} = \sin^{-1}\left(\frac{1}{6}\right) = 0.1674$$

Here's a unit circle for this example.



To find the second angle in this case we can notice that the line in the first quadrant makes an angle of 0.1674 with the positive x-axis and so the angle in the second quadrant will then make an angle of 0.1674 with the negative x-axis. So, if we start at the positive x-axis we rotate a half revolution and then back off 0.1674. Therefore, the angle that we're after is then, $\pi - 0.1674 = 2.9742$.

Here's the rest of the solution for this example. We're going to assume from this point on that you can do this work without much explanation.

$$\begin{aligned} \frac{x}{2} &= 0.1674 + 2\pi n \\ \frac{x}{2} &= 2.9742 + 2\pi n \end{aligned} \qquad \Rightarrow \qquad \begin{aligned} x &= 0.3348 + 4\pi n \\ x &= 5.9484 + 4\pi n \end{aligned} \qquad n = 0, \pm 1, \pm 2, \ldots$$

$$\begin{array}{lcll} n = -2 & : & x = \cancel{-24.7980} & \text{and} \quad -19.1844 \\ n = -1 & : & x = -12.2316 & \text{and} \quad -6.6180 \\ n = 0 & : & x = 0.3348 & \text{and} \quad 5.9484 \\ n = 1 & : & x = 12.9012 & \text{and} \quad 18.5148 \\ n = 2 & : & x = 25.4676 & \text{and} \quad \cancel{31.0812} \end{array}$$

The solutions to this equation are then,

$$x = -19.1844, \ -12.2316, \ -6.6180, \ 0.3348, \ 5.9484, \ 12.9012, \ 18.5148, \ 25.4676$$

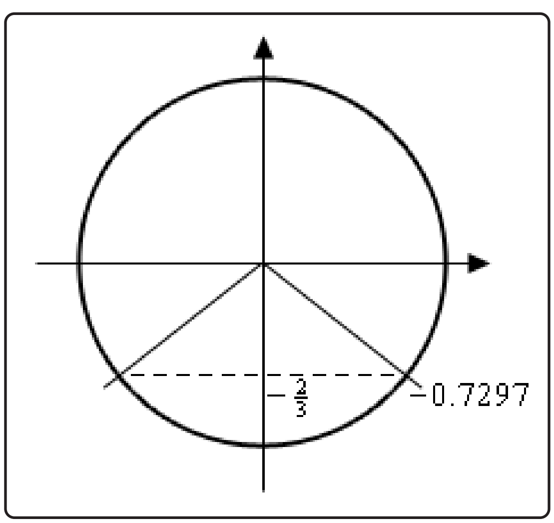
Example 4

Solve $3\sin(5z) = -2$ on $[0, 1]$.

Solution

You should be getting pretty good at these by now, so we won't be putting much explanation in for this one. Here we go.

$$\sin(5z) = -\frac{2}{3} \qquad \Rightarrow \qquad 5z = \sin^{-1}\left(-\frac{2}{3}\right) = -0.7297$$



Okay, with this one we're going to do a little more work than with the others. For the first angle we could use the answer our calculator gave us. However, it's easy to lose minus signs so we'll instead use $2\pi - 0.7297 = 5.5535$. Again, there is no reason to this other than a worry about losing the minus sign in the calculator answer. If you'd like to use the calculator answer you are more than welcome to. For the second angle we'll note that the lines in the third and fourth quadrant make an angle of 0.7297 with the x-axis. So, if we start at the positive x-axis we rotate a half revolution and then add on 0.7297 for the second angle. Therefore, the second angle is $\pi + 0.7297 = 3.8713$.

Here's the rest of the work for this example.

$$\begin{aligned} 5z &= 5.5535 + 2\pi n \\ 5z &= 3.8713 + 2\pi n \end{aligned} \quad \Rightarrow \quad \begin{aligned} z &= 1.1107 + \frac{2\pi n}{5} \\ z &= 0.7743 + \frac{2\pi n}{5} \end{aligned} \qquad n = 0, \pm 1, \pm 2, \ldots$$

$$\begin{aligned} n &= -1 \quad : \quad x = \cancel{-0.1460} \quad \text{and} \quad \cancel{-0.4823} \\ n &= 0 \quad : \quad \cancel{x = 1.1107} \quad \text{and} \quad 0.7743 \end{aligned}$$

So, in this case we get a single solution of 0.7743.

Note that in the previous example we only got a single solution. This happens on occasion so don't get worried about it. Also, note that it was the second angle that gave this solution and so if we'd just relied on our calculator without worrying about other angles we would not have gotten this solution. Again, it can't be stressed enough that while calculators are a great tool if we don't understand how to correctly interpret/use the result we can (and often will) get the solution wrong.

To this point we've only worked examples involving sine and cosine. Let's now work a couple of examples that involve other trig functions to see how they work.

Example 5

Solve $9\sin(2x) = -5\cos(2x)$ on $[-10, 0]$.

Solution

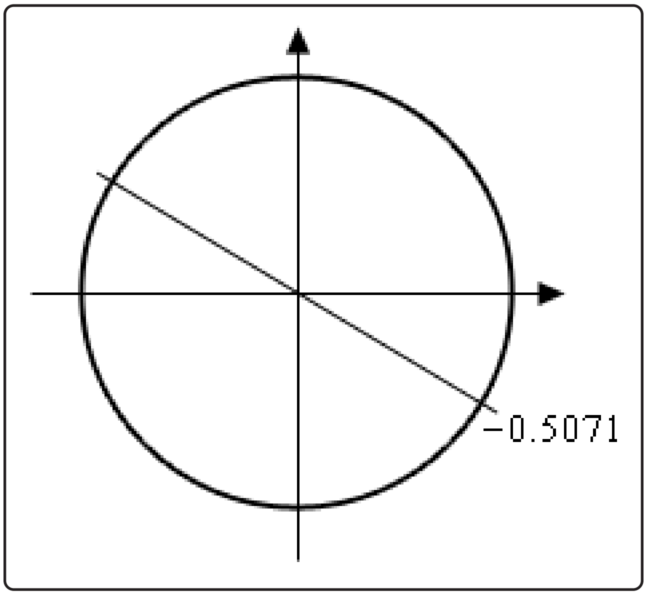
At first glance this problem seems to be at odds with the sentence preceding the example. However, it really isn't.

First, when we have more than one trig function in an equation we need a way to get equations that only involve one trig function. There are many ways of doing this that depend on the type of equation we're starting with. In this case we can simply divide both sides by a cosine and we'll get a single tangent in the equation. We can now see that this really is an equation that doesn't involve a sine or a cosine.

So, let's get started on this example.

$$\frac{\sin(2x)}{\cos(2x)} = \tan(2x) = -\frac{5}{9} \qquad \Rightarrow \qquad 2x = \tan^{-1}\left(-\frac{5}{9}\right) = -0.5071$$

Now, the unit circle doesn't involve tangents, however we can use it to illustrate the second angle in the range $[0, 2\pi]$.



The angles that we're looking for here are those whose quotient of $\frac{\text{sine}}{\text{cosine}}$ is the same. The second angle where we will get the same value of tangent will be exactly opposite of the given point. For this angle the values of sine and cosine are the same except they will have opposite signs. In the quotient however, the difference in signs will cancel out and we'll get the same value of tangent. So, the second angle will always be the first angle plus π.

Before getting the second angle let's also note that, like the previous example, we'll use the $2\pi - 0.5071 = 5.7761$ for the first angle. Again, this is only because of a concern about losing track of the minus sign in our calculator answer. We could just as easily do the work with the original angle our calculator gave us.

Now, this is where is seems like we're just randomly making changes and doing things for no reason. The second angle that we're going to use is,

$$\pi + (-0.5071) = \pi - 0.5071 = 2.6345$$

The fact that we used the calculator answer here seems to contradict the fact that we used a different angle for the first above. The reason for doing this here is to give a second angle that is in the range $[0, 2\pi]$. Had we used 5.7761 to find the second angle we'd get

$\pi + 5.7761 = 8.9177$. This is a perfectly acceptable answer; however, it is larger than 2π (6.2832) and the general rule of thumb is to keep the initial angles as small as possible.

Here are all the solutions to the equation.

$$\begin{aligned} 2x &= 5.7761 + 2\pi n \\ 2x &= 2.6345 + 2\pi n \end{aligned} \quad \Rightarrow \quad \begin{aligned} x &= 2.8881 + \pi n \\ x &= 1.3173 + \pi n \end{aligned} \quad n = 0, \pm 1, \pm 2, \ldots$$

$$\begin{array}{lcll} n = -4 & : & x = -9.6783 & \text{and} \quad \cancel{-11.2491} \\ n = -3 & : & x = -6.5367 & \text{and} \quad -8.1075 \\ n = -2 & : & x = -3.3951 & \text{and} \quad -4.9659 \\ n = -1 & : & x = -0.2535 & \text{and} \quad -1.8243 \\ n = 0 & : & \cancel{x = 2.8881} & \text{and} \quad \cancel{1.3173} \end{array}$$

The seven solutions to this equation are then,

$$-0.2535, \quad -1.8243, \quad -3.3951, \quad -4.9659, \quad -6.5367, \quad -8.1075, \quad -9.6783$$

Note as well that we didn't need to do the $n = 0$ computation since we could see from the given interval that we only wanted negative answers and these would clearly give positive answers.

Before moving on we need to address one issue about the previous example. The solution method used there is not the "standard" solution method. Because the second angle is just π plus the first and if we added π onto the second angle we'd be back at the line representing the first angle the more standard solution method is to just add πn onto the first angle.

If using the calculator answer this would give,

$$2x = -0.5071 + \pi n, \quad n = 0, \pm 1, \pm 2, \ldots$$

If using the positive angle corresponding to the calculator answer this would give,

$$2x = 5.7761 + \pi n, \quad n = 0, \pm 1, \pm 2, \ldots$$

Then dividing by 2 either of the following sets of solutions,

$$x = -0.2535 + \frac{\pi n}{2}, \quad n = 0, \pm 1, \pm 2, \ldots$$
$$x = 2.8881 + \frac{\pi n}{2}, \quad n = 0, \pm 1, \pm 2, \ldots$$

Either of these sets of solutions is identical to the set of solutions we got in the example (we'll leave it to you to plug in some n's and verify that). So, why did we not use the method in the previous example? Simple. The method in the previous example more closely mirrors the solution method for cosine and sine (*i.e.* they both, generally, give two sets of angles) and so for students that aren't comfortable with solving trig equations this gives a "consistent" solution method.

Many calculators today can only do inverse sine, inverse cosine, and inverse tangent. So, let's see an example that uses one of the other trig functions.

Example 6

Solve $7\sec(3t) = -10$.

Solution

We'll start this one in exactly the same way we've done all the others.

$$\sec(3t) = -\frac{10}{7} \qquad \Rightarrow \qquad 3t = \sec^{-1}\left(-\frac{10}{7}\right)$$

Now we reach the problem. As noted above, many calculators can't handle inverse secant so we're going to need a different solution method for this one. To finish the solution here we'll simply recall the definition of secant in terms of cosine and convert this into an equation involving cosine instead and we already know how to solve those kinds of trig equations.

$$\frac{1}{\cos(3t)} = \sec(3t) = -\frac{10}{7} \qquad \Rightarrow \qquad \cos(3t) = -\frac{7}{10}$$

Now, we solved this equation in the second example above so we won't redo our work here. The solution is,

$$\begin{aligned} t &= 0.7821 + \frac{2\pi n}{3} \\ t &= 1.3123 + \frac{2\pi n}{3} \end{aligned} \qquad n = 0, \pm 1, \pm 2, \ldots$$

We weren't given an interval in this problem so here is nothing else to do here.

For the remainder of the examples in this section we're not going to be finding solutions in an interval to save some space. If you followed the work from the first few examples in which we were given intervals you should be able to do any of the remaining examples if given an interval.

Also, we will no longer be including sketches of unit circles in the remaining solutions. We are going to assume that you can use the above sketches as guides for sketching unit circles to verify the claims in the following examples.

The next three examples don't require a calculator but are important enough or cause enough problems for students to include in this section in case you run across them and haven't seen them anywhere else.

Example 7

Solve cos $(4\theta) = -1$.

Solution

There really isn't too much to do with this problem. It is, however, different from all the others done to this point. All the others done to this point have had two angles in the interval $[0, 2\pi]$ that were solutions to the equation. This only has one. If you aren't sure you believe this sketch a quick unit circle and you'll see that in fact there is only one angle for which cosine is -1.

Here is the solution to this equation.

$$4\theta = \pi + 2\pi n \qquad \Rightarrow \qquad \theta = \frac{\pi}{4} + \frac{\pi n}{2} \qquad n = 0, \pm 1, \pm 2, \ldots$$

Example 8

Solve sin $\left(\frac{\alpha}{7}\right) = 0$.

Solution

Again, not much to this problem. Using a unit circle it isn't too hard to see that the solutions to this equation are,

$$\begin{aligned} \frac{\alpha}{7} &= 0 + 2\pi n \\ \frac{\alpha}{7} &= \pi + 2\pi n \end{aligned} \qquad \Rightarrow \qquad \begin{aligned} \alpha &= 14\pi n \\ \alpha &= 7\pi + 14\pi n \end{aligned} \qquad n = 0, \pm 1, \pm 2, \ldots$$

This next example has an important point that needs to be understood when solving some trig equations.

Example 9

Solve $\sin(3t) = 2$.

Solution

This example is designed to remind you of certain properties about sine and cosine. Recall that $-1 \leq \sin(\theta) \leq 1$ and $-1 \leq \cos(\theta) \leq 1$. Therefore, since sine will never be greater that 1 it definitely can't be 2. So, **THERE ARE NO SOLUTIONS** to this equation!

It is important to remember that not all trig equations will have solutions.

Because this document is also being prepared for viewing on the web we're going to split this section in two in order to keep the page size (and hence load time in a browser) to a minimum. In the next section we're going to take a look at some slightly more "complicated" equations. Although, as you'll see, they aren't as complicated as they may at first seem.

1.6 Solving Trig Equations with Calculators, Part II

Because this document is also being prepared for viewing on the web we split this section into two parts to keep the size of the pages to a minimum.

Also, as with the last few examples in the previous part of this section we are not going to be looking for solutions in an interval in order to save space. The important part of these examples is to find the solutions to the equation. If we'd been given an interval it would be easy enough to find the solutions that actually fall in the interval.

In all the examples in the previous section all the arguments, the $3t$, $\frac{\alpha}{7}$, *etc.*, were fairly simple. Let's take a look at an example that has a slightly more complicated looking argument.

Example 1

Solve $5\cos(2x-1) = -3$.

Solution

Note that the argument here is not really all that complicated but the addition of the "-1" often seems to confuse people so we need to a quick example with this kind of argument. The solution process is identical to all the problems we've done to this point so we won't be putting in much explanation. Here is the solution.

$$\cos(2x-1) = -\frac{3}{5} \qquad \Rightarrow \qquad 2x-1 = \cos^{-1}\left(-\frac{3}{5}\right) = 2.2143$$

This angle is in the second quadrant and so we can use either -2.2143 or $2\pi - 2.2143 = 4.0689$ for the second angle. As usual for these notes we'll use the positive one. Therefore the two angles are,

$$\begin{aligned} 2x-1 &= 2.2143 + 2\pi n \\ 2x-1 &= 4.0689 + 2\pi n \end{aligned} \qquad n = 0, \pm 1, \pm 2, \ldots$$

Now, we still need to find the actual values of x that are the solutions. These are found in the same manner as all the problems above. We'll first add 1 to both sides and then divide by 2. Doing this gives,

$$\begin{aligned} x &= 1.6072 + \pi n \\ x &= 2.5345 + \pi n \end{aligned} \qquad n = 0, \pm 1, \pm 2, \ldots$$

So, in this example we saw an argument that was a little different from those seen previously, but not all that different when it comes to working the problems so don't get too excited about it.

We now need to move into a different type of trig equation. All of the trig equations solved to this point (the previous example as well as the previous section) were, in some way, more or less the "standard" trig equation that is usually solved in a trig class. There are other types of equations involving trig functions however that we need to take a quick look at. The remaining examples show some of these different kinds of trig equations.

Example 2

Solve $2\cos(6y) + 11\cos(6y)\sin(3y) = 0$.

Solution

So, this definitely doesn't look like any of the equations we've solved to this point and initially the process is different as well. First, notice that there is a $\cos(6y)$ in each term, so let's factor that out and see what we have.

$$\cos(6y)\left(2 + 11\sin(3y)\right) = 0$$

We now have a product of two terms that is zero and so we know that we must have,

$$\cos(6y) = 0 \qquad \text{OR} \qquad 2 + 11\sin(3y) = 0$$

Now, at this point we have two trig equations to solve and each is identical to the type of equation we were solving earlier. Because of this we won't put in much detail about solving these two equations.

First, solving $\cos(6y) = 0$ gives,

$$\begin{array}{lcll} 6y = \dfrac{\pi}{2} + 2\pi n & & y = \dfrac{\pi}{12} + \dfrac{\pi n}{3} & \\ & \Rightarrow & & n = 0, \pm 1, \pm 2, \ldots \\ 6y = \dfrac{3\pi}{2} + 2\pi n & & y = \dfrac{\pi}{4} + \dfrac{\pi n}{3} & \end{array}$$

Next, solving $2 + 11\sin(3y) = 0$ gives,

$$\begin{array}{lcll} 3y = 6.1004 + 2\pi n & & y = 2.0335 + \dfrac{2\pi n}{3} & \\ & \Rightarrow & & n = 0, \pm 1, \pm 2, \ldots \\ 3y = 3.3244 + 2\pi n & & y = 1.1081 + \dfrac{2\pi n}{3} & \end{array}$$

Remember that in these notes we tend to take positive angles and so the first solution here is in fact $2\pi - 0.1828$ where our calculator gave us -0.1828 as the answer when using the inverse sine function.

The solutions to this equation are then,

$$\begin{aligned} y &= \frac{\pi}{12} + \frac{\pi n}{3} \\ y &= \frac{\pi}{4} + \frac{\pi n}{3} \qquad\qquad n = 0, \pm 1, \pm 2, \ldots \\ y &= 2.0335 + \frac{2\pi n}{3} \\ y &= 1.1081 + \frac{2\pi n}{3} \end{aligned}$$

This next example also involves "factoring" trig equations but in a slightly different manner than the previous example.

Example 3

Solve $4\sin^2\left(\frac{t}{3}\right) - 3\sin\left(\frac{t}{3}\right) = 1$.

Solution

Before solving this equation let's solve an apparently unrelated equation.

$$4x^2 - 3x = 1 \qquad \Rightarrow \qquad 4x^2 - 3x - 1 = (4x+1)(x-1) = 0 \qquad \Rightarrow \qquad x = -\frac{1}{4}, 1$$

This is an easy (or at least we hope it's easy at this point) equation to solve. The obvious question then is, why did we do this? We'll, if you compare the two equations you'll see that the only real difference is that the one we just solved has an x everywhere and the equation we want to solve has a sine. What this tells us is that we can work the two equations in exactly the same way.

We, will first "factor" the equation as follows,

$$4\sin^2\left(\frac{t}{3}\right) - 3\sin\left(\frac{t}{3}\right) - 1 = \left(4\sin\left(\frac{t}{3}\right) + 1\right)\left(\sin\left(\frac{t}{3}\right) - 1\right) = 0$$

Now, set each of the two factors equal to zero and solve for the sine,

$$\sin\left(\frac{t}{3}\right) = -\frac{1}{4} \qquad\qquad \sin\left(\frac{t}{3}\right) = 1$$

We now have two trig equations that we can easily (hopefully...) solve at this point. We'll leave the details to you to verify that the solutions to each of these and hence the solutions to the original equation are,

$$\begin{aligned} t &= 18.0915 + 6\pi n \\ t &= 10.1829 + 6\pi n \qquad n = 0, \pm 1, \pm 2, \ldots \\ t &= \frac{3\pi}{2} + 6\pi n \end{aligned}$$

The first two solutions are from the first equation and the third solution is from the second equation.

Let's work one more trig equation that involves solving a quadratic equation. However, this time, unlike the previous example this one won't factor and so we'll need to use the quadratic formula.

Example 4

Solve $8\cos^2(1-x) + 13\cos(1-x) - 5 = 0$.

Solution

Now, as mentioned prior to starting the example this quadratic does not factor. However, that doesn't mean all is lost. We can solve the following equation with the quadratic formula (you do remember this and how to use it right?),

$$8t^2 + 13t - 5 = 0 \qquad \Rightarrow \qquad t = \frac{-13 \pm \sqrt{329}}{16} = 0.3211, \quad -1.9461$$

So, if we can use the quadratic formula on this then we can also use it on the equation we're asked to solve. Doing this gives us,

$$\cos(1-x) = 0.3211 \qquad \text{OR} \qquad \cos(1-x) = -1.9461$$

Now, recall Example 9 from the previous section. In that example we noted that $-1 \le \cos(\theta) \le 1$ and so the second equation will have no solutions. Therefore, the solutions to the first equation will yield the only solutions to our original equation. Solving this gives the following set of solutions,

$$\begin{aligned} x &= -0.2439 - 2\pi n \\ x &= -4.0393 - 2\pi n \end{aligned} \qquad n = 0, \pm 1, \pm 2, \ldots$$

Note that we did get some negative numbers here and that does seem to violate the general form that we've been using in most of these examples. However, in this case the "−" are coming about when we solved for x after computing the inverse cosine in our calculator.

There is one more example in this section that we need to work that illustrates another way in which factoring can arise in solving trig equations. This equation is also the only one where the variable appears both inside and outside of the trig equation. Not all equations in this form can be easily solved, however some can so we want to do a quick example of one.

Example 5

Solve $5x\tan(8x) = 3x$.

Solution

First, before we even start solving we need to make one thing clear. **DO NOT CANCEL AN x FROM BOTH SIDES!!!** While this may seem like a natural thing to do it **WILL** cause us to lose a solution here.

So, to solve this equation we'll first get all the terms on one side of the equation and then factor an x out of the equation. If we can cancel an x from all terms then it can be factored out. Doing this gives,

$$5x\tan(8x) - 3x = x(5\tan(8x) - 3) = 0$$

Upon factoring we can see that we must have either,

$$x = 0 \qquad \text{OR} \qquad \tan(8x) = \frac{3}{5}$$

Note that if we'd canceled the x we would have missed the first solution. Now, we solved an equation with a tangent in it in Example 5 of the previous section so we'll not go into the details of this solution here. Here is the solution to the trig equation.

$$\begin{aligned} x &= 0.0676 + \frac{\pi n}{4} \\ & & n = 0, \pm 1, \pm 2, \ldots \\ x &= 0.4603 + \frac{\pi n}{4} \end{aligned}$$

The complete set of solutions then to the original equation are,

$$\begin{aligned} x &= 0 \\ x &= 0.0676 + \frac{\pi n}{4} & & n = 0, \pm 1, \pm 2, \ldots \\ x &= 0.4603 + \frac{\pi n}{4} \end{aligned}$$

1.7 Exponential Functions

In this section we're going to review one of the more common functions in both calculus and the sciences. However, before getting to this function let's take a much more general approach to things.

Let's start with $b > 0$, $b \neq 1$. An exponential function is then a function in the form,

$$f(x) = b^x$$

Note that we avoid $b = 1$ because that would give the constant function, $f(x) = 1$. We avoid $b = 0$ since this would also give a constant function and we avoid negative values of b for the following reason.

Let's, for a second, suppose that we did allow b to be negative and look at the following function.

$$g(x) = (-4)^x$$

Let's do some evaluation.

$$g(2) = (-4)^2 = 16 \qquad g\left(\frac{1}{2}\right) = (-4)^{\frac{1}{2}} = \sqrt{-4} = 2i$$

So, for some values of x we will get real numbers and for other values of x we will get complex numbers. We want to avoid this so if we require $b > 0$ this will not be a problem.

Let's take a look at a couple of exponential functions.

Example 1

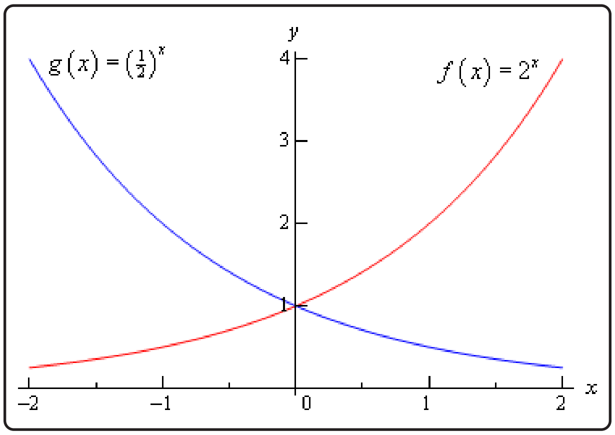
Sketch the graph of $f(x) = 2^x$ and $g(x) = \left(\frac{1}{2}\right)^x$.

Solution

Let's first get a table of values for these two functions.

x	$f(x)$	$g(x)$
-2	$f(-2) = 2^{-2} = \frac{1}{4}$	$g(-2) = \left(\frac{1}{2}\right)^{-2} = 4$
-1	$f(-1) = 2^{-1} = \frac{1}{2}$	$g(-1) = \left(\frac{1}{2}\right)^{-1} = 2$
0	$f(0) = 2^0 = 1$	$g(0) = \left(\frac{1}{2}\right)^0 = 1$
1	$f(1) = 2$	$g(1) = \frac{1}{2}$
2	$f(2) = 4$	$g(2) = \frac{1}{4}$

Here's the sketch of both of these functions.



This graph illustrates some very nice properties about exponential functions in general.

Properties of $f(x) = b^x$

1. $f(0) = 1$. The function will always take the value of 1 at $x = 0$.
2. $f(x) \neq 0$. An exponential function will never be zero.
3. $f(x) > 0$. An exponential function is always positive.
4. The previous two properties can be summarized by saying that the range of an exponential function is $(0, \infty)$.
5. The domain of an exponential function is $(-\infty, \infty)$. In other words, you can plug every x into an exponential function.
6. If $0 < b < 1$ then,
   (a) $f(x) \to 0$ as $x \to \infty$
   (b) $f(x) \to \infty$ as $x \to -\infty$
7. If $b > 1$ then,
   (a) $f(x) \to \infty$ as $x \to \infty$
   (b) $f(x) \to 0$ as $x \to -\infty$

These will all be very useful properties to recall at times as we move throughout this course (and later Calculus courses for that matter...).

There is a very important exponential function that arises naturally in many places. This function is called the **natural exponential function**. However, for most people, this is simply the exponential function.

Definition

The **natural exponential function** is $f(x) = \mathbf{e}^x$ where, $\mathbf{e} = 2.71828182845905\ldots$.

So, since $\mathbf{e} > 1$ we also know that $\mathbf{e}^x \to \infty$ as $x \to \infty$ and $\mathbf{e}^x \to 0$ as $x \to -\infty$.

Let's take a quick look at an example.

Example 2

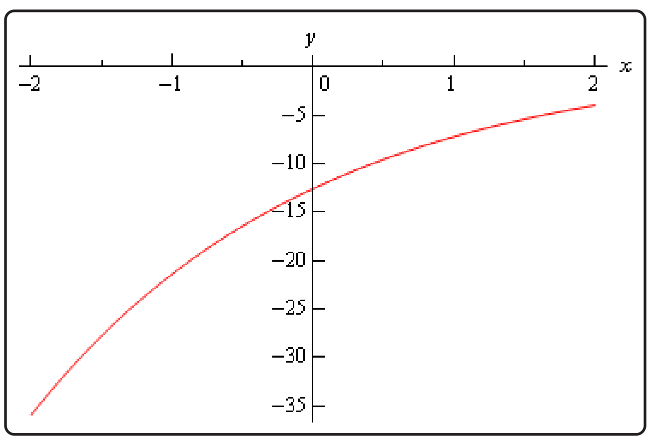
Sketch the graph of $h(t) = 1 - 5\mathbf{e}^{1-\frac{t}{2}}$.

Solution

Let's first get a table of values for this function.

t	-2	-1	0	1	2	3
$h(t)$	-35.9453	-21.4084	-12.5914	-7.2436	-4	-2.0327

Here is the sketch.



The main point behind this problem is to make sure you can do this type of evaluation so make sure that you can get the values that we graphed in this example. You will be asked to do this kind of evaluation on occasion in this class.

You will be seeing exponential functions in pretty much every chapter in this class so make sure that you are comfortable with them.

1.8 Logarithm Functions

In this section we'll take a look at a function that is related to the exponential functions we looked at in the last section. We will look at logarithms in this section. Logarithms are one of the functions that students fear the most. The main reason for this seems to be that they simply have never really had to work with them. Once they start working with them, students come to realize that they aren't as bad as they first thought.

We'll start with $b > 0$, $b \neq 1$ just as we did in the last section. Then we have

$$y = \log_b x \qquad \text{is equivalent to} \qquad x = b^y$$

The first is called logarithmic form and the second is called the exponential form. Remembering this equivalence is the key to evaluating logarithms. The number, b, is called the base.

Let's do some quick evaluations.

Example 1

Without a calculator give the exact value of each of the following logarithms.

(a) $\log_2 16$

(b) $\log_4 16$

(c) $\log_5 625$

(d) $\log_9 \frac{1}{531441}$

(e) $\log_{\frac{1}{6}} 36$

(f) $\log_{\frac{3}{2}} \frac{27}{8}$

Solution

To quickly evaluate logarithms the easiest thing to do is to convert the logarithm to exponential form. So, let's take a look at the first one.

(a) $\log_2 16$

First, let's convert to exponential form.

$$\log_2 16 = ? \qquad \text{is equivalent to} \qquad 2^? = 16$$

So, we're really asking 2 raised to what gives 16. Since 2 raised to 4 is 16 we get,

$$\log_2 16 = 4 \qquad \text{because} \qquad 2^4 = 16$$

We'll not do the remaining parts in quite this detail, but they will all work in this way.

(b) $\log_4 16$

$$\log_4 16 = 2 \qquad \text{because} \qquad 4^2 = 16$$

Note the difference between the first and second logarithm! The base is important! It can completely change the answer.

(c) $\log_5 625$

$$\log_5 625 = 4 \qquad \text{because} \qquad 5^4 = 625$$

(d) $\log_9 \frac{1}{531441}$

$$\log_9 \frac{1}{531441} = -6 \qquad \text{because} \qquad 9^{-6} = \frac{1}{9^6} = \frac{1}{531441}$$

(e) $\log_{\frac{1}{6}} 36$

$$\log_{\frac{1}{6}} 36 = -2 \qquad \text{because} \qquad \left(\frac{1}{6}\right)^{-2} = 6^2 = 36$$

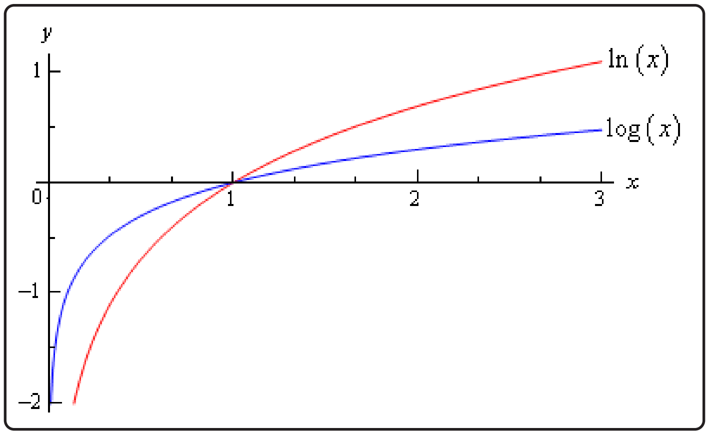
(f) $\log_{\frac{3}{2}} \frac{27}{8}$

$$\log_{\frac{3}{2}} \frac{27}{8} = 3 \qquad \text{because} \qquad \left(\frac{3}{2}\right)^{3} = \frac{27}{8}$$

There are a couple of special logarithms that arise in many places. These are,

$$\ln(x) = \log_{\mathbf{e}}(x) \qquad \text{This log is called the natural logarithm}$$
$$\log(x) = \log_{10}(x) \qquad \text{This log is called the common logarithm}$$

In the natural logarithm the base **e** is the same number as in the natural exponential logarithm that we saw in the last section. Here is a sketch of both of these logarithms.



From this graph we can get a couple of very nice properties about the natural logarithm that we will use many times in this and later Calculus courses.

$$\ln(x) \to \infty \text{ as } x \to \infty$$
$$\ln(x) \to -\infty \text{ as } x \to 0, \ x > 0$$

Let's take a look at a couple of more logarithm evaluations. Some of which deal with the natural or common logarithm and some of which don't.

Example 2

Without a calculator give the exact value of each of the following logarithms.

(a) $\ln \sqrt[3]{\mathbf{e}}$

(b) $\log 1000$

(c) $\log_{16} 16$

(d) $\log_{23} 1$

(e) $\log_2 \sqrt[7]{32}$

Solution

These work exactly the same as previous example so we won't put in too many details.

(a) $\ln \sqrt[3]{\mathbf{e}}$

$$\ln \sqrt[3]{\mathbf{e}} = \frac{1}{3} \qquad \text{because} \qquad \mathbf{e}^{\frac{1}{3}} = \sqrt[3]{\mathbf{e}}$$

(b) $\log 1000$

$$\log 1000 = 3 \qquad \text{because} \qquad 10^3 = 1000$$

(c) $\log_{16} 16$

$$\log_{16} 16 = 1 \qquad \text{because} \qquad 16^1 = 16$$

(d) $\log_{23} 1$

$$\log_{23} 1 = 0 \qquad \text{because} \qquad 23^0 = 1$$

(e) $\log_2 \sqrt[7]{32}$

$$\log_2 \sqrt[7]{32} = \frac{5}{7} \qquad \text{because} \qquad \sqrt[7]{32} = 32^{\frac{1}{7}} = \left(2^5\right)^{\frac{1}{7}} = 2^{\frac{5}{7}}$$

This last set of examples leads us to some of the basic properties of logarithms.

Properties

1. The domain of the logarithm function is $(0, \infty)$. In other words, we can only plug positive numbers into a logarithm! We can't plug in zero or a negative number.
2. The range of the logarithm function is $(-\infty, \infty)$.
3. $\log_b(b) = 1$
4. $\log_b(1) = 0$
5. $\log_b(b^x) = x$
6. $b^{\log_b(x)} = x$

The last two properties will be especially useful in the next section. Notice as well that these last two properties tell us that,

$$f(x) = b^x \qquad \text{and} \qquad g(x) = \log_b(x)$$

are inverses of each other.

Here are some more properties that are useful in the manipulation of logarithms.

More Properties

5. $\log_b(xy) = \log_b(x) + \log_b(y)$
6. $\log_b\left(\frac{x}{y}\right) = \log_b(x) - \log_b(y)$
7. $\log_b\left(x^r\right) = r\log_b(x)$

Note that there is no equivalent property to the first two for sums and differences. In other words,

$$\log_b(x+y) \neq \log_b(x) + \log_b(y)$$
$$\log_b(x-y) \neq \log_b(x) - \log_b(y)$$

Example 3

Write each of the following in terms of simpler logarithms.

(a) $\ln\left(x^3y^4z^5\right)$

(b) $\log_3\left(\frac{9x^4}{\sqrt{y}}\right)$

(c) $\log\left(\frac{x^2+y^2}{(x-y)^3}\right)$

Solution

What the instructions really mean here is to use as many of the properties of logarithms as we can to simplify things down as much as we can.

(a) $\ln\left(x^3y^4z^5\right)$

Property 7 above can be extended to products of more than two functions. Once we've used Property 7 we can then use Property 9.

$$\begin{aligned}\ln\left(x^3y^4z^5\right) &= \ln(x^3) + \ln(y^4) + \ln(z^5)\\ &= 3\ln(x) + 4\ln(y) + 5\ln(z)\end{aligned}$$

(b) $\log_3\left(\frac{9x^4}{\sqrt{y}}\right)$

When using property 8 above make sure that the logarithm that you subtract is the one that contains the denominator as its argument. Also, note that that we'll be converting the root to fractional exponents in the first step.

$$\log_3\left(\frac{9x^4}{\sqrt{y}}\right) = \log_3(9x^4) - \log_3(y^{\frac{1}{2}})$$
$$= \log_3(9) + \log_3(x^4) - \log_3(y^{\frac{1}{2}})$$
$$= 2 + 4\log_3(x) - \frac{1}{2}\log_3(y)$$

(c) $\log\left(\frac{x^2+y^2}{(x-y)^3}\right)$

The point to this problem is mostly the correct use of property 9 above.

$$\log\left(\frac{x^2+y^2}{(x-y)^3}\right) = \log\left(x^2+y^2\right) - \log\left(x-y\right)^3$$
$$= \log\left(x^2+y^2\right) - 3\log\left(x-y\right)$$

You can use Property 9 on the second term because the WHOLE term was raised to the 3, but in the first logarithm, only the individual terms were squared and not the term as a whole so the 2's must stay where they are!

The last topic that we need to look at in this section is the **change of base** formula for logarithms. The change of base formula is,

$$\log_b(x) = \frac{\log_a(x)}{\log_a(b)}$$

This is the most general change of base formula and will convert from base b to base a. However, the usual reason for using the change of base formula is to compute the value of a logarithm that is in a base that you can't easily deal with. Using the change of base formula means that you can write the logarithm in terms of a logarithm that you can deal with. The two most common change of base formulas are

$$\log_b(x) = \frac{\ln(x)}{\ln(b)} \qquad \text{and} \qquad \log_b(x) = \frac{\log(x)}{\log(b)}$$

In fact, often you will see one or the other listed as THE change of base formula!

In the first part of this section we computed the value of a few logarithms, but we could do these without the change of base formula because all the arguments could be written in terms of the base to a power. For instance,

$$\log_7(49) = 2 \qquad \text{because} \qquad 7^2 = 49$$

However, this only works because 49 can be written as a power of 7! We would need the change of base formula to compute $\log_7 50$.

$$\log_7(50) = \frac{\ln(50)}{\ln(7)} = \frac{3.91202300543}{1.94591014906} = 2.0103821378$$

OR

$$\log_7(50) = \frac{\log(50)}{\log(7)} = \frac{1.69897000434}{0.845098040014} = 2.0103821378$$

So, it doesn't matter which we use, we will get the same answer regardless of the logarithm that we use in the change of base formula.

Note as well that we could use the change of base formula on $\log_7 49$ if we wanted to as well.

$$\log_7(49) = \frac{\ln(49)}{\ln(7)} = \frac{3.89182029811}{1.94591014906} = 2$$

This is a lot of work however, and is probably not the best way to deal with this.

So, in this section we saw how logarithms work and took a look at some of the properties of logarithms. We will run into logarithms on occasion so make sure that you can deal with them when we do run into them.

1.9 Exponential and Logarithm Equations

In this section we'll take a look at solving equations with exponential functions or logarithms in them.

We'll start with equations that involve exponential functions. The main property that we'll need for these equations is,

$$\log_b(b^x) = x$$

Example 1

Solve $7 + 15\mathbf{e}^{1-3z} = 10$.

Solution

The first step is to get the exponential all by itself on one side of the equation with a coefficient of one.

$$\begin{aligned} 7 + 15\mathbf{e}^{1-3z} &= 10 \\ 15\mathbf{e}^{1-3z} &= 3 \\ \mathbf{e}^{1-3z} &= \frac{1}{5} \end{aligned}$$

Now, we need to get the z out of the exponent so we can solve for it. To do this we will use the property above. Since we have an **e** in the equation we'll use the natural logarithm. First, we take the logarithm of both sides and then use the property to simplify the equation.

$$\begin{aligned} \ln\left(\mathbf{e}^{1-3z}\right) &= \ln\left(\frac{1}{5}\right) \\ 1 - 3z &= \ln\left(\frac{1}{5}\right) \end{aligned}$$

All we need to do now is solve this equation for z.

$$\begin{aligned} 1 - 3z &= \ln\left(\frac{1}{5}\right) \\ -3z &= -1 + \ln\left(\frac{1}{5}\right) \\ z &= -\frac{1}{3}\left(-1 + \ln\left(\frac{1}{5}\right)\right) = 0.8698126372 \end{aligned}$$

Example 2

Solve $10^{t^2-t} = 100$.

Solution

Now, in this case it looks like the best logarithm to use is the common logarithm since left hand side has a base of 10. There's no initial simplification to do, so just take the log of both sides and simplify.

$$\begin{aligned} \log 10^{t^2-t} &= \log 100 = \log 10^2 = 2 \\ t^2 - t &= 2 \end{aligned}$$

At this point, we've just got a quadratic that can be solved

$$\begin{aligned} t^2 - t - 2 &= 0 \\ (t-2)(t+1) &= 0 \end{aligned}$$

So, it looks like the solutions in this case are $t = 2$ and $t = -1$.

Now that we've seen a couple of equations where the variable only appears in the exponent we need to see an example with variables both in the exponent and out of it.

Example 3

Solve $x - x\mathbf{e}^{5x+2} = 0$.

Solution

The first step is to factor an x out of both terms.

DO NOT DIVIDE AN x FROM BOTH TERMS!!!!

Note that it is very tempting to "simplify" the equation by dividing an x out of both terms. However, if you do that you'll miss a solution as we'll see.

$$\begin{aligned} x - x\mathbf{e}^{5x+2} &= 0 \\ x\left(1 - \mathbf{e}^{5x+2}\right) &= 0 \end{aligned}$$

So, it's now a little easier to deal with. From this we can see that we get one of two possibilities.

$$x = 0 \qquad \text{OR} \qquad 1 - \mathbf{e}^{5x+2} = 0$$

The first possibility has nothing more to do, except notice that if we had divided both sides by an x we would have missed this one so be careful. In the second possibility we've got a little more to do. This is an equation similar to the first two that we did in this section.

$$\begin{aligned}\mathbf{e}^{5x+2} &= 1\\ 5x+2 &= \ln 1\\ 5x+2 &= 0\\ x &= -\frac{2}{5}\end{aligned}$$

Don't forget that $\ln 1 = 0$.

So, the two solutions are $x = 0$ and $x = -\frac{2}{5}$.

The next equation is a more complicated (looking at least...) example similar to the previous one.

Example 4

Solve $5\left(x^2-4\right) = \left(x^2-4\right)\mathbf{e}^{7-x}$.

Solution

As with the previous problem do NOT divide an x^2-4 out of both sides. Doing this will lose solutions even though it "simplifies" the equation. Note however, that if you can divide a term out then you can also factor it out if the equation is written properly.

So, the first step here is to move everything to one side of the equation and then to factor out the x^2-4.

$$\begin{aligned}5\left(x^2-4\right)-\left(x^2-4\right)\mathbf{e}^{7-x} &= 0\\ \left(x^2-4\right)\left(5-\mathbf{e}^{7-x}\right) &= 0\end{aligned}$$

At this point all we need to do is set each factor equal to zero and solve each.

$$\begin{aligned}x^2-4 &= 0 & \qquad 5-\mathbf{e}^{7-x} &= 0\\ x &= \pm 2 & \mathbf{e}^{7-x} &= 5\\ & & 7-x &= \ln(5)\\ & & x &= 7-\ln(5) = 5.390562088\end{aligned}$$

The three solutions are then $x = \pm 2$ and $x = 5.3906$.

As a final example let's take a look at an equation that contains two different exponentials.

Example 5

Solve $4\mathbf{e}^{1+3x} - 9\mathbf{e}^{5-2x} = 0$.

Solution

The first step here is to get one exponential on each side and then we'll divide both sides by one of them (which doesn't matter for the most part) so we'll have a quotient of two exponentials. The quotient can then be simplified and we'll finally get both coefficients on the other side. Doing all of this gives,

$$\begin{aligned} 4\mathbf{e}^{1+3x} &= 9\mathbf{e}^{5-2x} \\ \frac{\mathbf{e}^{1+3x}}{\mathbf{e}^{5-2x}} &= \frac{9}{4} \\ \mathbf{e}^{1+3x-(5-2x)} &= \frac{9}{4} \\ \mathbf{e}^{5x-4} &= \frac{9}{4} \end{aligned}$$

Note that while we said that it doesn't really matter which exponential we divide out by doing it the way we did here we'll avoid a negative coefficient on the x. Not a major issue, but those minus signs on coefficients are really easy to lose on occasion.

This is now in a form that we can deal with so here's the rest of the solution.

$$\begin{aligned} \mathbf{e}^{5x-4} &= \frac{9}{4} \\ 5x-4 &= \ln\left(\frac{9}{4}\right) \\ 5x &= 4+\ln\left(\frac{9}{4}\right) \\ x &= \frac{1}{5}\left(4+\ln\left(\frac{9}{4}\right)\right) = 0.9621860432 \end{aligned}$$

This equation has a single solution of $x = 0.9622$.

Now let's take a look at some equations that involve logarithms. The main property that we'll be using to solve these kinds of equations is,

$$b^{\log_b x} = x$$

Example 6

Solve $3 + 2\ln\left(\frac{x}{7} + 3\right) = -4$.

Solution

This first step in this problem is to get the logarithm by itself on one side of the equation with a coefficient of 1.

$$\begin{aligned} 2\ln\left(\frac{x}{7} + 3\right) &= -7 \\ \ln\left(\frac{x}{7} + 3\right) &= -\frac{7}{2} \end{aligned}$$

Now, we need to get the x out of the logarithm and the best way to do that is to "exponentiate" both sides using **e**. In other words,

$$\mathbf{e}^{\ln\left(\frac{x}{7}+3\right)} = \mathbf{e}^{-\frac{7}{2}}$$

So, using the property above with **e**, since there is a natural logarithm in the equation, we get,

$$\frac{x}{7} + 3 = \mathbf{e}^{-\frac{7}{2}}$$

Now all that we need to do is solve this for x.

$$\begin{aligned} \frac{x}{7} + 3 &= \mathbf{e}^{-\frac{7}{2}} \\ \frac{x}{7} &= -3 + \mathbf{e}^{-\frac{7}{2}} \\ x &= 7\left(-3 + \mathbf{e}^{-\frac{7}{2}}\right) = -20.78861832 \end{aligned}$$

At this point we might be tempted to say that we're done and move on. However, we do need to be careful. Recall from the previous section that we can't plug a negative number into a logarithm. This, by itself, doesn't mean that our answer won't work since its negative. What we need to do is plug it into the logarithm and make sure that $\frac{x}{7} + 3$ will not be negative. I'll leave it to you to verify that this is in fact positive upon plugging our solution into the logarithm and so $x = -20.78861832$ is a solution to the equation.

Let's now take a look at a more complicated equation. Often there will be more than one logarithm in the equation. When this happens we will need to use one or more of the following properties to combine all the logarithms into a single logarithm. Once this has been done we can proceed as we did in the previous example.

$$\log_b(xy) = \log_b(x) + \log_b(y) \qquad \log_b\left(\frac{x}{y}\right) = \log_b(x) - \log_b(y) \qquad \log_b\left(x^r\right) = r\log_b(x)$$

Example 7

Solve $2\ln\left(\sqrt{x}\right) - \ln\left(1-x\right) = 2$.

Solution

First get the two logarithms combined into a single logarithm.

$$\begin{aligned} 2\ln\left(\sqrt{x}\right) - \ln\left(1-x\right) &= 2 \\ \ln\left(\left(\sqrt{x}\right)^2\right) - \ln\left(1-x\right) &= 2 \\ \ln\left(x\right) - \ln\left(1-x\right) &= 2 \\ \ln\left(\frac{x}{1-x}\right) &= 2 \end{aligned}$$

Now, exponentiate both sides and solve for x.

$$\begin{aligned} \frac{x}{1-x} &= \mathbf{e}^2 \\ x &= \mathbf{e}^2\left(1-x\right) \\ x &= \mathbf{e}^2 - \mathbf{e}^2 x \\ x\left(1+\mathbf{e}^2\right) &= \mathbf{e}^2 \\ x &= \frac{\mathbf{e}^2}{1+\mathbf{e}^2} = 0.8807970780 \end{aligned}$$

The solution work here was a little messy but this is work that you will need to be able to do on occasion so make sure you can do it!

Finally, we just need to make sure that the solution, $x = 0.8807970780$, doesn't produce negative numbers in both of the original logarithms. It doesn't, so this is in fact our solution to this problem.

Let's take a look at another example.

Example 8

Solve $\log x + \log (x-3) = 1$.

Solution

As with the last example, first combine the logarithms into a single logarithm.

$$\begin{aligned} \log x + \log (x-3) &= 1 \\ \log (x(x-3)) &= 1 \end{aligned}$$

Now exponentiate, using 10 this time instead of **e** because we've got common logs in the equation, both sides.

$$\begin{aligned} 10^{\log(x^2-3x)} &= 10^1 \\ x^2 - 3x &= 10 \\ x^2 - 3x - 10 &= 0 \\ (x-5)(x+2) &= 0 \end{aligned}$$

So, potential solutions are $x = 5$ and $x = -2$. Note, however that if we plug $x = -2$ into either of the two original logarithms we would get negative numbers so this can't be a solution. We can however, use $x = 5$.

Therefore, the solution to this equation is $x = 5$.

When solving equations with logarithms it is important to check your potential solutions to make sure that they don't generate logarithms of negative numbers or zero. It is also important to make sure that you do the checks in the **original** equation. If you check them in the second logarithm above (after we've combined the two logs) both solutions will appear to work! This is because in combining the two logarithms we've actually changed the problem. In fact, it is this change that introduces the extra solution that we couldn't use!

Also, be careful in solving equations containing logarithms to not get locked into the idea that you will get two potential solutions and only one of these will work. It is possible to have problems where both are solutions and where neither are solutions.

There is one more problem that we should work.

Example 9

Solve $\ln(x-2)+\ln(x+1)=2$.

Solution

The first step of this problem is the same as we've been doing up to this point. So, let's combine the logarithms.

$$\begin{aligned} \ln((x-2)(x+1)) &= 2 \\ \ln\left(x^2-x-2\right) &= 2 \end{aligned}$$

Now we can exponentiate both sides with respect to **e** to eliminate the logarithm. Doing this along with a little simplification gives,

$$\begin{aligned} x^2-x-2 &= \mathbf{e}^2 \\ x^2-x-2-\mathbf{e}^2 &= 0 \end{aligned}$$

We've reached the point of this problem. We need to solve this quadratic and without the $\mathbf{e}^2$ everyone would be able to do that. However, with the $\mathbf{e}^2$ people tend to decide that they can't do it.

This is just a quadratic equation and everyone in this class should be able to solve that. The only difference between this quadratic equation and those you are probably used to seeing is that there are numbers in it that are not integers, or at worst, fractions. In this case the constant in the quadratic is just $-2-\mathbf{e}^2$ and so all we need to do is use the quadratic formula to get the solutions.

The solutions to this quadratic equation are,

$$x=\frac{1\pm\sqrt{1-4(1)\left(-2-\mathbf{e}^2\right)}}{2}=\frac{1\pm\sqrt{9+4\mathbf{e}^2}}{2}=-2.6047,\quad 3.6047$$

Do not get excited about the "messy" solutions to this quadratic. We will get these kinds of solutions on occasion.

The last step to this problem is to check the two solutions to the quadratic equation in the original equation. Doing that we can see that the first solution, -2.6047, will give negative numbers in the logarithms and so can't be a solution. On the other hand, the second solution, 3.6047, does not give negative numbers in the logarithms and so is okay.

The solution to the original equation is $x=3.6047$.

1.10 Common Graphs

The purpose of this section is to make sure that you're familiar with the graphs of many of the basic functions that you're liable to run across in a calculus class.

Example 1

Graph $y = -\frac{2}{5}x + 3$.

Solution

This is a line in the slope intercept form

$$y = mx + b$$

In this case the line has a y intercept of $(0, b)$ and a slope of m. Recall that slope can be thought of as

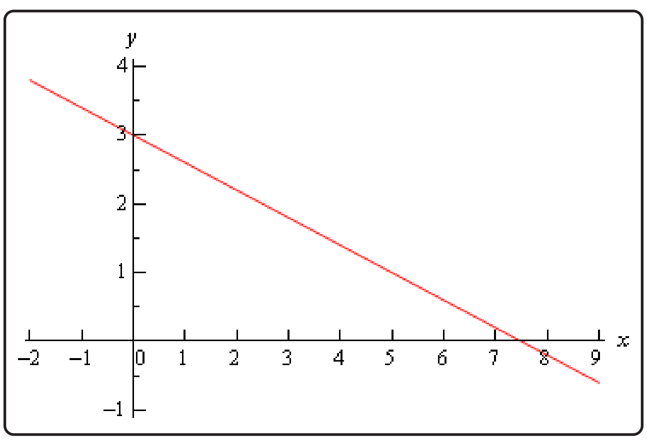
$$m = \frac{\text{rise}}{\text{run}}$$

Note that if the slope is negative we tend to think of the rise as a fall.

The slope allows us to get a second point on the line. Once we have any point on the line and the slope we move right by *run* and up/down by *rise* depending on the sign. This will be a second point on the line.

In this case we know $(0, 3)$ is a point on the line and the slope is $-\frac{2}{5}$. So starting at $(0, 3)$ we'll move 5 to the right (*i.e.* $0 \to 5$) and down 2 (*i.e.* $3 \to 1$) to get $(5, 1)$ as a second point on the line. Once we've got two points on a line all we need to do is plot the two points and connect them with a line.

Here's the sketch for this line.



Example 2

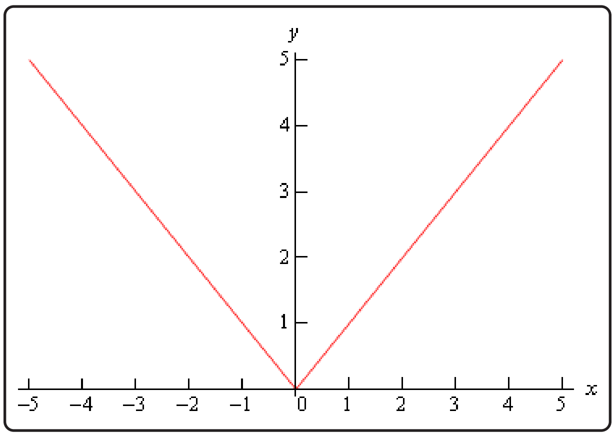
Graph $f(x) = |x|$.

Solution

There really isn't much to this problem outside of reminding ourselves of what absolute value is. Recall that the absolute value function is defined as,

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

The graph is then,



Example 3

Graph $f(x) = -x^2 + 2x + 3$.

Solution

This is a parabola in the general form.

$$f(x) = ax^2 + bx + c$$

In this form, the x-coordinate of the vertex (the highest or lowest point on the parabola) is $x = -\frac{b}{2a}$ and the y-coordinate is $y = f\left(-\frac{b}{2a}\right)$. So, for our parabola the coordinates of the vertex will be.

$$x = -\frac{2}{2(-1)} = 1$$

$$y = f(1) = -(1)^2 + 2(1) + 3 = 4$$

So, the vertex for this parabola is $(1, 4)$.

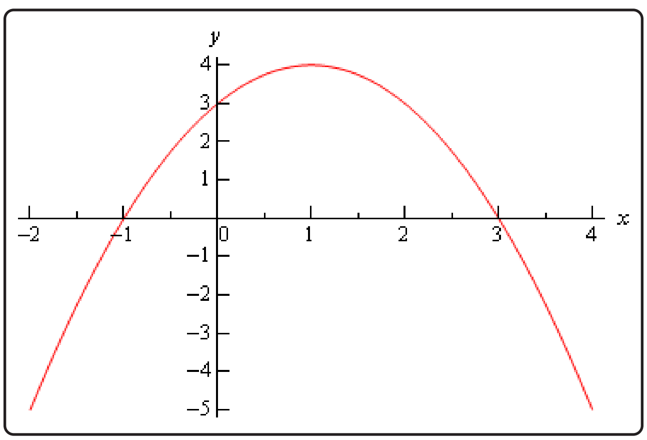
We can also determine which direction the parabola opens from the sign of a. If a is positive the parabola opens up and if a is negative the parabola opens down. In our case the parabola opens down.

Now, because the vertex is above the x-axis and the parabola opens down we know that we'll have x-intercepts (*i.e.* values of x for which we'll have $f(x) = 0$) on this graph. So, we'll solve the following.

$$\begin{aligned} -x^2 + 2x + 3 &= 0 \\ x^2 - 2x - 3 &= 0 \\ (x-3)(x+1) &= 0 \end{aligned}$$

So, we will have x-intercepts at $x = -1$ and $x = 3$. Notice that to make our life easier in the solution process we multiplied everything by -1 to get the coefficient of the x^2 positive. This made the factoring easier.

Here's a sketch of this parabola.



Example 4

Graph $f(y) = y^2 - 6y + 5$.

Solution

Most people come out of an Algebra class capable of dealing with functions in the form $y = f(x)$. However, many functions that you will have to deal with in a Calculus class are in the form $x = f(y)$ and can only be easily worked with in that form. So, you need to get used to working with functions in this form.

The nice thing about these kinds of function is that if you can deal with functions in the form $y = f(x)$ then you can deal with functions in the form $x = f(y)$ even if you aren't that familiar with them.

Let's first consider the equation.

$$y = x^2 - 6x + 5$$

This is a parabola that opens up and has a vertex of $(3, -4)$, as we know from our work in the previous example.

For our function we have essentially the same equation except the x and y's are switched around. In other words, we have a parabola in the form,

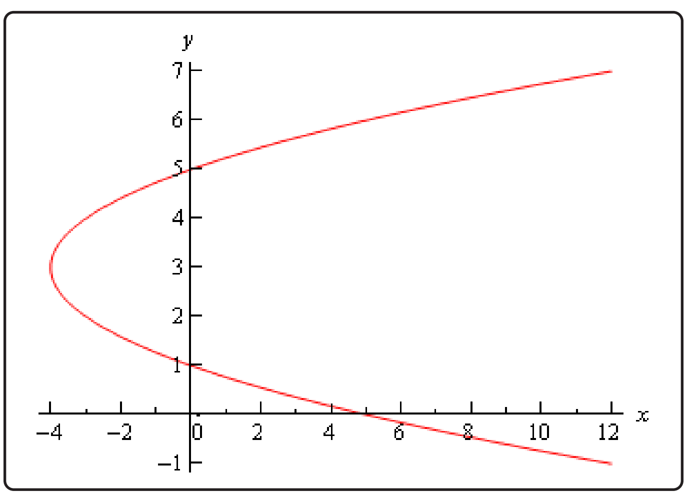
$$x = ay^2 + by + c$$

This is the general form of this kind of parabola and this will be a parabola that opens left or right depending on the sign of a. The y-coordinate of the vertex is given by $y = -\frac{b}{2a}$ and we find the x-coordinate by plugging this into the equation. So, you can see that this is very similar to the type of parabola that you're already used to dealing with.

Now, let's get back to the example. Our function is a parabola that opens to the right (a is positive) and has a vertex at $(-4, 3)$. The vertex is to the left of the y-axis and opens to the right so we'll need the y-intercepts (*i.e.* values of y for which we'll have $f(y) = 0$)). We find these just like we found x-intercepts in the previous problem.

$$\begin{aligned} y^2 - 6y + 5 &= 0 \\ (y-5)(y-1) &= 0 \end{aligned}$$

So, our parabola will have y-intercepts at $y = 1$ and $y = 5$. Here's a sketch of the graph.



Example 5

Graph $x^2 + 2x + y^2 - 8y + 8 = 0$.

Solution

To determine just what kind of graph we've got here we need to complete the square on both the x and the y.

$$\begin{aligned} x^2 + 2x + y^2 - 8y + 8 &= 0 \\ x^2 + 2x + 1 - 1 + y^2 - 8y + 16 - 16 + 8 &= 0 \\ (x+1)^2 + (y-4)^2 &= 9 \end{aligned}$$

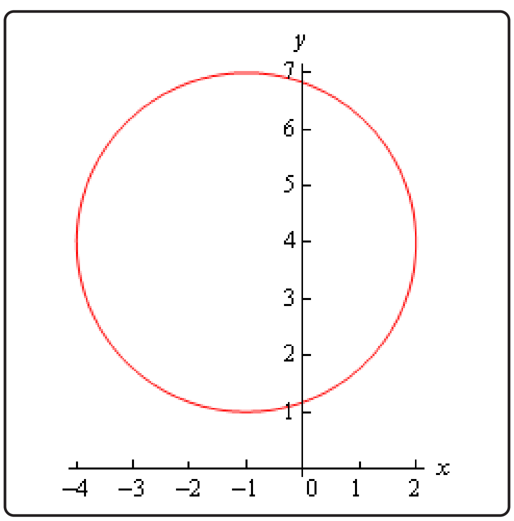
Recall that to complete the square we take the half of the coefficient of the x (or the y), square this and then add and subtract it to the equation.

Upon doing this we see that we have a circle and it's now written in standard form.

$$(x-h)^2 + (y-k)^2 = r^2$$

When circles are in this form we can easily identify the center (h, k) and radius r. Once we have these we can graph the circle simply by starting at the center and moving right, left, up and down by r to get the rightmost, leftmost, top most and bottom most points respectively.

Our circle has a center at $(-1, 4)$ and a radius of 3. Here's a sketch of this circle.



Example 6

Graph $\frac{(x-2)^2}{9} + 4(y+2)^2 = 1$.

Solution

This is an ellipse. The standard form of the ellipse is

$$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$$

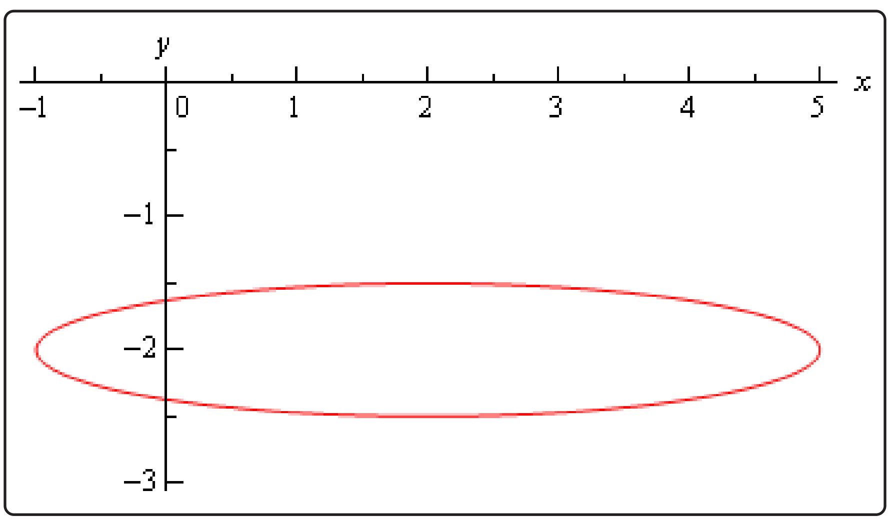
This is an ellipse with center (h, k) and the right most and left most points are a distance of a away from the center and the top most and bottom most points are a distance of b away from the center.

The ellipse for this problem has center $(2, -2)$ and has $a = 3$ and $b = \frac{1}{2}$. Note that to get the b we're really rewriting the equation as,

$$\frac{(x-2)^2}{9} + \frac{(y+2)^2}{1/4} = 1$$

to get it into standard from.

Here's a sketch of the ellipse.



Example 7

Graph $\dfrac{(x+1)^2}{9} - \dfrac{(y-2)^2}{4} = 1$.

Solution

This is a hyperbola. There are actually two standard forms for a hyperbola. Here are the basics for each form.

Form	$\dfrac{(x-h)^2}{a^2} - \dfrac{(y-k)^2}{b^2} = 1$	$\dfrac{(y-k)^2}{b^2} - \dfrac{(x-h)^2}{a^2} = 1$
Center	(h,k)	(h,k)
Opens	Opens right and left	Opens up and down
Vertices	a units right and left of center	b units up and down from center
Slope of Asymptotes	$\pm\dfrac{b}{a}$	$\pm\dfrac{b}{a}$

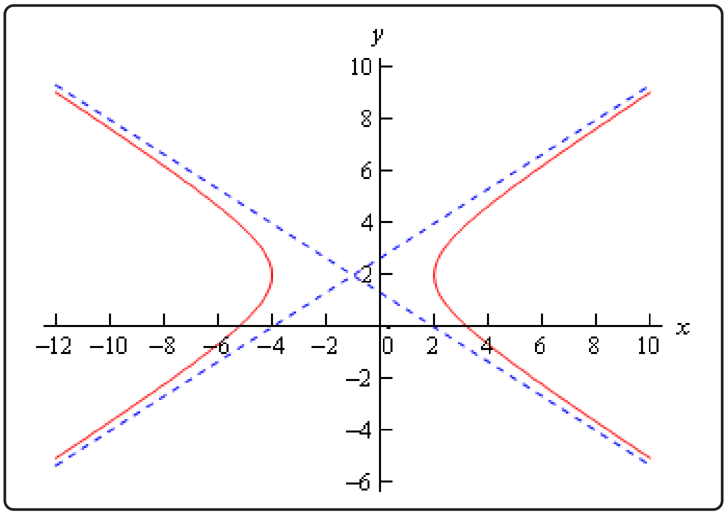
So, what does all this mean? First, notice that one of the terms is positive and the other is negative. This will determine which direction the two parts of the hyperbola open. If the x term is positive the hyperbola opens left and right. Likewise, if the y term is positive the parabola opens up and down.

Both have the same "center". Note that hyperbolas don't really have a center in the sense that circles and ellipses have centers. The center is the starting point in graphing a hyperbola. It tells us how to get to the vertices and how to get the asymptotes set up.

The asymptotes of a hyperbola are two lines that intersect at the center and have the slopes listed above. As you move farther out from the center the graph will get closer and closer to the asymptotes.

For the equation listed here the hyperbola will open left and right. Its center is $(-1,2)$. The two vertices are $(-4,2)$ and $(2,2)$. The asymptotes will have slopes $\pm\frac{2}{3}$.

Here is a sketch of this hyperbola. Note that the asymptotes are denoted by the two dashed lines.



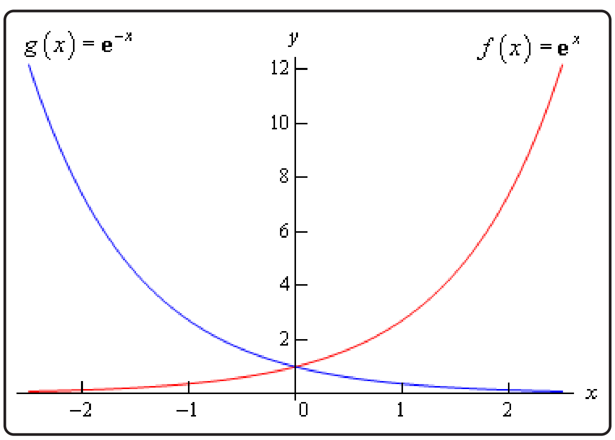
Example 8

Graph $f(x) = \mathbf{e}^x$ and $g(x) = \mathbf{e}^{-x}$.

Solution

There really isn't a lot to this problem other than making sure that both of these exponentials are graphed somewhere.

These will both show up with some regularity in later sections and their behavior as x goes to both plus and minus infinity will be needed and from this graph we can clearly see this behavior.

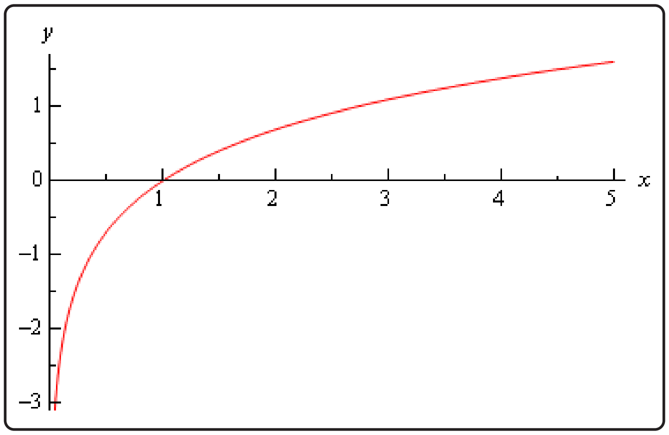


Example 9

Graph $f(x) = \ln(x)$.

Solution

This has already been graphed once in this review, but this puts it here with all the other "important" graphs.

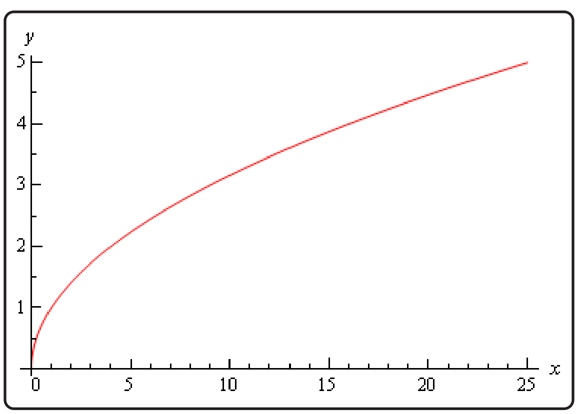


Example 10

Graph $y = \sqrt{x}$.

Solution

This one is fairly simple, we just need to make sure that we can graph it when need be.



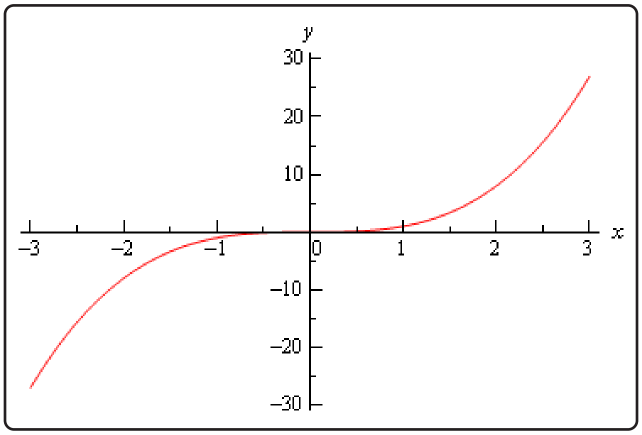
Remember that the domain of the square root function is $x \geq 0$.

Example 11

Graph $y = x^3$.

Solution

Again, there really isn't much to this other than to make sure it's been graphed somewhere so we can say we've done it.

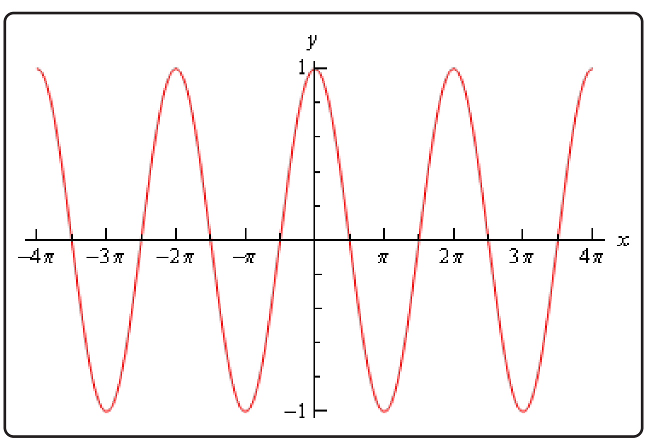


Example 12

Graph $y = \cos(x)$.

Solution

There really isn't a whole lot to this one. Here's the graph for $-4\pi \le x \le 4\pi$.



Let's also note here that we can put all values of x into cosine (which won't be the case for most of the trig functions) and so the domain is all real numbers. Also note that

$$-1 \leq \cos(x) \leq 1$$

It is important to notice that cosine will never be larger than 1 or smaller than -1. This will be useful on occasion in a calculus class. In general we can say that

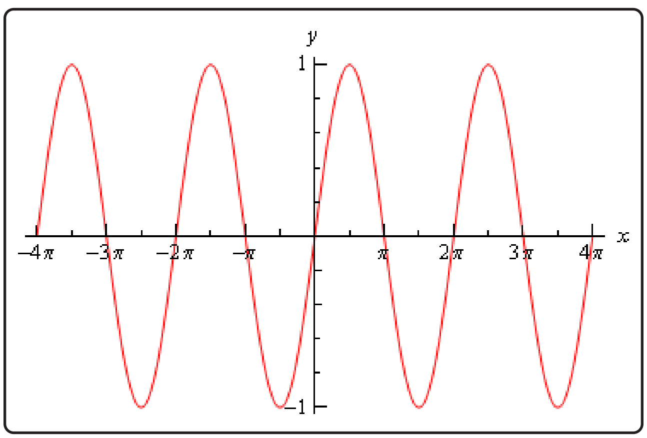
$$-R \leq R\cos(\omega x) \leq R$$

Example 13

Graph $y = \sin(x)$.

Solution

As with the previous problem there really isn't a lot to do other than graph it. Here is the graph for $-4\pi \leq x \leq 4\pi$.



From this graph we can see that sine has the same range that cosine does. In general

$$-R \leq R\sin(\omega x) \leq R$$

As with cosine, sine itself will never be larger than 1 and never smaller than -1. Also the domain of sine is all real numbers.

Example 14

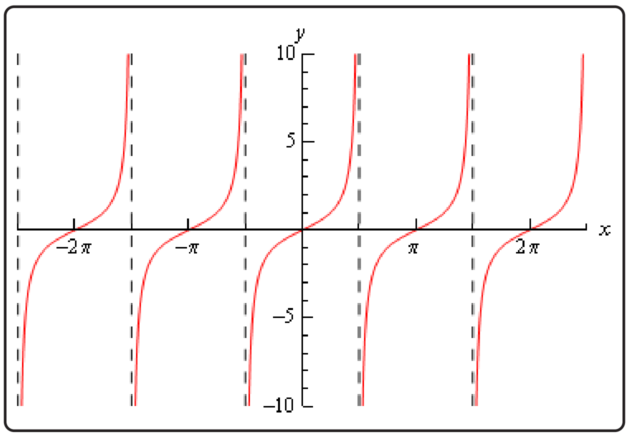
Graph $y = \tan(x)$.

Solution

In the case of tangent we have to be careful when plugging x's in since tangent doesn't exist wherever cosine is zero (remember that $\tan x = \frac{\sin x}{\cos x}$). Tangent will not exist at

$$x = \cdots, -\frac{5\pi}{2}, -\frac{3\pi}{2}, -\frac{\pi}{2}, \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots$$

and the graph will have asymptotes at these points. Here is the graph of tangent on the range $-\frac{5\pi}{2} < x < \frac{5\pi}{2}$.



Example 15

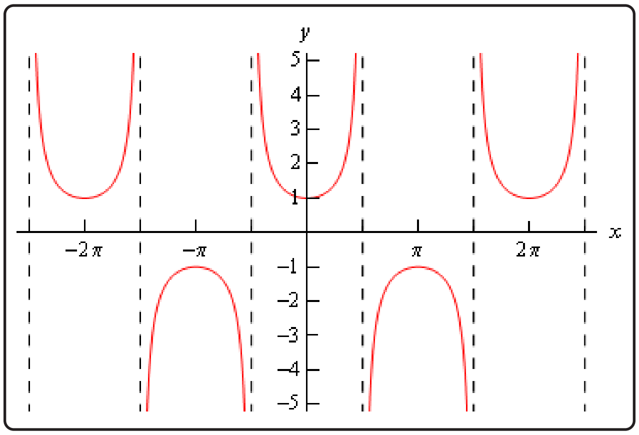
Graph $y = \sec(x)$.

Solution

As with tangent we will have to avoid x's for which cosine is zero (remember that $\sec x = \frac{1}{\cos x}$). Secant will not exist at

$$x = \cdots, -\frac{5\pi}{2}, -\frac{3\pi}{2}, -\frac{\pi}{2}, \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots$$

and the graph will have asymptotes at these points. Here is the graph of secant on the range $-\frac{5\pi}{2} < x < \frac{5\pi}{2}$.



Notice that the graph is always greater than 1 or less than -1. This should not be terribly surprising. Recall that

$$-1 \le \cos(x) \le 1$$

So, one divided by something less than one will be greater than 1. Also, $^1/_{\pm 1} = \pm 1$ and so we get the following ranges for secant.

$$\sec(\omega x) \ge 1 \qquad \text{and} \qquad \sec(\omega x) \le -1$$

Note that we did not graph cotangent or cosecant here. However, they are similar to the graphs of tangent and secant and you should be able to do quick sketches of them given the work above if needed.

Finally, note that we did not cover any of the basic transformations that are often used in graphing functions here. The practice problems for this section have quite a few problems designed to help you remember them. If you know the basic transformations it often makes graphing a much simpler process so if you are not comfortable with them you should work through the practice problems for this section.

2 Limits

The topic that we will be examining in this chapter is that of Limits. This is the first of three major topics that we will be covering in this course. While we will be spending the least amount of time on limits in comparison to the other two topics limits are very important in the study of Calculus. We will be seeing limits in a variety of places once we move out of this chapter. In particular we will see that limits are part of the formal definition of the other two major topics.

In this chapter we will discuss just what a limit tells us about a function as well as how they can be used to get the rate of change of a function as well as the slope of the line tangent to the graph of a function (although we'll be seeing other, easier, ways of doing these later). We will investigate limit properties as well as how a variety of techniques to employ when attempting to compute a limit. We will also look at limits whose "value" is infinity and how to compute limits at infinity.

In addition, we'll introduce the concept of continuity and how continuity is used in the Intermediate Value Theorem. The Intermediate Value Theorem is an important idea that has a variety of "real world" applications including showing that a function has a root (*i.e.* is equal to zero) in some interval.

Finally, we'll close out the chapter with the formal/precise definition of the Limit, sometimes called the delta-epsilon definition.

2.1 Tangent Lines and Rates of Change

In this section we are going to take a look at two fairly important problems in the study of calculus. There are two reasons for looking at these problems now.

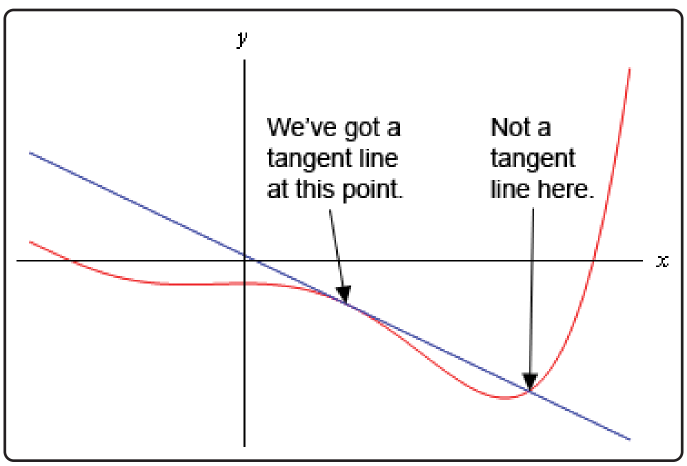
First, both of these problems will lead us into the study of limits, which is the topic of this chapter after all. Looking at these problems here will allow us to start to understand just what a limit is and what it can tell us about a function.

Secondly, the rate of change problem that we're going to be looking at is one of the most important concepts that we'll encounter in the second chapter of this course. In fact, it's probably one of the most important concepts that we'll encounter in the whole course. So, looking at it now will get us to start thinking about it from the very beginning.

Tangent Lines

The first problem that we're going to take a look at is the tangent line problem. Before getting into this problem it would probably be best to define a tangent line.

A tangent line to the function $f(x)$ at the point $x = a$ is a line that just touches the graph of the function at the point in question and is "parallel" (in some way) to the graph at that point. Take a look at the graph below.



In this graph the line is a tangent line at the indicated point because it just touches the graph at that point and is also "parallel" to the graph at that point. Likewise, at the second point shown, the line does just touch the graph at that point, but it is not "parallel" to the graph at that point and so it's not a tangent line to the graph at that point.

At the second point shown (the point where the line isn't a tangent line) we will sometimes call the line a **secant line**.

We've used the word parallel a couple of times now and we should probably be a little careful with it. In general, we will think of a line and a graph as being parallel at a point if they are both moving in the same direction at that point. So, in the first point above the graph and the line are moving in the same direction and so we will say they are parallel at that point. At the second point, on the other hand, the line and the graph are not moving in the same direction so they aren't parallel at that point.

Okay, now that we've gotten the definition of a tangent line out of the way let's move on to the tangent line problem. That's probably best done with an example.

Example 1

Find the tangent line to $f(x) = 15 - 2x^2$ at $x = 1$.

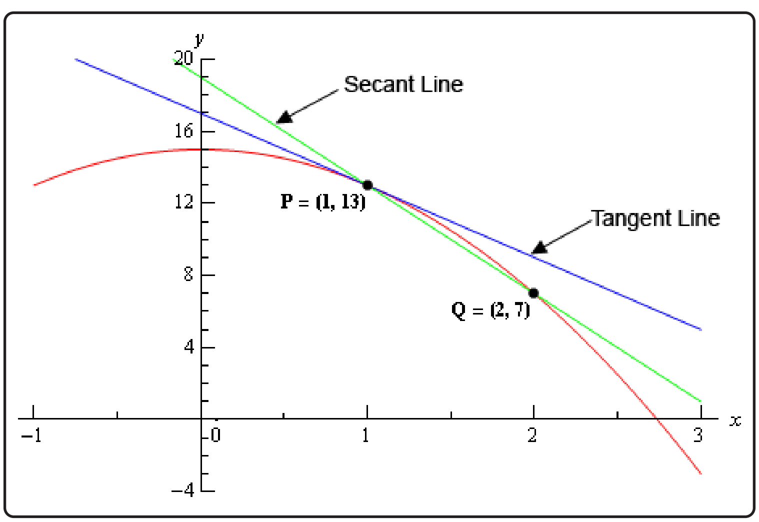
Solution

We know from algebra that to find the equation of a line we need either two points on the line or a single point on the line and the slope of the line. Since we know that we are after a tangent line we do have a point that is on the line. The tangent line and the graph of the function must touch at x = 1 so the point $(1, f(1)) = (1, 13)$ must be on the line.

Now we reach the problem. This is all that we know about the tangent line. In order to find the tangent line we need either a second point or the slope of the tangent line. Since the only reason for needing a second point is to allow us to find the slope of the tangent line let's just concentrate on seeing if we can determine the slope of the tangent line.

At this point in time all that we're going to be able to do is to get an estimate for the slope of the tangent line, but if we do it correctly we should be able to get an estimate that is in fact the actual slope of the tangent line. We'll do this by starting with the point that we're after, let's call it $P = (1, 13)$. We will then pick another point that lies on the graph of the function, let's call that point $Q = (x, f(x))$.

For the sake of argument let's take $x = 2$ and so the second point will be $Q = (2, 7)$. Below is a graph of the function, the tangent line and the secant line that connects P and Q.

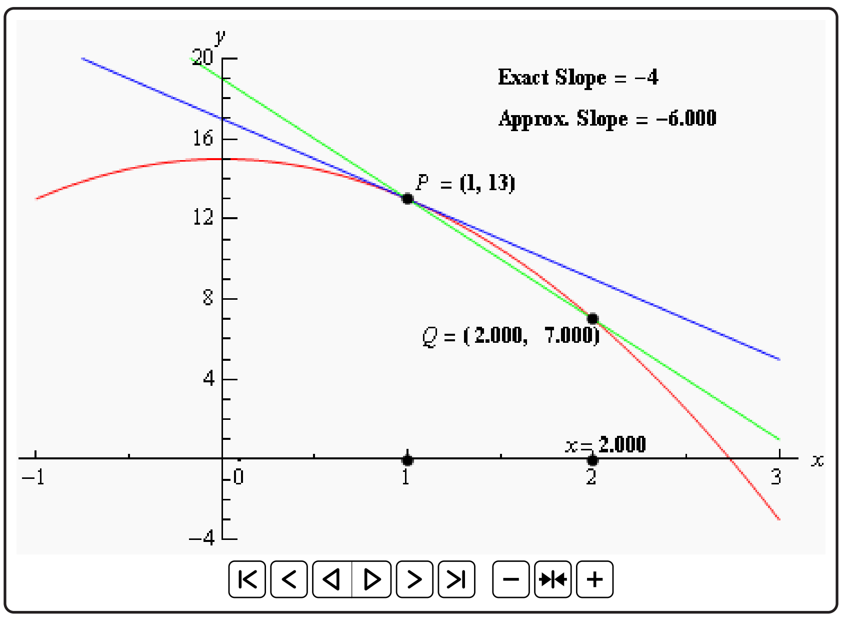


We can see from this graph that the secant and tangent lines are somewhat similar and so the slope of the secant line should be somewhat close to the actual slope of the tangent line. So, as an estimate of the slope of the tangent line we can use the slope of the secant line, let's call it m_{PQ}, which is,

$$m_{PQ} = \frac{f(2) - f(1)}{2 - 1} = \frac{7 - 13}{1} = -6$$

Now, if we weren't too interested in accuracy we could say this is good enough and use this as an estimate of the slope of the tangent line. However, we would like an estimate that is at least somewhat close the actual value. So, to get a better estimate we can take an x that is closer to $x = 1$ and redo the work above to get a new estimate on the slope. We could then take a third value of x even closer yet and get an even better estimate.

In other words, as we take Q closer and closer to P the slope of the secant line connecting Q and P should be getting closer and closer to the slope of the tangent line. If you are viewing this on the web, the image below shows this process.



As you can see (animation won't work on all pdf viewers unfortunately) as we moved Q in closer and closer to P the secant lines does start to look more and more like the tangent line and so the approximate slopes (*i.e.* the slopes of the secant lines) are getting closer and closer to the exact slope. Also, do not worry about how I got the exact or approximate slopes. We'll be computing the approximate slopes shortly and we'll be able to compute the exact slope in a few sections.

In this figure we only looked at Q's that were to the right of P, but we could have just as easily used Q's that were to the left of P and we would have received the same results. In fact, we should always take a look at Q's that are on both sides of P. In this case the same thing is happening on both sides of P. However, we will eventually see that doesn't have to happen. Therefore, we should always take a look at what is happening on both sides of the point in question when doing this kind of process.

So, let's see if we can come up with the approximate slopes we showed above, and hence an estimation of the slope of the tangent line. In order to simplify the process a little let's get a formula for the slope of the line between P and Q, m_{PQ}, that will work for any x that we choose to work with. We can get a formula by finding the slope between P and Q using the "general" form of $Q = (x, f(x))$.

$$m_{PQ} = \frac{f(x) - f(1)}{x - 1} = \frac{15 - 2x^2 - 13}{x - 1} = \frac{2 - 2x^2}{x - 1}$$

Now, let's pick some values of x getting closer and closer to $x = 1$, plug in and get some slopes.

x	m_{PQ}	x	m_{PQ}
2	-6	0	-2
1.5	-5	0.5	-3
1.1	-4.2	0.9	-3.8
1.01	-4.02	0.99	-3.98
1.001	-4.002	0.999	-3.998
1.0001	-4.0002	0.9999	-3.9998

So, if we take x's to the right of 1 and move them in very close to 1 it appears the slope of the secant lines is approaching -4. Likewise, if we take x's to the left of 1 and move them in very close to 1 the slope of the secant lines again appears to be approaching -4.

Based on this evidence it seems that the slopes of the secant lines are approaching -4 as we move in towards $x = 1$, so we will estimate that the slope of the tangent line is also -4. As noted above, this is the correct value and we will be able to prove this eventually.

Now, the equation of the line that goes through $(a, f(a))$ is given by

$$y = f(a) + m(x - a)$$

Therefore, the equation of the tangent line to $f(x) = 15 - 2x^2$ at $x = 1$ is

$$y = 13 - 4(x - 1) = -4x + 17$$

There are a couple of important points to note about our work above. First, we looked at points that were on both sides of $x = 1$. In this kind of process it is important to never assume that what is happening on one side of a point will also be happening on the other side as well. We should always look at what is happening on both sides of the point. In this example we could sketch a graph and from that guess that what is happening on one side will also be happening on the other, but we will usually not have the graphs in front of us or be able to easily get them.

Next, notice that when we say we're going to move in close to the point in question we do mean that we're going to move in very close and we also used more than just a couple of points. We should never try to determine a trend based on a couple of points that aren't really all that close to the point in question.

The next thing to notice is really a warning more than anything. The values of m_{PQ} in this example were fairly "nice" and it was pretty clear what value they were approaching after a couple of computations. In most cases this will not be the case. Most values will be far "messier" and you'll often need quite a few computations to be able to get an estimate. You should always use at least four points, on each side to get the estimate. Two points is never sufficient to get a good estimate and three points will also often not be sufficient to get a good estimate. Generally, you keeping picking points closer and closer to the point you are looking at until the change in the value between two successive points is getting very small.

Last, we were after something that was happening at $x = 1$ and we couldn't actually plug $x = 1$ into our formula for the slope. Despite this limitation we were able to determine some information about what was happening at $x = 1$ simply by looking at what was happening around $x = 1$. This is more important than you might at first realize and we will be discussing this point in detail in later sections.

Before moving on let's do a quick review of just what we did in the above example. We wanted the tangent line to $f(x)$ at a point $x = a$. First, we know that the point $P = (a, f(a))$ will be on the tangent line. Next, we'll take a second point that is on the graph of the function, call it $Q = (x, f(x))$ and compute the slope of the line connecting P and Q as follows,

$$m_{PQ} = \frac{f(x) - f(a)}{x - a}$$

We then take values of x that get closer and closer to $x = a$ (making sure to look at x's on both sides of $x = a$ and use this list of values to estimate the slope of the tangent line, m.

The tangent line will then be,

$$y = f(a) + m(x - a)$$

Rates of Change

The next problem that we need to look at is the rate of change problem. As mentioned earlier, this will turn out to be one of the most important concepts that we will look at throughout this course.

Here we are going to consider a function, $f(x)$, that represents some quantity that varies as x varies. For instance, maybe $f(x)$ represents the amount of water in a holding tank after x minutes. Or maybe $f(x)$ is the distance traveled by a car after x hours. In both of these example we used x

to represent time. Of course x doesn't have to represent time, but it makes for examples that are easy to visualize.

What we want to do here is determine just how fast $f(x)$ is changing at some point, say $x = a$. This is called the **instantaneous rate of change** or sometimes just **rate of change** of $f(x)$ at $x = a$.

As with the tangent line problem all that we're going to be able to do at this point is to estimate the rate of change. So, let's continue with the examples above and think of $f(x)$ as something that is changing in time and x being the time measurement. Again, x doesn't have to represent time but it will make the explanation a little easier. While we can't compute the instantaneous rate of change at this point we can find the average rate of change.

To compute the average rate of change of $f(x)$ at $x = a$ all we need to do is to choose another point, say x, and then the average rate of change will be,

$$\begin{aligned} A.R.C. &= \frac{\text{change in } f(x)}{\text{change in } x} \\ &= \frac{f(x) - f(a)}{x - a} \end{aligned}$$

Then to estimate the instantaneous rate of change at $x = a$ all we need to do is to choose values of x getting closer and closer to $x = a$ (don't forget to choose them on both sides of $x = a$) and compute values of $A.R.C.$ We can then estimate the instantaneous rate of change from that.

Let's take a look at an example.

Example 2

Suppose that the amount of air in a balloon after t hours is given by

$$V(t) = t^3 - 6t^2 + 35$$

Estimate the instantaneous rate of change of the volume after 5 hours.

Solution

Okay. The first thing that we need to do is get a formula for the average rate of change of the volume. In this case this is,

$$A.R.C. = \frac{V(t) - V(5)}{t - 5} = \frac{t^3 - 6t^2 + 35 - 10}{t - 5} = \frac{t^3 - 6t^2 + 25}{t - 5}$$

To estimate the instantaneous rate of change of the volume at $t = 5$ we just need to pick values of t that are getting closer and closer to $t = 5$. Here is a table of values of t and the average rate of change for those values.

t	$A.R.C.$	t	$A.R.C.$
6	25.0	4	7.0
5.5	19.75	4.5	10.75
5.1	15.91	4.9	14.11
5.01	15.0901	4.99	14.9101
5.001	15.009001	4.999	14.991001
5.0001	15.00090001	4.9999	14.99910001

So, from this table it looks like the average rate of change is approaching 15 and so we can estimate that the instantaneous rate of change is 15 at this point.

So, just what does this tell us about the volume at $t = 5$? Let's put some units on the answer from above. This might help us to see what is happening to the volume at this point. Let's suppose that the units on the volume were in cm^3. The units on the rate of change (both average and instantaneous) are then cm^3/hr.

We have estimated that at $t = 5$ the volume is changing at a rate of 15 cm^3/hr. This means that at $t = 5$ the volume is changing in such a way that, if the rate were constant, then an hour later there would be 15 cm^3 more air in the balloon than there was at $t = 5$.

We do need to be careful here however. In reality there probably won't be 15 cm^3 more air in the balloon after an hour. The rate at which the volume is changing is generally not constant so we can't make any real determination as to what the volume will be in another hour. What we can say is that the volume is increasing, since the instantaneous rate of change is positive, and if we had rates of change for other values of t we could compare the numbers and see if the rate of change is faster or slower at the other points.

For instance, at $t = 4$ the instantaneous rate of change is 0 cm^3/hr and at $t = 3$ the instantaneous rate of change is -9 cm^3/hr. We'll leave it to you to check these rates of change. In fact, that would be a good exercise to see if you can build a table of values that will support our claims on these rates of change.

Anyway, back to the example. At $t = 4$ the rate of change is zero and so at this point in time the volume is not changing at all. That doesn't mean that it will not change in the future. It just means that exactly at $t = 4$ the volume isn't changing. Likewise, at $t = 3$ the volume is decreasing since the rate of change at that point is negative. We can also say that, regardless of the increasing/decreasing aspects of the rate of change, the volume of the balloon is changing faster at $t = 5$ than it is at $t = 3$ since 15 is larger than 9.

We will be talking a lot more about rates of change when we get into the next chapter.

Velocity Problem

Let's briefly look at the velocity problem. Many calculus books will treat this as its own problem. We however, like to think of this as a special case of the rate of change problem. In the velocity problem we are given a position function of an object, $f(t)$, that gives the position of an object at time t. Then to compute the instantaneous velocity of the object we just need to recall that the velocity is nothing more than the rate at which the position is changing.

In other words, to estimate the instantaneous velocity we would first compute the average velocity,

$$\begin{aligned} A.V. &= \frac{\text{change in position}}{\text{time traveled}} \\ &= \frac{f(t) - f(a)}{t - a} \end{aligned}$$

and then take values of t closer and closer to $t = a$ and use these values to estimate the instantaneous velocity.

Change of Notation

There is one last thing that we need to do in this section before we move on. The main point of this section was to introduce us to a couple of key concepts and ideas that we will see throughout the first portion of this course as well as get us started down the path towards limits.

Before we move into limits officially let's go back and do a little work that will relate both (or all three if you include velocity as a separate problem) problems to a more general concept.

First, notice that whether we wanted the tangent line, instantaneous rate of change, or instantaneous velocity each of these came down to using exactly the same formula. Namely,

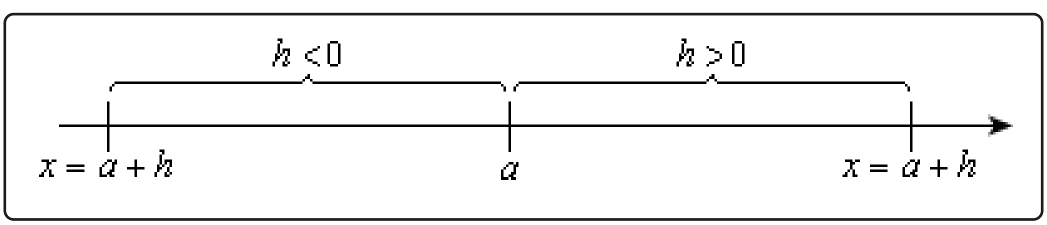
$$\frac{f(x) - f(a)}{x - a} \tag{2.1}$$

This should suggest that all three of these problems are then really the same problem. In fact this is the case as we will see in the next chapter. We are really working the same problem in each of these cases the only difference is the interpretation of the results.

In preparation for the next section where we will discuss this in much more detail we need to do a quick change of notation. It's easier to do here since we've already invested a fair amount of time into these problems.

In all of these problems we wanted to determine what was happening at $x = a$. To do this we chose another value of x and plugged into Equation 2.1. For what we were doing here that is probably most intuitive way of doing it. However, when we start looking at these problems as a single problem Equation 2.1 will not be the best formula to work with.

What we'll do instead is to first determine how far from $x = a$ we want to move and then define our new point based on that decision. So, if we want to move a distance of h from $x = a$ the new point would be $x = a + h$. This is shown in the sketch below.



As we saw in our work above it is important to take values of x that are both sides of $x = a$. This way of choosing new value of x will do this for us as we can see in the sketch above. If $h > 0$ we will get value of x that are to the right of $x = a$ and if $h < 0$ we will get values of x that are to the left of $x = a$ and both are given by $x = a + h$.

Now, with this new way of getting a second value of x Equation 2.1 will become,

$$\frac{f(x)-f(a)}{x-a} = \frac{f(a+h)-f(a)}{a+h-a} = \frac{f(a+h)-f(a)}{h}$$

Now, this is for a specific value of x, *i.e.* $x = a$ and we'll rarely be looking at these at specific values of x. So, we take the final step in the above equation and replace the a with x to get,

$$\frac{f(x+h)-f(x)}{h}$$

This gives us a formula for a general value of x and on the surface it might seem that this is going to be an overly complicated way of dealing with this stuff. However, as we will see it will often be easier to deal with this form than it will be to deal with the original form, Equation 2.1.

2.2 The Limit

In the previous section we looked at a couple of problems and in both problems we had a function (slope in the tangent problem case and average rate of change in the rate of change problem) and we wanted to know how that function was behaving at some point $x = a$. At this stage of the game we no longer care where the functions came from and we no longer care if we're going to see them down the road again or not. All that we need to know or worry about is that we've got these functions and we want to know something about them.

To answer the questions in the last section we choose values of x that got closer and closer to $x = a$ and we plugged these into the function. We also made sure that we looked at values of x that were on both the left and the right of $x = a$. Once we did this we looked at our table of function values and saw what the function values were approaching as x got closer and closer to $x = a$ and used this to guess the value that we were after.

This process is called **taking a limit** and we have some notation for this. The limit notation for the two problems from the last section is,

$$\lim_{x \to 1} \frac{2 - 2x^2}{x - 1} = -4 \qquad\qquad \lim_{t \to 5} \frac{t^3 - 6t^2 + 25}{t - 5} = 15$$

In this notation we will note that we always give the function that we're working with and we also give the value of x (or t) that we are moving in towards.

In this section we are going to take an intuitive approach to limits and try to get a feel for what they are and what they can tell us about a function. With that goal in mind we are not going to get into how we actually compute limits yet. We will instead rely on what we did in the previous section as well as another approach to guess the value of the limits.

Both approaches that we are going to use in this section are designed to help us understand just what limits are. In general, we don't typically use the methods in this section to compute limits and in many cases can be very difficult to use to even estimate the value of a limit and/or will give the wrong value on occasion. We will look at actually computing limits in a couple of sections.

Let's first start off with the following "definition" of a limit.

Definition

We say that the limit of $f(x)$ is L as x approaches a and write this as

$$\lim_{x \to a} f(x) = L$$

provided we can make $f(x)$ as close to L as we want for all x sufficiently close to a, from both sides, without actually letting x be a.

This is not the exact, precise definition of a limit. If you would like to see the more precise and mathematical definition of a limit you should check out the The Definition of a Limit section at the end of this chapter. The definition given above is more of a "working" definition. This definition helps us to get an idea of just what limits are and what they can tell us about functions.

So just what does this definition mean? Well let's suppose that we know that the limit does in fact exist. According to our "working" definition we can then decide how close to L that we'd like to make $f(x)$. For sake of argument let's suppose that we want to make $f(x)$ no more than 0.001 away from L. This means that we want one of the following

$$\begin{array}{ll} f(x) - L < 0.001 & \text{if } f(x) \text{ is larger than L} \\ L - f(x) < 0.001 & \text{if } f(x) \text{ is smaller than L} \end{array}$$

Now according to the "working" definition this means that if we get x sufficiently close to a we can make one of the above true. However, it actually says a little more. It says that somewhere out there in the world is a value of x, say X, so that for all x's that are closer to a than X then one of the above statements will be true.

This is a fairly important idea. There are many functions out there in the world that we can make as close to L for specific values of x that are close to a, but there will be other values of x closer to a that give functions values that are nowhere near close to L. In order for a limit to exist once we get $f(x)$ as close to L as we want for some x then it will need to stay in that close to L (or get closer) for all values of x that are closer to a. We'll see an example of this later in this section.

In somewhat simpler terms the definition says that as x gets closer and closer to $x = a$ (from both sides of course...) then $f(x)$ **must** be getting closer and closer to L. Or, as we move in towards $x = a$ then $f(x)$ **must** be moving in towards L.

It is important to note once again that we must look at values of x that are on both sides of $x = a$. We should also note that we are not allowed to use $x = a$ in the definition. We will often use the information that limits give us to get some information about what is going on right at $x = a$, but the limit itself is not concerned with what is actually going on at $x = a$. The limit is only concerned with what is going on around the point $x = a$. This is an important concept about limits that we need to keep in mind.

An alternative notation that we will occasionally use in denoting limits is

$$f(x) \to L \quad \text{as} \quad x \to a$$

How do we use this definition to help us estimate limits? We do exactly what we did in the previous section. We take x's on both sides of $x = a$ that move in closer and closer to a and we plug these into our function. We then look to see if we can determine what number the function values are moving in towards and use this as our estimate.

Let's work an example.

Example 1

Estimate the value of the following limit.

$$\lim_{x \to 2} \frac{x^2 + 4x - 12}{x^2 - 2x}$$

Solution

Notice that we did say estimate the value of the limit. Again, we are not going to directly compute limits in this section. The point of this section is to give us a better idea of how limits work and what they can tell us about the function.

So, with that in mind we are going to work this in pretty much the same way that we did in the last section. We will choose values of x that get closer and closer to $x = 2$ and plug these values into the function. Doing this gives the following table of values.

x	$f(x)$	x	$f(x)$
2.5	3.4	1.5	5.0
2.1	3.857142857	1.9	4.157894737
2.01	3.985074627	1.99	4.015075377
2.001	3.998500750	1.999	4.001500750
2.0001	3.999850007	1.9999	4.000150008
2.00001	3.999985000	1.99999	4.000015000

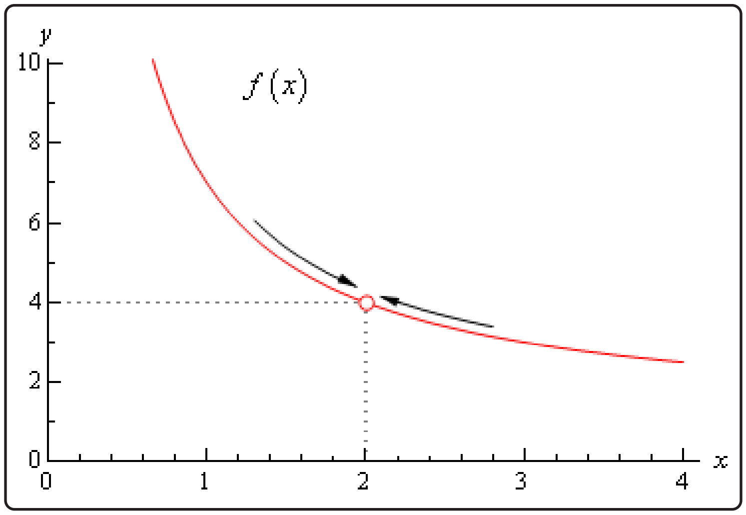
Note that we made sure and picked values of x that were on both sides of $x = 2$ and that we moved in very close to $x = 2$ to make sure that any trends that we might be seeing are in fact correct.

Also notice that we can't actually plug in $x = 2$ into the function as this would give us a division by zero error. This is not a problem since the limit doesn't care what is happening at the point in question.

From this table it appears that the function is going to 4 as x approaches 2, so

$$\lim_{x \to 2} \frac{x^2 + 4x - 12}{x^2 - 2x} = 4$$

Let's think a little bit more about what's going on here. Let's graph the function from the last example. The graph of the function in the range of x's that were interested in is shown below.



First, notice that there is a rather large open dot at $x = 2$. This is there to remind us that the function (and hence the graph) doesn't exist at $x = 2$.

As we were plugging in values of x into the function we are in effect moving along the graph in towards the point as $x = 2$. This is shown in the graph by the two arrows on the graph that are moving in towards the point.

When we are computing limits the question that we are really asking is what y value is our graph approaching as we move in towards $x = a$ on our graph. We are **NOT** asking what y value the graph takes at the point in question. In other words, we are asking what the graph is doing **around** the point $x = a$. In our case we can see that as x moves in towards 2 (from both sides) the function is approaching $y = 4$ even though the function itself doesn't even exist at $x = 2$. Therefore, we can say that the limit is in fact 4.

So, what have we learned about limits? Limits are asking what the function is doing **around** $x = a$ and are **not** concerned with what the function is actually doing at $x = a$. This is a good thing as many of the functions that we'll be looking at won't even exist at $x = a$ as we saw in our last example.

Let's work another example to drive this point home.

Example 2

Estimate the value of the following limit.

$$\lim_{x \to 2} g(x) \qquad \text{where,} \qquad g(x) = \begin{cases} \dfrac{x^2 + 4x - 12}{x^2 - 2x} & \text{if } x \neq 2 \\ 6 & \text{if } x = 2 \end{cases}$$

Solution

The first thing to note here is that this is exactly the same function as the first example with the exception that we've now given it a value for $x = 2$. So, let's first note that

$$g(2) = 6$$

As far as estimating the value of this limit goes, nothing has changed in comparison to the first example. We could build up a table of values as we did in the first example or we could take a quick look at the graph of the function. Either method will give us the value of the limit.

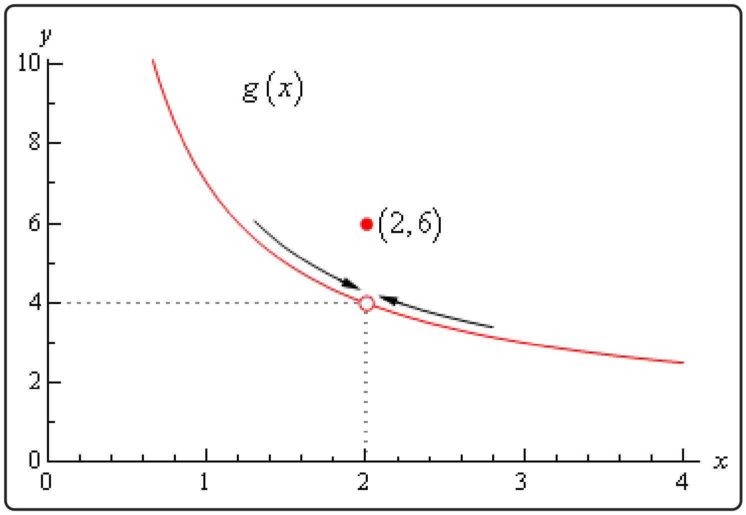
Let's first take a look at a table of values and see what that tells us. Notice that the presence of the value for the function at $x = 2$ will not change our choices for x. We only choose values of x that are getting closer to $x = 2$ but we never take $x = 2$. In other words, the table of values that we used in the first example will be exactly the same table that we'll use here. So, since we've already got it down once there is no reason to redo it here.

From this table it is again clear that the limit is,

$$\lim_{x \to 2} g(x) = 4$$

The limit is **NOT** 6! Remember from the discussion after the first example that limits do not care what the function is actually doing at the point in question. Limits are only concerned with what is going on **around** the point. Since the only thing about the function that we actually changed was its behavior at $x = 2$ this will not change the limit.

Let's also take a quick look at this function's graph to see if this says the same thing.



Again, we can see that as we move in towards $x = 2$ on our graph the function is still approaching a y value of 4. Remember that we are only asking what the function is doing **around** $x = 2$ and we don't care what the function is actually doing at $x = 2$. The graph then also supports the conclusion that the limit is,

$$\lim_{x \to 2} g(x) = 4$$

Let's make the point one more time just to make sure we've got it. Limits are **not** concerned with what is going on at $x = a$. Limits are only concerned with what is going on **around** $x = a$. We keep saying this, but it is a very important concept about limits that we must always keep in mind. So, we will take every opportunity to remind ourselves of this idea.

Since limits aren't concerned with what is actually happening at $x = a$ we will, on occasion, see situations like the previous example where the limit at a point and the function value at a point are different. This won't always happen of course. There are times where the function value and the limit at a point are the same and we will eventually see some examples of those. It is important however, to not get excited about things when the function and the limit do not take the same value at a point. It happens sometimes so we will need to be able to deal with those cases when they arise.

Let's take a look another example to try and beat this idea into the ground.

Example 3

Estimate the value of the following limit.

$$\lim_{\theta \to 0} \frac{1 - \cos(\theta)}{\theta}$$

Solution

First don't get excited about the θ in function. It's just a letter, just like x is a letter! It's a Greek letter, but it's a letter and you will be asked to deal with Greek letters on occasion so it's a good idea to start getting used to them at this point.

Now, also notice that if we plug in $\theta = 0$ that we will get division by zero and so the function doesn't exist at this point. Actually, we get $0/0$ at this point, but because of the division by zero this function does not exist at $\theta = 0$.

So, as we did in the first example let's get a table of values and see what if we can guess what value the function is heading in towards.

θ	$f(\theta)$	θ	$f(\theta)$
1	0.45969769	-1	-0.45969769
0.1	0.04995835	-0.1	-0.04995835
0.01	0.00499996	-0.01	-0.00499996
0.001	0.00049999	-0.001	-0.00049999

Okay, it looks like the function is moving in towards a value of zero as θ moves in towards 0, from both sides of course.

Therefore, the we will guess that the limit has the value,

$$\lim_{\theta \to 0} \frac{1 - \cos(\theta)}{\theta} = 0$$

So, once again, the limit had a value even though the function didn't exist at the point we were interested in.

It's now time to work a couple of more examples that will lead us into the next idea about limits that we're going to want to discuss.

Example 4

Estimate the value of the following limit.

$$\lim_{t \to 0} \cos\left(\frac{\pi}{t}\right)$$

Solution

Let's build up a table of values and see what's going on with our function in this case.

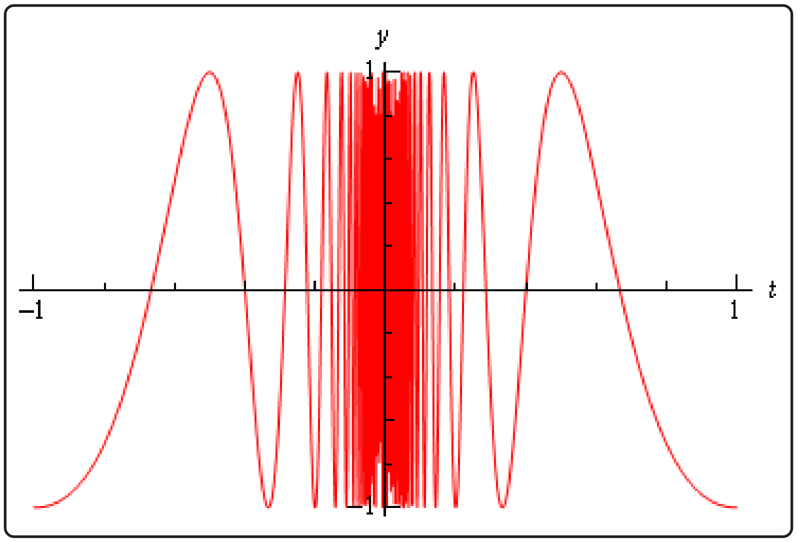
t	$f(t)$	t	$f(t)$
1	-1	-1	-1
0.1	1	-0.1	1
0.01	1	-0.01	1
0.001	1	-0.001	1

Now, if we were to guess the limit from this table we would guess that the limit is 1. However, if we did make this guess we would be wrong. Consider any of the following function evaluations.

$$f\left(\frac{1}{2001}\right) = -1 \qquad f\left(\frac{2}{2001}\right) = 0 \qquad f\left(\frac{4}{4001}\right) = \frac{\sqrt{2}}{2}$$

In all three of these function evaluations we evaluated the function at a number that is less than 0.001 and got three totally different numbers. Recall that the definition of the limit that we're working with requires that the function be approaching a single value (our guess) as t gets closer and closer to the point in question. It doesn't say that only some of the function values must be getting closer to the guess. It says that all the function values must be getting closer and closer to our guess.

To see what's happening here a graph of the function would be convenient.



From this graph we can see that as we move in towards $t = 0$ the function starts oscillating wildly and in fact the oscillations increases in speed the closer to $t = 0$ that we get. Recall from our definition of the limit that in order for a limit to exist the function must be settling down in towards a single value as we get closer to the point in question.

This function clearly does not settle in towards a single number and so this limit **does not exist**!

This last example points out the drawback of just picking values of the variable and using a table of function values to estimate the value of a limit. The values of the variable that we chose in the previous example were valid and in fact were probably values that many would have picked. In fact, they were exactly the same values we used in the problem before this one and they worked in that problem!

When using a table of values there will always be the possibility that we aren't choosing the correct values and that we will guess incorrectly for our limit. This is something that we should always keep in mind when doing this to guess the value of limits. In fact, this is such a problem that after this section we will never use a table of values to guess the value of a limit again.

This last example also has shown us that limits do not have to exist. To that point we've only seen limits that existed, but that just doesn't always have to be the case.

Let's take a look at one more example in this section.

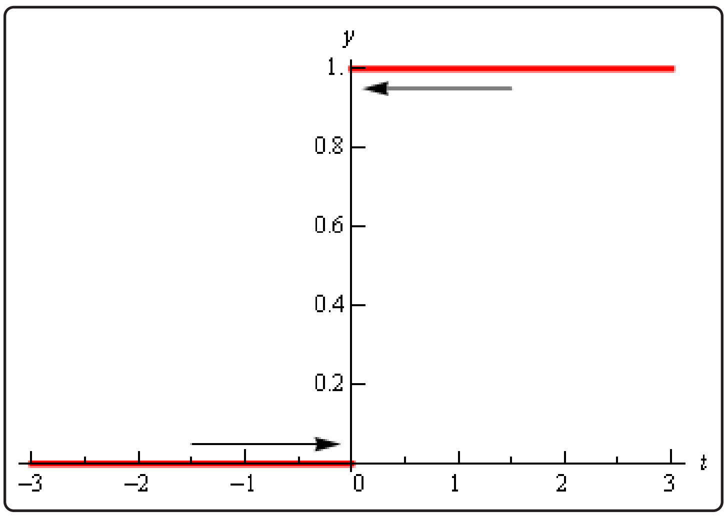
Example 5

Estimate the value of the following limit.

$$\lim_{t \to 0} H(t) \qquad \text{where,} \qquad H(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0 \end{cases}$$

Solution

This function is often called either the **Heaviside** or **step** function. We could use a table of values to estimate the limit, but it's probably just as quick in this case to use the graph so let's do that. Below is the graph of this function.



We can see from the graph that if we approach $t = 0$ from the right side the function is moving in towards a y value of 1. Well actually it's just staying at 1, but in the terminology that we've been using in this section it's moving in towards 1...

Also, if we move in towards $t = 0$ from the left the function is moving in towards a y value of 0.

According to our definition of the limit the function needs to move in towards a single value as we move in towards $t = a$ (from both sides). This isn't happening in this case and so in this example we will also say that the limit doesn't exist.

Note that the limit in this example is a little different from the previous example. In the previous example the function did not settle down to a single number as we moved in towards $t = 0$. In this example however, the function does settle down to a single number as $t = 0$ on either side. The problem is that the number is different on each side of $t = 0$. This is an idea that we'll look at in a little more detail in the next section.

Let's summarize what we (hopefully) learned in this section. In the first three examples we saw that limits do not care what the function is actually doing at the point in question. They only are concerned with what is happening around the point. In fact, we can have limits at $x = a$ even if the function itself does not exist at that point. Likewise, even if a function exists at a point there is no reason (at this point) to think that the limit will have the same value as the function at that point. Sometimes the limit and the function will have the same value at a point and other times they won't have the same value.

Next, in the third and fourth examples we saw the main reason for not using a table of values to guess the value of a limit. In those examples we used exactly the same set of values, however they only worked in one of the examples. Using tables of values to guess the value of limits is simply not a good way to get the value of a limit. This is the only section in which we will do this. Tables of values should always be your last choice in finding values of limits.

The last two examples showed us that not all limits will in fact exist. We should not get locked into the idea that limits will always exist. In most calculus courses we work with limits that almost always exist and so it's easy to start thinking that limits always exist. Limits don't always exist and so don't get into the habit of assuming that they will.

Finally, we saw in the fourth example that the only way to deal with the limit was to graph the function. Sometimes this is the only way, however this example also illustrated the drawback of using graphs. In order to use a graph to guess the value of the limit you need to be able to actually sketch the graph. For many functions this is not that easy to do.

There is another drawback in using graphs. Even if you have the graph it's only going to be useful if the y value is approaching an integer. If the y value is approaching say $\frac{-15}{123}$ there is no way that you're going to be able to guess that value from the graph and we are usually going to want exact values for our limits.

So, while graphs of functions can, on occasion, make your life easier in guessing values of limits they are again probably not the best way to get values of limits. They are only going to be useful if you can get your hands on it and the value of the limit is a "nice" number.

The natural question then is why did we even talk about using tables and/or graphs to estimate limits if they aren't the best way. There were a couple of reasons.

First, they can help us get a better understanding of what limits are and what they can tell us. If we don't do at least a couple of limits in this way we might not get all that good of an idea on just what limits are.

The second reason for doing limits in this way is to point out their drawback so that we aren't tempted to use them all the time!

We will eventually talk about how we really do limits. However, there is one more topic that we need to discuss before doing that. Since this section has already gone on for a while we will talk about this in the next section.

2.3 One-Sided Limits

In the final two examples in the previous section we saw two limits that did not exist. However, the reason for each of the limits not existing was different for each of the examples.

We saw that

$$\lim_{t \to 0} \cos\left(\frac{\pi}{t}\right)$$

did not exist because the function did not settle down to a single value as t approached $t = 0$. The closer to $t = 0$ we moved the more wildly the function oscillated and in order for a limit to exist the function must settle down to a single value.

However, we saw that

$$\lim_{t \to 0} H(t) \qquad \text{where,} \qquad H(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0 \end{cases}$$

did not exist not because the function didn't settle down to a single number as we moved in towards $t = 0$, but instead because it settled into two different numbers depending on which side of $t = 0$ we were on.

In this case the function was a very well-behaved function, unlike the first function. The only problem was that, as we approached $t = 0$, the function was moving in towards different numbers on each side. We would like a way to differentiate between these two examples.

We do this with **one-sided limits**. As the name implies, with one-sided limits we will only be looking at one side of the point in question. Here are the definitions for the two one sided limits.

Right-handed limit

We say

$$\lim_{x \to a^+} f(x) = L$$

provided we can make $f(x)$ as close to L as we want for all x sufficiently close to a with $x > a$ without actually letting x be a.

Left-handed limit

We say

$$\lim_{x \to a^-} f(x) = L$$

provided we can make $f(x)$ as close to L as we want for all x sufficiently close to a with $x < a$ without actually letting x be a.

Note that the change in notation is very minor and in fact might be missed if you aren't paying attention. The only difference is the bit that is under the "lim" part of the limit. For the right-handed limit we now have $x \to a^+$ (note the "+") which means that we know will only look at $x > a$. Likewise, for the left-handed limit we have $x \to a^-$ (note the "-") which means that we will only be looking at $x < a$.

Also, note that as with the "normal" limit (*i.e.* the limits from the previous section) we still need the function to settle down to a single number in order for the limit to exist. The only difference this time is that the function only needs to settle down to a single number on either the right side of $x = a$ or the left side of $x = a$ depending on the one-sided limit we're dealing with.

So, when we are looking at limits it's now important to pay very close attention to see whether we are doing a normal limit or one of the one-sided limits. Let's now take a look at the some of the problems from the last section and look at one-sided limits instead of the normal limit.

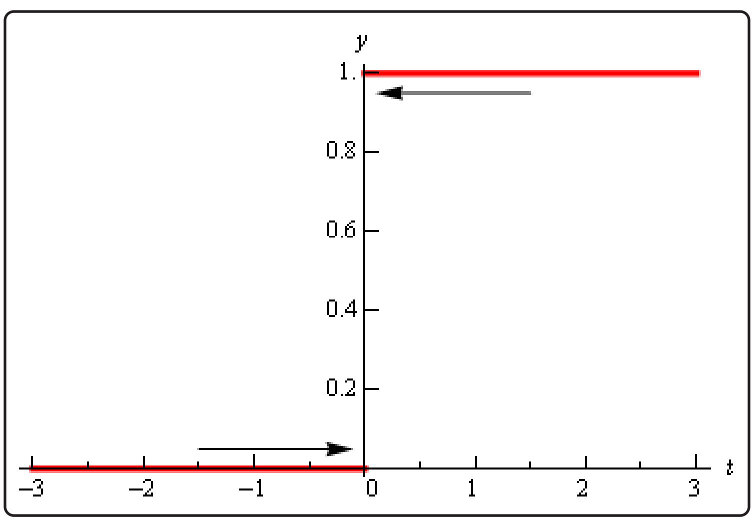
Example 1

Estimate the value of the following limits.

$$\lim_{t \to 0^+} H(t) \quad \text{and} \quad \lim_{t \to 0^-} H(t) \qquad \text{where} \qquad H(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0 \end{cases}$$

Solution

To remind us what this function looks like here's the graph.



So, we can see that if we stay to the right of $t = 0$ (*i.e.* $t > 0$) then the function is moving in towards a value of 1 as we get closer and closer to $t = 0$, but staying to the right. We can therefore say that the right-handed limit is,

$$\lim_{t \to 0^+} H(t) = 1$$

Likewise, if we stay to the left of $t = 0$ (*i.e* $t < 0$) the function is moving in towards a value of 0 as we get closer and closer to $t = 0$, but staying to the left. Therefore, the left-handed limit is,

$$\lim_{t \to 0^-} H(t) = 0$$

In this example we do get one-sided limits even though the normal limit itself doesn't exist.

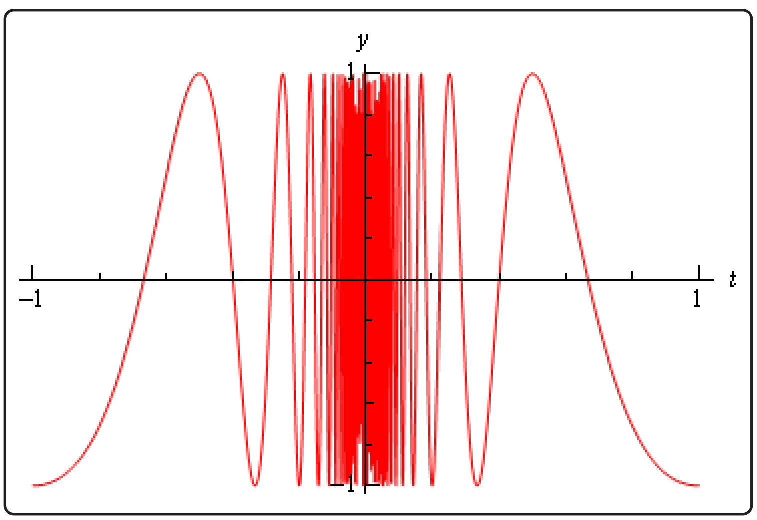
Example 2

Estimate the value of the following limits.

$$\lim_{t \to 0^+} \cos\left(\frac{\pi}{t}\right) \qquad \lim_{t \to 0^-} \cos\left(\frac{\pi}{t}\right)$$

Solution

From the graph of this function shown below,



we can see that both of the one-sided limits suffer the same problem that the normal limit did in the previous section. The function does not settle down to a single number on either side of $t = 0$. Therefore, neither the left-handed nor the right-handed limit will exist in this case.

So, one-sided limits don't have to exist just as normal limits aren't guaranteed to exist.

Let's take a look at another example from the previous section.

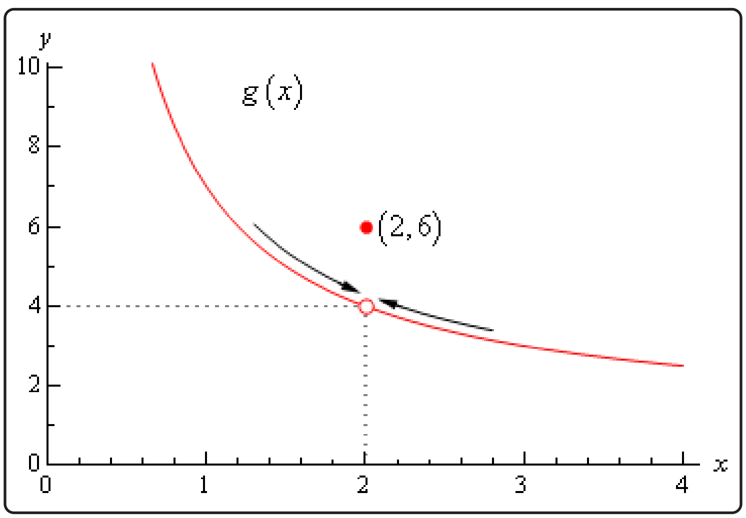
Example 3

Estimate the value of the following limits.

$$\lim_{x \to 2^+} g(x) \qquad \text{and} \qquad \lim_{x \to 2^-} g(x) \qquad \text{where} \qquad g(x) = \begin{cases} \dfrac{x^2 + 4x - 12}{x^2 - 2x} & \text{if } x \neq 2 \\ 6 & \text{if } x = 2 \end{cases}$$

Solution

So, as we've done with the previous two examples, let's remind ourselves of the graph of this function.



In this case regardless of which side of $x = 2$ we are on the function is always approaching a value of 4 and so we get,

$$\lim_{x \to 2^+} g(x) = 4 \qquad \qquad \lim_{x \to 2^-} g(x) = 4$$

Note that one-sided limits do not care about what's happening at the point any more than normal limits do. They are still only concerned with what is going on around the point. The only real difference between one-sided limits and normal limits is the range of x's that we look at when determining the value of the limit.

Now let's take a look at the first and last example in this section to get a very nice fact about the relationship between one-sided limits and normal limits. In the last example the one-sided limits as well as the normal limit existed and all three had a value of 4. In the first example the two one-sided limits both existed, but did not have the same value and the normal limit did not exist.

The relationship between one-sided limits and normal limits can be summarized by the following fact.

Fact

Given a function $f(x)$ if,

$$\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x) = L$$

then the normal limit will exist and

$$\lim_{x \to a} f(x) = L$$

Likewise, if

$$\lim_{x \to a} f(x) = L$$

then,

$$\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x) = L$$

This fact can be turned around to also say that if the two one-sided limits have different values, *i.e.*,

$$\lim_{x \to a^+} f(x) \neq \lim_{x \to a^-} f(x)$$

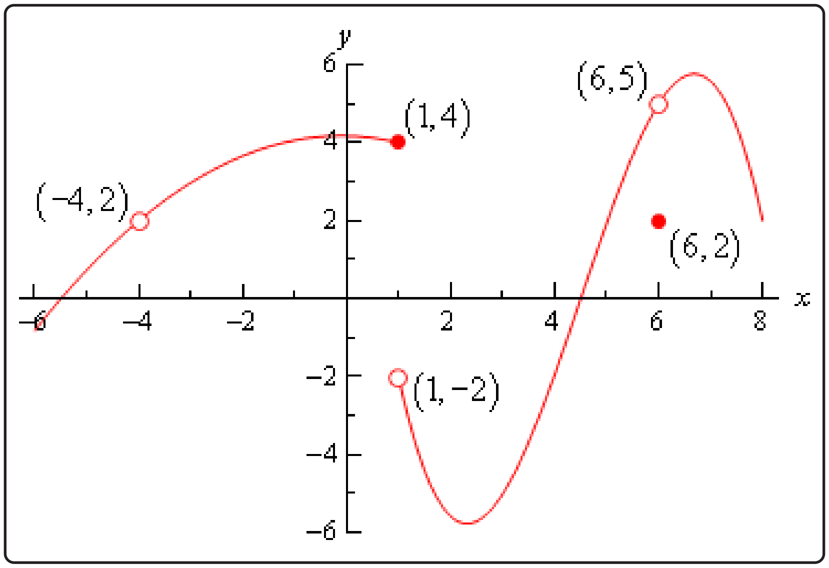
then the normal limit will not exist.

This should make some sense. If the normal limit did exist then by the fact the two one-sided limits would have to exist and have the same value by the above fact. So, if the two one-sided limits have different values (or don't even exist) then the normal limit simply can't exist.

Let's take a look at one more example to make sure that we've got all the ideas about limits down that we've looked at in the last couple of sections.

Example 4

Given the following graph,



compute each of the following.

(a) $f(-4)$ **(b)** $\lim\limits_{x \to -4^-} f(x)$ **(c)** $\lim\limits_{x \to -4^+} f(x)$ **(d)** $\lim\limits_{x \to -4} f(x)$

(e) $f(1)$ **(f)** $\lim\limits_{x \to 1^-} f(x)$ **(g)** $\lim\limits_{x \to 1^+} f(x)$ **(h)** $\lim\limits_{x \to 1} f(x)$

(i) $f(6)$ **(j)** $\lim\limits_{x \to 6^-} f(x)$ **(k)** $\lim\limits_{x \to 6^+} f(x)$ **(l)** $\lim\limits_{x \to 6} f(x)$

Solution

(a) $f(-4)$ doesn't exist. There is no closed dot for this value of x and so the function doesn't exist at this point.

(b) $\lim\limits_{x \to -4^-} f(x) = 2$ The function is approaching a value of 2 as x moves in towards -4 from the left.

(c) $\lim\limits_{x \to -4^+} f(x) = 2$ The function is approaching a value of 2 as x moves in towards -4 from the right.

(d) $\lim\limits_{x \to -4} f(x) = 2$ We can do this one of two ways. Either we can use the fact here and notice that the two one-sided limits are the same and so the normal limit must exist and have the same value as the one-sided limits or just get the answer from the graph.

Also recall that a limit can exist at a point even if the function doesn't exist at that point.

(e) $f(1) = 4$. The function will take on the y value where the closed dot is.

(f) $\lim\limits_{x \to 1^-} f(x) = 4$ The function is approaching a value of 4 as x moves in towards 1 from the left.

(g) $\lim\limits_{x \to 1^+} f(x) = -2$ The function is approaching a value of -2 as x moves in towards 1 from the right. Remember that the limit does NOT care about what the function is actually doing at the point, it only cares about what the function is doing around the point. In this case, always staying to the right of $x = 1$, the function is approaching a value of -2 and so the limit is -2. The limit is not 4, as that is value of the function at the point and again the limit doesn't care about that!

(h) $\lim\limits_{x \to 1} f(x)$ doesn't exist. The two one-sided limits both exist, however they are different and so the normal limit doesn't exist.

(i) $f(6) = 2$. The function will take on the y value where the closed dot is.

(j) $\lim\limits_{x \to 6^-} f(x) = 5$ The function is approaching a value of 5 as x moves in towards 6 from the left.

(k) $\lim\limits_{x \to 6^+} f(x) = 5$ The function is approaching a value of 5 as x moves in towards 6 from the right.

(l) $\lim\limits_{x \to 6} f(x) = 5$ Again, we can use either the graph or the fact to get this. Also, once more remember that the limit doesn't care what is happening at the point and so it's possible for the limit to have a different value than the function at a point. When dealing with limits we've always got to remember that limits simply do not care about what the function is doing at the point in question. Limits are only concerned with what the function is doing around the point.

Hopefully over the last couple of sections you've gotten an idea on how limits work and what they can tell us about functions. Some of these ideas will be important in later sections so it's important that you have a good grasp on them.

2.4 Limit Properties

The time has almost come for us to actually compute some limits. However, before we do that we will need some properties of limits that will make our life somewhat easier. So, let's take a look at those first. The proof of some of these properties can be found in the Proof of Various Limit Properties section of the Extras appendix.

Properties

First, we will assume that $\lim\limits_{x\to a} f(x)$ and $\lim\limits_{x\to a} g(x)$ exist and that c is any constant. Then,

1. $\lim\limits_{x\to a} [cf(x)] = c \lim\limits_{x\to a} f(x)$
   In other words, we can "factor" a multiplicative constant out of a limit.

2. $\lim\limits_{x\to a} [f(x) \pm g(x)] = \lim\limits_{x\to a} f(x) \pm \lim\limits_{x\to a} g(x)$
   So, to take the limit of a sum or difference all we need to do is take the limit of the individual parts and then put them back together with the appropriate sign. This is also not limited to two functions. This fact will work no matter how many functions we've got separated by "+" or "-".

3. $\lim\limits_{x\to a} [f(x)\, g(x)] = \lim\limits_{x\to a} f(x) \quad \lim\limits_{x\to a} g(x)$
   We take the limits of products in the same way that we can take the limit of sums or differences. Just take the limit of the pieces and then put them back together. Also, as with sums or differences, this fact is not limited to just two functions.

4. $\lim\limits_{x\to a} \left[\dfrac{f(x)}{g(x)}\right] = \dfrac{\lim\limits_{x\to a} f(x)}{\lim\limits_{x\to a} g(x)},$ provided $\lim\limits_{x\to a} g(x) \neq 0$
   As noted in the statement we only need to worry about the limit in the denominator being zero when we do the limit of a quotient. If it were zero we would end up with a division by zero error and we need to avoid that.

5. $\lim\limits_{x\to a} [f(x)]^n = \left[\lim\limits_{x\to a} f(x)\right]^n,$ where n is any real number
   In this property n can be any real number (positive, negative, integer, fraction, irrational, zero, *etc.*). In the case that n is an integer this rule can be thought of as an extended case of **3**.

   For example, consider the case of $n = 2$.

$$\begin{aligned}
\lim_{x\to a} [f(x)]^2 &= \lim_{x\to a} [f(x)\, f(x)] \\
&= \lim_{x\to a} f(x) \lim_{x\to a} f(x) && \text{using property 3} \\
&= \left[\lim_{x\to a} f(x)\right]^2
\end{aligned}$$

   The same can be done for any integer n.

6. $\lim\limits_{x \to a} \left[\sqrt[n]{f(x)}\right] = \sqrt[n]{\lim\limits_{x \to a} f(x)}$
   This is just a special case of the previous example.

$$\begin{aligned} \lim_{x \to a} \left[\sqrt[n]{f(x)}\right] &= \lim_{x \to a} \left[f(x)\right]^{\frac{1}{n}} \\ &= \left[\lim_{x \to a} f(x)\right]^{\frac{1}{n}} \\ &= \sqrt[n]{\lim_{x \to a} f(x)} \end{aligned}$$

7. $\lim\limits_{x \to a} c = c, \quad c$ is any real number
   In other words, the limit of a constant is just the constant. You should be able to convince yourself of this by drawing the graph of $f(x) = c$.

8. $\lim\limits_{x \to a} x = a$
   As with the last one you should be able to convince yourself of this by drawing the graph of $f(x) = x$.

9. $\lim\limits_{x \to a} x^n = a^n$
   This is really just a special case of property **5** using $f(x) = x$.

Note that all these properties also hold for the two one-sided limits as well we just didn't write them down with one sided limits to save on space.

Let's compute a limit or two using these properties. The next couple of examples will lead us to some truly useful facts about limits that we will use on a continual basis.

Example 1

Compute the value of the following limit.

$$\lim_{x \to -2} \left(3x^2 + 5x - 9\right)$$

Solution

This first time through we will use only the properties above to compute the limit.

First, we will use property **2** to break up the limit into three separate limits. We will then use property **1** to bring the constants out of the first two limits. Doing this gives us,

$$\begin{aligned} \lim_{x \to -2} \left(3x^2 + 5x - 9\right) &= \lim_{x \to -2} 3x^2 + \lim_{x \to -2} 5x - \lim_{x \to -2} 9 \\ &= 3 \lim_{x \to -2} x^2 + 5 \lim_{x \to -2} x - \lim_{x \to -2} 9 \end{aligned}$$

We can now use properties **7** through **9** to actually compute the limit.

$$\begin{aligned}\lim_{x\to -2}\left(3x^2+5x-9\right) &= 3\lim_{x\to -2}x^2+5\lim_{x\to -2}x-\lim_{x\to -2}9\\ &= 3(-2)^2+5\left(-2\right)-9\\ &= -7\end{aligned}$$

Now, let's notice that if we had defined

$$p\left(x\right)=3x^2+5x-9$$

then the proceeding example would have been,

$$\begin{aligned}\lim_{x\to -2}p\left(x\right) &= \lim_{x\to -2}\left(3x^2+5x-9\right)\\ &= 3(-2)^2+5\left(-2\right)-9\\ &= -7\\ &= p\left(-2\right)\end{aligned}$$

In other words, in this case we see that the limit is the same value that we'd get by just evaluating the function at the point in question. This seems to violate one of the main concepts about limits that we've seen to this point.

In the previous two sections we made a big deal about the fact that limits do not care about what is happening at the point in question. They only care about what is happening around the point. So how does the previous example fit into this since it appears to violate this main idea about limits?

Despite appearances the limit still doesn't care about what the function is doing at $x = -2$. In this case the function that we've got is simply "nice enough" so that what is happening around the point is exactly the same as what is happening at the point. Eventually we will formalize up just what is meant by "nice enough". At this point let's not worry too much about what "nice enough" is. Let's just take advantage of the fact that some functions will be "nice enough", whatever that means.

The function in the last example was a polynomial. It turns out that all polynomials are "nice enough" so that what is happening around the point is exactly the same as what is happening at the point. This leads to the following fact.

Fact

If $p(x)$ is a polynomial then,

$$\lim_{x\to a}p\left(x\right)=p\left(a\right)$$

By the end of this section we will generalize this out considerably to most of the functions that we'll be seeing throughout this course.

Let's take a look at another example.

Example 2

Evaluate the following limit.

$$\lim_{z \to 1} \frac{6 - 3z + 10z^2}{-2z^4 + 7z^3 + 1}$$

Solution

First notice that we can use property **4** to write the limit as,

$$\lim_{z \to 1} \frac{6 - 3z + 10z^2}{-2z^4 + 7z^3 + 1} = \frac{\lim\limits_{z \to 1} 6 - 3z + 10z^2}{\lim\limits_{z \to 1} -2z^4 + 7z^3 + 1}$$

Well, actually we should be a little careful. We can do that provided the limit of the denominator isn't zero. As we will see however, it isn't in this case so we're okay.

Now, both the numerator and denominator are polynomials so we can use the fact above to compute the limits of the numerator and the denominator and hence the limit itself.

$$\begin{aligned} \lim_{z \to 1} \frac{6 - 3z + 10z^2}{-2z^4 + 7z^3 + 1} &= \frac{6 - 3\,(1) + 10(1)^2}{-2(1)^4 + 7(1)^3 + 1} \\ &= \frac{13}{6} \end{aligned}$$

Notice that the limit of the denominator wasn't zero and so our use of property **4** was legitimate.

In the previous example, as with polynomials, all we really did was evaluate the function at the point in question. So, it appears that there is a fairly large class of functions for which this can be done. Let's generalize the fact from above a little.

Fact

Provided $f(x)$ is "nice enough" we have,

$$\lim_{x \to a} f(x) = f(a) \qquad \lim_{x \to a^-} f(x) = f(a) \qquad \lim_{x \to a^+} f(x) = f(a)$$

Again, we will formalize up just what we mean by “nice enough” eventually. At this point all we want to do is worry about which functions are “nice enough”. Some functions are “nice enough” for all x while others will only be “nice enough” for certain values of x. It will all depend on the function.

As noted in the statement, this fact also holds for the two one-sided limits as well as the normal limit.

Here is a list of some of the more common functions that are “nice enough”.

- Polynomials are nice enough for all x’s.
- If $f(x) = \dfrac{p(x)}{q(x)}$ then $f(x)$ will be nice enough provided both $p(x)$ and $q(x)$ are nice enough and if we don’t get division by zero at the point we’re evaluating at.
- $\cos(x)$, $\sin(x)$ are nice enough for all x’s
- $\sec(x)$, $\tan(x)$ are nice enough provided $x \neq \ldots, -\frac{5\pi}{2}, -\frac{3\pi}{2}, \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots$ In other words secant and tangent are nice enough everywhere cosine isn’t zero. To see why recall that these are both really rational functions and that cosine is in the denominator of both then go back up and look at the second bullet above.
- $\csc(x)$, $\cot(x)$ are nice enough provided $x \neq \ldots, -2\pi,\ -\pi,\ 0,\ \pi,\ 2\pi, \ldots$ In other words cosecant and cotangent are nice enough everywhere sine isn’t zero.
- $\sqrt[n]{x}$ is nice enough for all x if n is odd.
- $\sqrt[n]{x}$ is nice enough for $x \geq 0$ if n is even. Here we require $x \geq 0$ to avoid having to deal with complex values.
- a^x, $\mathbf{e}^x$ are nice enough for all x’s.
- $\log_b x$, $\ln x$ are nice enough for $x > 0$. Remember we can only plug positive numbers into logarithms and not zero or negative numbers.
- Any sum, difference or product of the above functions will also be nice enough. Quotients will be nice enough provided we don’t get division by zero upon evaluating the limit.

The last bullet is important. This means that for any combination of these functions all we need to do is evaluate the function at the point in question, making sure that none of the restrictions are violated. This means that we can now do a large number of limits.

Example 3

Evaluate the following limit.

$$\lim_{x \to 3} \left(-\sqrt[5]{x} + \frac{\mathbf{e}^x}{1 + \ln(x)} + \sin(x)\cos(x) \right)$$

Solution

This is a combination of several of the functions listed above and none of the restrictions are violated so all we need to do is plug in $x = 3$ into the function to get the limit.

$$\begin{aligned} \lim_{x \to 3} \left(-\sqrt[5]{x} + \frac{\mathbf{e}^x}{1 + \ln(x)} + \sin(x)\cos(x) \right) &= -\sqrt[5]{3} + \frac{\mathbf{e}^3}{1 + \ln(3)} + \sin(3)\cos(3) \\ &= 8.1854272743 \end{aligned}$$

Not a very pretty answer, but we can now do the limit.

2.5 Computing Limits

In the previous section we saw that there is a large class of functions that allows us to use

$$\lim_{x \to a} f(x) = f(a)$$

to compute limits. However, there are also many limits for which this won't work easily. The purpose of this section is to develop techniques for dealing with some of these limits that will not allow us to just use this fact.

Let's first go back and take a look at one of the first limits that we looked at and compute its exact value and verify our guess for the limit.

Example 1

Evaluate the following limit.

$$\lim_{x \to 2} \frac{x^2 + 4x - 12}{x^2 - 2x}$$

Solution

First let's notice that if we try to plug in $x = 2$ we get,

$$\lim_{x \to 2} \frac{x^2 + 4x - 12}{x^2 - 2x} = \frac{0}{0}$$

So, we can't just plug in $x = 2$ to evaluate the limit. So, we're going to have to do something else.

The first thing that we should always do when evaluating limits is to simplify the function as much as possible. In this case that means factoring both the numerator and denominator. Doing this gives,

$$\begin{aligned}\lim_{x \to 2} \frac{x^2 + 4x - 12}{x^2 - 2x} &= \lim_{x \to 2} \frac{(x-2)(x+6)}{x(x-2)} \\ &= \lim_{x \to 2} \frac{x+6}{x}\end{aligned}$$

So, upon factoring we saw that we could cancel an $x - 2$ from both the numerator and the denominator. Upon doing this we now have a new rational expression that we can plug $x = 2$ into because we lost the division by zero problem. Therefore, the limit is,

$$\lim_{x \to 2} \frac{x^2 + 4x - 12}{x^2 - 2x} = \lim_{x \to 2} \frac{x+6}{x} = \frac{8}{2} = 4$$

Note that this is in fact what we guessed the limit to be.

Before leaving this example let's discuss the fact that we couldn't plug $x = 2$ into our original limit but once we did the simplification we just plugged in $x = 2$ to get the answer. At first glance this may appear to be a contradiction.

In the original limit we couldn't plug in $x = 2$ because that gave us the $0/0$ situation that we couldn't do anything with. Upon doing the simplification we can note that,

$$\frac{x^2 + 4x - 12}{x^2 - 2x} = \frac{x + 6}{x} \qquad \text{provided } x \neq 2$$

In other words, the two equations give identical values except at $x = 2$ and because limits are only concerned with that is going on around the point $x = 2$ the limit of the two equations will be equal. More importantly, in the simplified version we get a "nice enough" equation and so what is happening around $x = 2$ is identical to what is happening at $x = 2$.

We can therefore take the limit of the simplified version simply by plugging in $x = 2$ even though we couldn't plug $x = 2$ into the original equation and the value of the limit of the simplified equation will be the same as the limit of the original equation.

On a side note, the $0/0$ we initially got in the previous example is called an **indeterminate form**. This means that we don't really know what it will be until we do some more work. Typically, zero in the denominator means it's undefined. However, that will only be true if the numerator isn't also zero. Also, zero in the numerator usually means that the fraction is zero, unless the denominator is also zero. Likewise, anything divided by itself is 1, unless we're talking about zero.

So, there are really three competing "rules" here and it's not clear which one will win out. It's also possible that none of them will win out and we will get something totally different from undefined, zero, or one. We might, for instance, get a value of 4 out of this, to pick a number completely at random.

When simply evaluating an equation $0/0$ is undefined. However, in taking the limit, if we get $0/0$ we can get a variety of answers and the only way to know which on is correct is to actually compute the limit.

There are many more kinds of indeterminate forms and we will be discussing indeterminate forms at length in the next chapter.

Let's take a look at a couple of more examples.

Example 2

Evaluate the following limit.

$$\lim_{h\to 0}\frac{2(-3+h)^2-18}{h}$$

Solution

In this case we also get $0/0$ and factoring is not really an option. However, there is still some simplification that we can do.

$$\begin{aligned}\lim_{h\to 0}\frac{2(-3+h)^2-18}{h} &= \lim_{h\to 0}\frac{2\left(9-6h+h^2\right)-18}{h}\\ &= \lim_{h\to 0}\frac{18-12h+2h^2-18}{h}\\ &= \lim_{h\to 0}\frac{-12h+2h^2}{h}\end{aligned}$$

So, upon multiplying out the first term we get a little cancellation and now notice that we can factor an h out of both terms in the numerator which will cancel against the h in the denominator and the division by zero problem goes away and we can then evaluate the limit.

$$\begin{aligned}\lim_{h\to 0}\frac{2(-3+h)^2-18}{h} &= \lim_{h\to 0}\frac{-12h+2h^2}{h}\\ &= \lim_{h\to 0}\frac{h\left(-12+2h\right)}{h}\\ &= \lim_{h\to 0}\ -12+2h=-12\end{aligned}$$

Example 3

Evaluate the following limit.

$$\lim_{t\to 4}\frac{t-\sqrt{3t+4}}{4-t}$$

Solution

This limit is going to be a little more work than the previous two. Once again however note that we get the indeterminate form $0/0$ if we try to just evaluate the limit. Also note that neither of the two examples will be of any help here, at least initially. We can't factor the equation and we can't just multiply something out to get the equation to simplify.

When there is a square root in the numerator or denominator we can try to rationalize and see if that helps. Recall that rationalizing makes use of the fact that

$$(a+b)(a-b) = a^2 - b^2$$

So, if either the first and/or the second term have a square root in them the rationalizing will eliminate the root(s). This *might* help in evaluating the limit.

Let's try rationalizing the numerator in this case.

$$\lim_{t\to 4} \frac{t-\sqrt{3t+4}}{4-t} = \lim_{t\to 4} \frac{\left(t-\sqrt{3t+4}\right)}{(4-t)} \frac{\left(t+\sqrt{3t+4}\right)}{\left(t+\sqrt{3t+4}\right)}$$

Remember that to rationalize we just take the numerator (since that's what we're rationalizing), change the sign on the second term and multiply the numerator and denominator by this new term.

Next, we multiply the numerator out being careful to watch minus signs.

$$\begin{aligned}\lim_{t\to 4} \frac{t-\sqrt{3t+4}}{4-t} &= \lim_{t\to 4} \frac{t^2-(3t+4)}{(4-t)\left(t+\sqrt{3t+4}\right)} \\ &= \lim_{t\to 4} \frac{t^2-3t-4}{(4-t)\left(t+\sqrt{3t+4}\right)}\end{aligned}$$

Notice that we didn't multiply the denominator out as well. Most students come out of an Algebra class having it beaten into their heads to always multiply this stuff out. However, in this case multiplying out will make the problem very difficult and in the end you'll just end up factoring it back out anyway.

At this stage we are almost done. Notice that we can factor the numerator so let's do that.

$$\lim_{t\to 4} \frac{t-\sqrt{3t+4}}{4-t} = \lim_{t\to 4} \frac{(t-4)(t+1)}{(4-t)\left(t+\sqrt{3t+4}\right)}$$

Now all we need to do is notice that if we factor a "-1" out of the first term in the denominator we can do some canceling. At that point the division by zero problem will go away and we can evaluate the limit

$$\begin{aligned}\lim_{t\to 4} \frac{t-\sqrt{3t+4}}{4-t} &= \lim_{t\to 4} \frac{(t-4)(t+1)}{-(t-4)\left(t+\sqrt{3t+4}\right)} \\ &= \lim_{t\to 4} \frac{t+1}{-\left(t+\sqrt{3t+4}\right)} \\ &= -\frac{5}{8}\end{aligned}$$

Note that if we had multiplied the denominator out we would not have been able to do this canceling and in all likelihood would not have even seen that some canceling could have been done.

So, we've taken a look at a couple of limits in which evaluation gave the indeterminate form 0/0 and we now have a couple of things to try in these cases.

Let's take a look at another kind of problem that can arise in computing some limits involving piecewise functions.

Example 4

Given the function,

$$g(y) = \begin{cases} y^2 + 5 & \text{if } y < -2 \\ 1 - 3y & \text{if } y \geq -2 \end{cases}$$

Compute the following limits.

(a) $\lim\limits_{y \to 6} g(y)$

(b) $\lim\limits_{y \to -2} g(y)$

Solution

(a) $\lim\limits_{y \to 6} g(y)$

In this case there really isn't a whole lot to do. In doing limits recall that we must always look at what's happening on both sides of the point in question as we move in towards it. In this case $y = 6$ is completely inside the second interval for the function and so there are values of y on both sides of $y = 6$ that are also inside this interval. This means that we can just use the fact to evaluate this limit.

$$\begin{aligned} \lim_{y \to 6} g(y) &= \lim_{y \to 6} (1 - 3y) \\ &= -17 \end{aligned}$$

(b) $\lim\limits_{y \to -2} g(y)$

This part is the real point to this problem. In this case the point that we want to take the limit for is the cutoff point for the two intervals. In other words, we can't just plug $y = -2$ into the second portion because this interval does not contain values of y to the left of $y = -2$ and we need to know what is happening on both sides of the point.

To do this part we are going to have to remember the fact from the section on one-sided limits that says that if the two one-sided limits exist and are the same then the normal limit will also exist and have the same value.

Notice that both of the one-sided limits can be done here since we are only going to be looking at one side of the point in question. So, let's do the two one-sided limits and see what we get.

$$\begin{aligned}
\lim_{y \to -2^-} g(y) &= \lim_{y \to -2^-} (y^2 + 5) && \text{since } y \to -2^- \text{ implies } y < -2 \\
&= 9 \\
\lim_{y \to -2^+} g(y) &= \lim_{y \to -2^+} (1 - 3y) && \text{since } y \to -2^+ \text{ implies } y > -2 \\
&= 7
\end{aligned}$$

So, in this case we can see that,

$$\lim_{y \to -2^-} g(y) = 9 \neq 7 = \lim_{y \to -2^+} g(y)$$

and so since the two one sided limits aren't the same

$$\lim_{y \to -2} g(y)$$

doesn't exist.

Note that a very simple change to the function will make the limit at $y = -2$ exist so don't get in into your head that limits at these cutoff points in piecewise function don't ever exist as the following example will show.

Example 5

Evaluate the following limit.

$$\lim_{y \to -2} g(y) \qquad \text{where, } g(y) = \begin{cases} y^2 + 5 & \text{if } y < -2 \\ 3 - 3y & \text{if } y \geq -2 \end{cases}$$

Solution

The two one-sided limits this time are,

$$\begin{aligned}
\lim_{y \to -2^-} g(y) &= \lim_{y \to -2^-} (y^2 + 5) && \text{since } y \to -2^- \text{ implies } y < -2 \\
&= 9 \\
\lim_{y \to -2^+} g(y) &= \lim_{y \to -2^+} (3 - 3y) && \text{since } y \to -2^+ \text{ implies } y > -2 \\
&= 9
\end{aligned}$$

The one-sided limits are the same so we get,

$$\lim_{y \to -2} g(y) = 9$$

There is one more limit that we need to do. However, we will need a new fact about limits that will help us to do this.

Fact

If $f(x) \le g(x)$ for all x on $[a, b]$ (except possibly at $x = c$) and $a \le c \le b$ then,

$$\lim_{x \to c} f(x) \le \lim_{x \to c} g(x)$$

Note that this fact should make some sense to you if we assume that both functions are nice enough. If both of the functions are "nice enough" to use the limit evaluation fact then we have,

$$\lim_{x \to c} f(x) = f(c) \le g(c) = \lim_{x \to c} g(x)$$

The inequality is true because we know that c is somewhere between a and b and in that range we also know $f(x) \le g(x)$.

Note that we don't really need the two functions to be nice enough for the fact to be true, but it does provide a nice way to give a quick "justification" for the fact.

Also, note that we said that we assumed that $f(x) \le g(x)$ for all x on $[a, b]$ (except possibly at $x = c$). Because limits do not care what is actually happening at $x = c$ we don't really need the inequality to hold at that specific point. We only need it to hold around $x = c$ since that is what the limit is concerned about.

We can take this fact one step farther to get the following theorem.

Squeeze Theorem

Suppose that for all x on $[a, b]$ (except possibly at $x = c$) we have,

$$f(x) \le h(x) \le g(x)$$

Also suppose that,

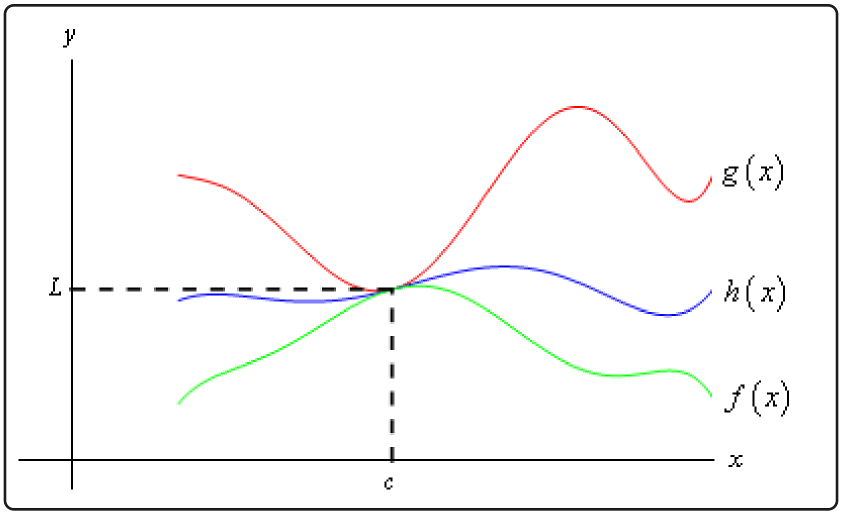
$$\lim_{x \to c} f(x) = \lim_{x \to c} g(x) = L$$

for some $a \le c \le b$. Then,

$$\lim_{x \to c} h(x) = L$$

As with the previous fact we only need to know that $f(x) \le h(x) \le g(x)$ is true around $x = c$ because we are working with limits and they are only concerned with what is going on around $x = c$ and not what is actually happening at $x = c$.

Now, if we again assume that all three functions are nice enough (again this isn't required to make the Squeeze Theorem true, it only helps with the visualization) then we can get a quick sketch of what the Squeeze Theorem is telling us. The following figure illustrates what is happening in this theorem.



From the figure we can see that if the limits of $f(x)$ and $g(x)$ are equal at $x = c$ then the function values must also be equal at $x = c$ (this is where we're using the fact that we assumed the functions were "nice enough", which isn't really required for the Theorem). However, because $h(x)$ is "squeezed" between $f(x)$ and $g(x)$ at this point then $h(x)$ must have the same value. Therefore, the limit of $h(x)$ at this point must also be the same.

The Squeeze theorem is also known as the **Sandwich Theorem** and the **Pinching Theorem**.

So, how do we use this theorem to help us with limits? Let's take a look at the following example to see the theorem in action.

Example 6

Evaluate the following limit.

$$\lim_{x \to 0} x^2 \cos\left(\frac{1}{x}\right)$$

Solution

In this example none of the previous examples can help us. There's no factoring or simplifying to do. We can't rationalize and one-sided limits won't work. There's even a question as to whether this limit will exist since we have division by zero inside the cosine at $x = 0$.

The first thing to notice is that we know the following fact about cosine.

$$-1 \le \cos(x) \le 1$$

Our function doesn't have just an x in the cosine, but as long as we avoid $x = 0$ we can say the same thing for our cosine.

$$-1 \le \cos\left(\frac{1}{x}\right) \le 1$$

It's okay for us to ignore $x = 0$ here because we are taking a limit and we know that limits don't care about what's actually going on at the point in question, $x = 0$ in this case.

Now if we have the above inequality for our cosine we can just multiply everything by an x^2 and get the following.

$$-x^2 \le x^2 \cos\left(\frac{1}{x}\right) \le x^2$$

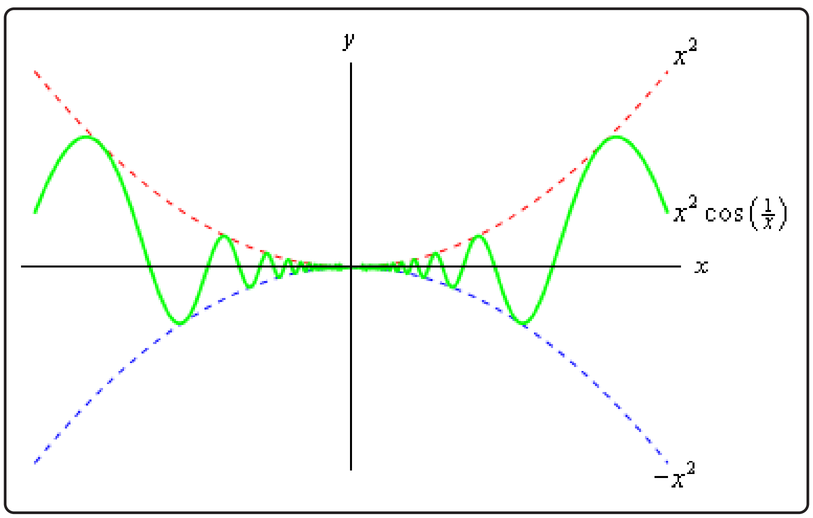
In other words we've managed to squeeze the function that we were interested in between two other functions that are very easy to deal with. So, the limits of the two outer functions are.

$$\lim_{x\to 0} x^2 = 0 \qquad\qquad \lim_{x\to 0} \left(-x^2\right) = 0$$

These are the same and so by the Squeeze theorem we must also have,

$$\lim_{x\to 0} x^2 \cos\left(\frac{1}{x}\right) = 0$$

We can verify this with the graph of the three functions. This is shown below.



In this section we've seen several tools that we can use to help us to compute limits in which we can't just evaluate the function at the point in question. As we will see many of the limits that we'll be doing in later sections will require one or more of these tools.

2.6 Infinite Limits

In this section we will take a look at limits whose value is infinity or minus infinity. These kinds of limit will show up fairly regularly in later sections and in other courses and so you'll need to be able to deal with them when you run across them.

The first thing we should probably do here is to define just what we mean when we say that a limit has a value of infinity or minus infinity.

Definition

We say

$$\lim_{x \to a} f(x) = \infty$$

if we can make $f(x)$ arbitrarily large for all x sufficiently close to $x = a$, from both sides, without actually letting $x = a$.

We say

$$\lim_{x \to a} f(x) = -\infty$$

if we can make $f(x)$ arbitrarily large and negative for all x sufficiently close to $x = a$, from both sides, without actually letting $x = a$.

These definitions can be appropriately modified for the one-sided limits as well. To see a more precise and mathematical definition of this kind of limit see the The Definition of the Limit section at the end of this chapter.

Let's start off with a fairly typical example illustrating infinite limits.

Example 1

Evaluate each of the following limits.

$$\lim_{x \to 0^+} \frac{1}{x} \qquad \lim_{x \to 0^-} \frac{1}{x} \qquad \lim_{x \to 0} \frac{1}{x}$$

Solution

So, we're going to be taking a look at a couple of one-sided limits as well as the normal limit here. In all three cases notice that we can't just plug in $x = 0$. If we did we would get division by zero. Also recall that the definitions above can be easily modified to give similar definitions for the two one-sided limits which we'll be needing here.

Now, there are several ways we could proceed here to get values for these limits. One way is to plug in some points and see what value the function is approaching. In the preceding section we said that we were no longer going to do this, but in this case it is a good way to illustrate just what's going on with this function.

So, here is a table of values of x's from both the left and the right. Using these values we'll be able to estimate the value of the two one-sided limits and once we have that done we can use the fact that the normal limit will exist only if the two one-sided limits exist and have the same value.

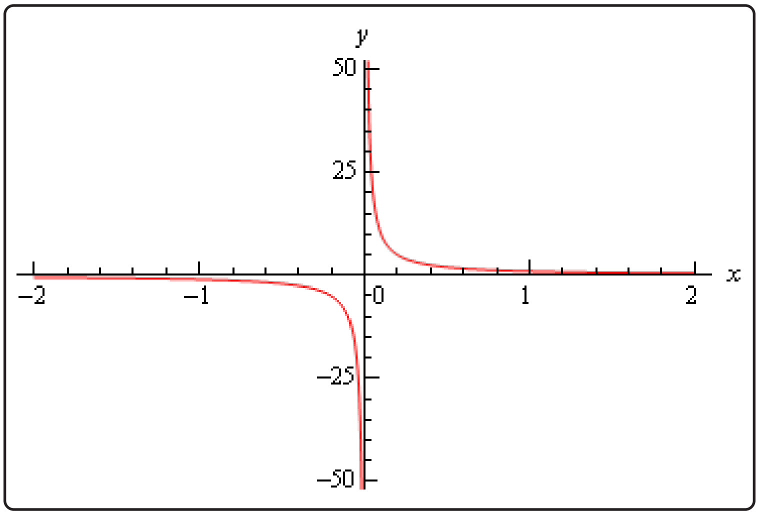
x	$\frac{1}{x}$	x	$\frac{1}{x}$
-0.1	-10	0.1	10
-0.01	-100	0.01	100
-0.001	-1000	0.001	1000
-0.0001	-10000	0.0001	10000

From this table we can see that as we make x smaller and smaller the function $\frac{1}{x}$ gets larger and larger and will retain the same sign that x originally had. It should make sense that this trend will continue for any smaller value of x that we chose to use. The function is a constant (one in this case) divided by an increasingly small number. The resulting fraction should be an increasingly large number and as noted above the fraction will retain the same sign as x.

We can make the function as large and positive as we want for all x's sufficiently close to zero while staying positive (*i.e.* on the right). Likewise, we can make the function as large and negative as we want for all x's sufficiently close to zero while staying negative (*i.e.* on the left). So, from our definition above it looks like we should have the following values for the two one sided limits.

$$\lim_{x \to 0^+} \frac{1}{x} = \infty \qquad \lim_{x \to 0^-} \frac{1}{x} = -\infty$$

Another way to see the values of the two one sided limits here is to graph the function. Again, in the previous section we mentioned that we won't do this too often as most functions are not something we can just quickly sketch out as well as the problems with accuracy in reading values off the graph. In this case however, it's not too hard to sketch a graph of the function and, in this case as we'll see accuracy is not really going to be an issue. So, here is a quick sketch of the graph.



So, we can see from this graph that the function does behave much as we predicted that it would from our table values. The closer x gets to zero from the right the larger (in the positive sense) the function gets, while the closer x gets to zero from the left the larger (in the negative sense) the function gets.

Finally, the normal limit, in this case, will not exist since the two one-sided limits have different values.

So, in summary here are the values of the three limits for this example.

$$\lim_{x\to 0^+}\frac{1}{x}=\infty \qquad \lim_{x\to 0^-}\frac{1}{x}=-\infty \qquad \lim_{x\to 0}\frac{1}{x}\text{ doesn't exist}$$

For most of the remaining examples in this section we'll attempt to "talk our way through" each limit. This means that we'll see if we can analyze what should happen to the function as we get very close to the point in question without actually plugging in any values into the function. For most of the following examples this kind of analysis shouldn't be all that difficult to do. We'll also verify our analysis with a quick graph.

So, let's do a couple more examples.

Example 2

Evaluate each of the following limits.

$$\lim_{x\to 0^+}\frac{6}{x^2} \qquad \lim_{x\to 0^-}\frac{6}{x^2} \qquad \lim_{x\to 0}\frac{6}{x^2}$$

Solution

As with the previous example let's start off by looking at the two one-sided limits. Once we have those we'll be able to determine a value for the normal limit.

So, let's take a look at the right-hand limit first and as noted above let's see if we can figure out what each limit will be doing without actually plugging in any values of x into the function. As we take smaller and smaller values of x, while staying positive, squaring them will only make them smaller (recall squaring a number between zero and one will make it smaller) and of course it will stay positive. So, we have a positive constant divided by an increasingly small positive number. The result should then be an increasingly large positive number. It looks like we should have the following value for the right-hand limit in this case,

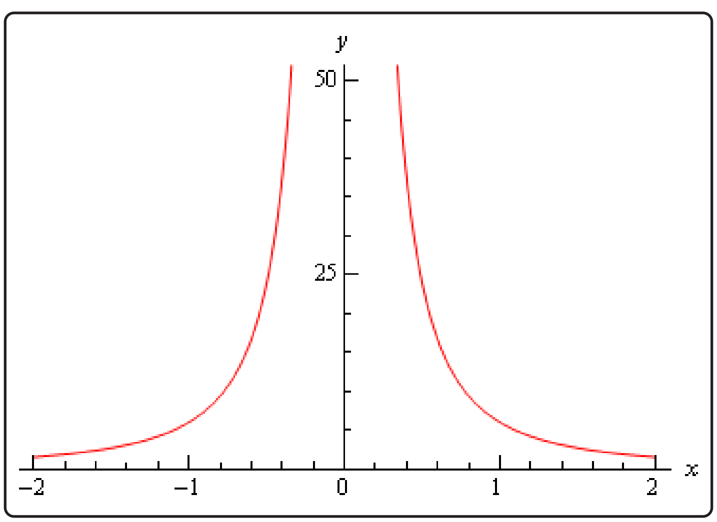
$$\lim_{x\to 0^+} \frac{6}{x^2} = \infty$$

Now, let's take a look at the left-hand limit. In this case we're going to take smaller and smaller values of x, while staying negative this time. When we square them they'll get smaller, but upon squaring the result is now positive. So, we have a positive constant divided by an increasingly small positive number. The result, as with the right-hand limit, will be an increasingly large positive number and so the left-hand limit will be,

$$\lim_{x\to 0^-} \frac{6}{x^2} = \infty$$

Now, in this example, unlike the first one, the normal limit will exist and be infinity since the two one-sided limits both exist and have the same value. So, in summary here are all the limits for this example as well as a quick graph verifying the limits.

$$\lim_{x\to 0^+} \frac{6}{x^2} = \infty \qquad \lim_{x\to 0^-} \frac{6}{x^2} = \infty \qquad \lim_{x\to 0} \frac{6}{x^2} = \infty$$



With this next example we'll move away from just an x in the denominator, but as we'll see in the next couple of examples they work pretty much the same way.

Example 3

Evaluate each of the following limits.

$$\lim_{x \to -2^+} \frac{-4}{x+2} \qquad \lim_{x \to -2^-} \frac{-4}{x+2} \qquad \lim_{x \to -2} \frac{-4}{x+2}$$

Solution

Let's again start with the right-hand limit. With the right-hand limit we know that we have,

$$x > -2 \qquad \Rightarrow \qquad x + 2 > 0$$

Also, as x gets closer and closer to -2 then $x + 2$ will be getting closer and closer to zero, while staying positive as noted above. So, for the right-hand limit, we'll have a negative constant divided by an increasingly small positive number. The result will be an increasingly large and negative number. So, it looks like the right-hand limit will be negative infinity.

For the left-hand limit we have,

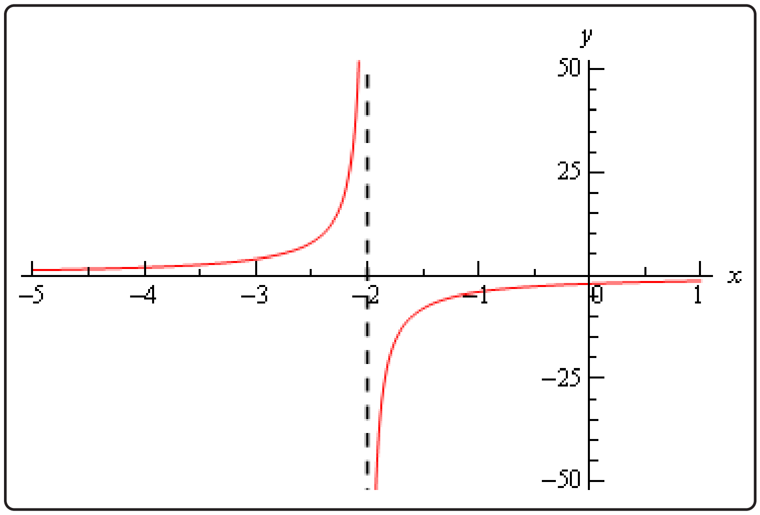
$$x < -2 \qquad \Rightarrow \qquad x + 2 < 0$$

and $x + 2$ will get closer and closer to zero (and be negative) as x gets closer and closer to -2. In this case then we'll have a negative constant divided by an increasingly small negative number. The result will then be an increasingly large positive number and so it looks like the left-hand limit will be positive infinity.

Finally, since two one sided limits are not the same the normal limit won't exist.

Here are the official answers for this example as well as a quick graph of the function for verification purposes.

$$\lim_{x \to -2^+} \frac{-4}{x+2} = -\infty \qquad \lim_{x \to -2^-} \frac{-4}{x+2} = \infty \qquad \lim_{x \to -2} \frac{-4}{x+2} \text{ doesn't exist}$$



At this point we should briefly acknowledge the idea of vertical asymptotes. Each of the three previous graphs have had one. Recall from an Algebra class that a vertical asymptote is a vertical line (the dashed line at $x = -2$ in the previous example) in which the graph will go towards infinity and/or minus infinity on one or both sides of the line.

In an Algebra class they are a little difficult to define other than to say pretty much what we just said. Now that we have infinite limits under our belt we can easily define a vertical asymptote as follows,

Definition

The function $f(x)$ will have a vertical asymptote at $x = a$ if we have any of the following limits at $x = a$.

$$\lim_{x \to a^-} f(x) = \pm\infty \qquad \lim_{x \to a^+} f(x) = \pm\infty \qquad \lim_{x \to a} f(x) = \pm\infty$$

Note that it only requires one of the above limits for a function to have a vertical asymptote at $x = a$.

Using this definition we can see that the first two examples had vertical asymptotes at $x = 0$ while the third example had a vertical asymptote at $x = -2$.

We aren't really going to do a lot with vertical asymptotes here but wanted to mention them at this point since we'd reached a good point to do that.

Let's now take a look at a couple more examples of infinite limits that can cause some problems on occasion.

Example 4

Evaluate each of the following limits.

$$\lim_{x \to 4^+} \frac{3}{(4-x)^3} \qquad \lim_{x \to 4^-} \frac{3}{(4-x)^3} \qquad \lim_{x \to 4} \frac{3}{(4-x)^3}$$

Solution

Let's start with the right-hand limit. For this limit we have,

$$x > 4 \qquad \Rightarrow \qquad 4 - x < 0 \qquad \Rightarrow \qquad (4-x)^3 < 0$$

also, $4 - x \to 0$ as $x \to 4$. So, we have a positive constant divided by an increasingly small negative number. The results will be an increasingly large negative number and so it looks like the right-hand limit will be negative infinity.

For the left-handed limit we have,

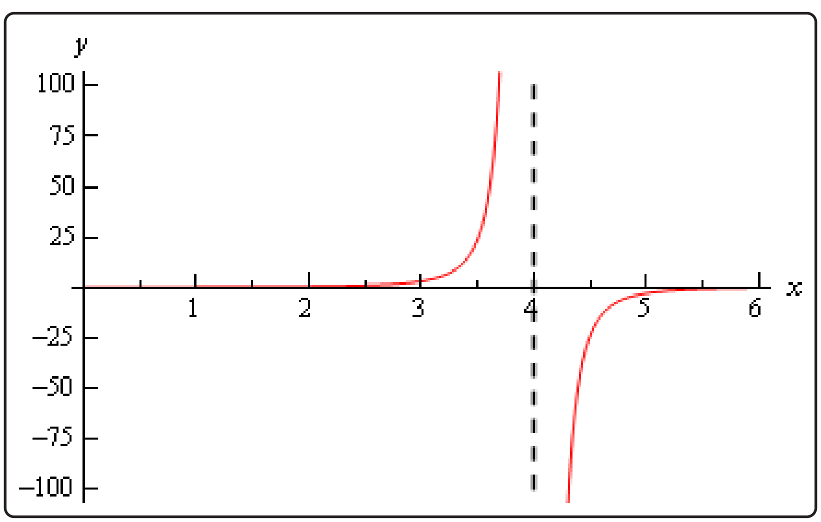
$$x < 4 \qquad \Rightarrow \qquad 4 - x > 0 \qquad \Rightarrow \qquad (4-x)^3 > 0$$

and we still have, $4 - x \to 0$ as $x \to 4$. In this case we have a positive constant divided by an increasingly small positive number. The results will be an increasingly large positive number and so it looks like the left-hand limit will be positive infinity.

The normal limit will not exist since the two one-sided limits are not the same. The official answers to this example are then,

$$\lim_{x \to 4^+} \frac{3}{(4-x)^3} = -\infty \qquad \lim_{x \to 4^-} \frac{3}{(4-x)^3} = \infty \qquad \lim_{x \to 4} \frac{3}{(4-x)^3} \text{ doesn't exist}$$

Here is a quick sketch to verify our limits.



All the examples to this point have had a constant in the numerator and we should probably take a quick look at an example that doesn't have a constant in the numerator.

Example 5

Evaluate each of the following limits.

$$\lim_{x\to 3^+} \frac{2x}{x-3} \qquad \lim_{x\to 3^-} \frac{2x}{x-3} \qquad \lim_{x\to 3} \frac{2x}{x-3}$$

Solution

Let's take a look at the right-handed limit first. For this limit we'll have,

$$x > 3 \qquad \Rightarrow \qquad x - 3 > 0$$

The main difference here with this example is the behavior of the numerator as we let x get closer and closer to 3. In this case we have the following behavior for both the numerator and denominator.

$$x - 3 \to 0 \text{ and } 2x \to 6 \text{ as } x \to 3$$

So, as we let x get closer and closer to 3 (always staying on the right of course) the numerator, while not a constant, is getting closer and closer to a positive constant while the denominator is getting closer and closer to zero and will be positive since we are on the right side.

This means that we'll have a numerator that is getting closer and closer to a non-zero and positive constant divided by an increasingly smaller positive number and so the result should be an increasingly larger positive number. The right-hand limit should then be positive infinity.

For the left-hand limit we'll have,

$$x < 3 \qquad \Rightarrow \qquad x - 3 < 0$$

As with the right-hand limit we'll have the following behaviors for the numerator and the denominator,

$$x - 3 \to 0 \text{ and } 2x \to 6 \text{ as } x \to 3$$

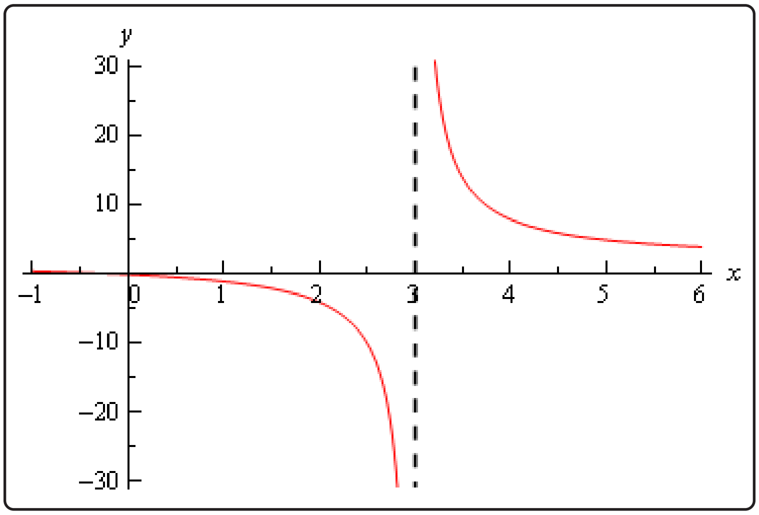
The main difference in this case is that the denominator will now be negative. So, we'll have a numerator that is approaching a positive, non-zero constant divided by an increasingly small negative number. The result will be an increasingly large and negative number.

The formal answers for this example are then,

$$\lim_{x\to 3^+} \frac{2x}{x-3} = \infty \qquad \lim_{x\to 3^-} \frac{2x}{x-3} = -\infty \qquad \lim_{x\to 3} \frac{2x}{x-3} \text{ doesn't exist}$$

As with most of the examples in this section the normal limit does not exist since the two one-sided limits are not the same.

Here's a quick graph to verify our limits.



So far all we've done is look at limits of rational expressions, let's do a couple of quick examples with some different functions.

Example 6

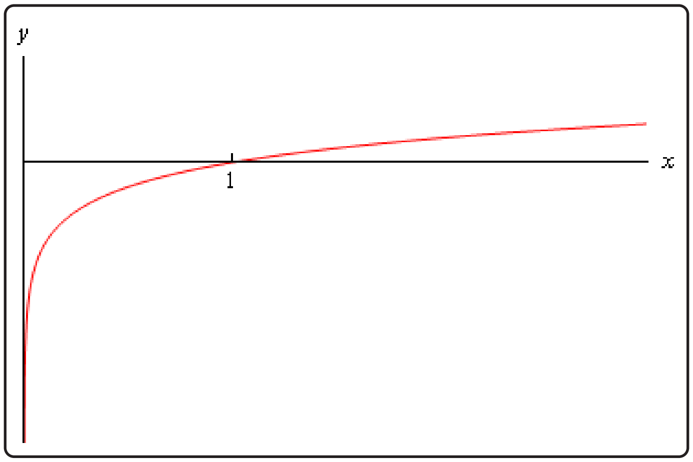
Evaluate

$$\lim_{x \to 0^+} \ln(x)$$

Solution

First, notice that we can only evaluate the right-handed limit here. We know that the domain of any logarithm is only the positive numbers and so we can't even talk about the left-handed limit because that would necessitate the use of negative numbers. Likewise, since we can't deal with the left-handed limit then we can't talk about the normal limit.

This limit is pretty simple to get from a quick sketch of the graph.



From this we can see that,

$$\lim_{x \to 0^+} \ln(x) = -\infty$$

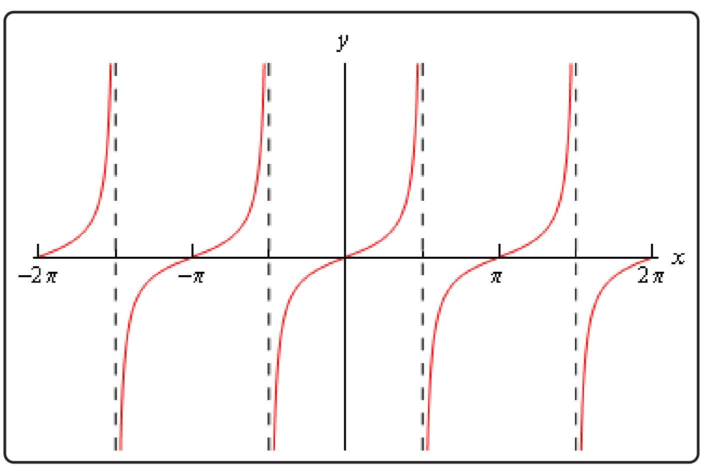
Example 7

Evaluate both of the following limits.

$$\lim_{x \to \frac{\pi}{2}^+} \tan(x) \qquad \lim_{x \to \frac{\pi}{2}^-} \tan(x)$$

Solution

Here's a quick sketch of the graph of the tangent function.



From this it's easy to see that we have the following values for each of these limits,

$$\lim_{x \to \frac{\pi}{2}^+} \tan(x) = -\infty \qquad \lim_{x \to \frac{\pi}{2}^-} \tan(x) = \infty$$

Note that the normal limit will not exist because the two one-sided limits are not the same.

We'll leave this section with a few facts about infinite limits.

Facts

Given the functions $f(x)$ and $g(x)$ suppose we have,

$$\lim_{x \to c} f(x) = \infty \qquad \lim_{x \to c} g(x) = L$$

for some real numbers c and L. Then,

1. $\lim\limits_{x \to c} [f(x) \pm g(x)] = \infty$
2. If $L > 0$ then $\lim\limits_{x \to c} [f(x)\, g(x)] = \infty$
3. If $L < 0$ then $\lim\limits_{x \to c} [f(x)\, g(x)] = -\infty$
4. $\lim\limits_{x \to c} \dfrac{g(x)}{f(x)} = 0$

To see the proof of this set of facts see the Proof of Various Limit Properties section in the Extras appendix.

Note as well that the above set of facts also holds for one-sided limits. They will also hold if $\lim\limits_{x \to c} f(x) = -\infty$, with a change of sign on the infinities in the first three parts. The proofs of these changes to the facts are nearly identical to the proof of the original facts and so are left to the you.

2.7 Limits at Infinity, Part I

In the previous section we saw limits that were infinity and it's now time to take a look at limits at infinity. By limits at infinity we mean one of the following two limits.

$$\lim_{x \to \infty} f(x) \qquad\qquad \lim_{x \to -\infty} f(x)$$

In other words, we are going to be looking at what happens to a function if we let x get very large in either the positive or negative sense. Also, as we'll soon see, these limits may also have infinity as a value.

First, let's note that the set of Facts from the Infinite Limit section also hold if we replace the $\lim\limits_{x \to c}$ with $\lim\limits_{x \to \infty}$ or $\lim\limits_{x \to -\infty}$. The proof of this is nearly identical to the proofof the original set of facts with only minor modifications to handle the change in the limit and so is left to you. We won't need these facts much over the next couple of sections but they will be required on occasion.

In fact, many of the limits that we're going to be looking at we will need the following two facts.

Fact 1

1. If r is a positive rational number and c is any real number then,

$$\lim_{x \to \infty} \frac{c}{x^r} = 0$$

2. If r is a positive rational number, c is any real number and x^r is defined for $x < 0$ then,

$$\lim_{x \to -\infty} \frac{c}{x^r} = 0$$

The first part of this fact should make sense if you think about it. Because we are requiring $r > 0$ we know that x^r will stay in the denominator. Next as we increase x then x^r will also increase. So, we have a constant divided by an increasingly large number and so the result will be increasingly small. Or, in the limit we will get zero.

The second part is nearly identical except we need to worry about x^r being defined for negative x. This condition is here to avoid cases such as $r = \frac{1}{2}$. If this r were allowed we'd be taking the square root of negative numbers which would be complex and we want to avoid that at this level.

Note as well that the sign of c will not affect the answer. Regardless of the sign of c we'll still have a constant divided by a very large number which will result in a very small number and the larger x get the smaller the fraction gets. The sign of c will affect which direction the fraction approaches zero (*i.e.* from the positive or negative side) but it still approaches zero.

If you think about it this is really a special case of the last Fact from the Facts in the previous section. However, to see a direct proof of this fact see the Proof of Various Limit Properties section in the Extras appendix.

Let's start off the examples with one that will lead us to a nice idea that we'll use on a regular basis about limits at infinity for polynomials.

Example 1

Evaluate each of the following limits.

(a) $\lim\limits_{x\to\infty}\left(2x^4 - x^2 - 8x\right)$

(b) $\lim\limits_{t\to-\infty}\left(\frac{1}{3}t^5 + 2t^3 - t^2 + 8\right)$

Solution

(a) $\lim\limits_{x\to\infty}\left(2x^4 - x^2 - 8x\right)$

Our first thought here is probably to just "plug" infinity into the polynomial and "evaluate" each term to determine the value of the limit. It is pretty simple to see what each term will do in the limit and so this seems like an obvious step, especially since we've been doing that for other limits in previous sections.

So, let's see what we get if we do that. As x approaches infinity, then x to a power can only get larger and the coefficient on each term (the first and third) will only make the term even larger. So, if we look at what each term is doing in the limit we get the following,

$$\lim_{x\to\infty}\left(2x^4 - x^2 - 8x\right) = \infty - \infty - \infty$$

Now, we've got a small, but easily fixed, problem to deal with. We are probably tempted to say that the answer is zero (because we have an infinity minus an infinity) or maybe $-\infty$ (because we're subtracting two infinities off of one infinity). However, in both cases we'd be wrong. This is one of those **indeterminate forms** that we first started seeing in a previous section.

Infinities just don't always behave as real numbers do when it comes to arithmetic. Without more work there is simply no way to know what $\infty - \infty$ will be and so we really need to be careful with this kind of problem. To read a little more about this see the Types of Infinity section in the Extras appendix.

So, we need a way to get around this problem. What we'll do here is factor the largest power of x out of the whole polynomial as follows,

$$\lim_{x\to\infty}\left(2x^4 - x^2 - 8x\right) = \lim_{x\to\infty}\left[x^4\left(2 - \frac{1}{x^2} - \frac{8}{x^3}\right)\right]$$

If you're not sure you agree with the factoring above (there's a chance you haven't really been asked to do this kind of factoring prior to this) then recall that to check all you need to do is multiply the x^4 back through the parenthesis to verify it was done correctly. Also, an easy way to remember how to do this kind of factoring is to note that the second term is just the original polynomial divided by x^4. This will always work when factoring a power of x out of a polynomial.

Now for each of the terms we have,

$$\lim_{x\to\infty} x^4 = \infty \qquad\qquad \lim_{x\to\infty}\left(2 - \frac{1}{x^2} - \frac{8}{x^3}\right) = 2$$

The first limit is clearly infinity and for the second limit we'll use the fact above on the last two terms. Therefore using Fact 2 from the previous section we see value of the limit will be,

$$\lim_{x\to\infty}\left(2x^4 - x^2 - 8x\right) = \infty$$

(b) $\lim\limits_{t\to-\infty}\left(\frac{1}{3}t^5 + 2t^3 - t^2 + 8\right)$

We'll work this part much quicker than the previous part. All we need to do is factor out the largest power of t to get the following,

$$\lim_{t\to-\infty}\left(\frac{1}{3}t^5 + 2t^3 - t^2 + 8\right) = \lim_{t\to-\infty}\left[t^5\left(\frac{1}{3} + \frac{2}{t^2} - \frac{1}{t^3} + \frac{8}{t^5}\right)\right]$$

Remember that all you need to do to get the factoring correct is divide the original polynomial by the power of t we're factoring out, t^5 in this case.

Now all we need to do is take the limit of the two terms. In the first don't forget that since we're going out towards $-\infty$ and we're raising t to the 5th power that the limit will be negative (negative number raised to an odd power is still negative). In the second term we'll again make heavy use of the fact above to see that is a finite number.

Therefore, using a modification of the Facts from the previous section the value of the limit is,

$$\lim_{t\to-\infty}\left(\frac{1}{3}t^5 + 2t^3 - t^2 + 8\right) = -\infty$$

Okay, now that we've seen how a couple of polynomials work we can give a simple fact about polynomials in general.

Fact 2

If $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ is a polynomial of degree n (*i.e.* $a_n \neq 0$) then,

$$\lim_{x\to\infty} p(x) = \lim_{x\to\infty} a_n x^n \qquad\qquad \lim_{x\to-\infty} p(x) = \lim_{x\to-\infty} a_n x^n$$

What this fact is really saying is that when we take a limit at infinity for a polynomial all we need to really do is look at the term with the largest power and ask what that term is doing in the limit since the polynomial will have the same behavior.

You can see the proof in the Proof of Various Limit Properties section in the Extras appendix.

Let's now move into some more complicated limits.

Example 2

Evaluate both of the following limits.

$$\lim_{x\to\infty} \frac{2x^4 - x^2 + 8x}{-5x^4 + 7} \qquad \lim_{x\to-\infty} \frac{2x^4 - x^2 + 8x}{-5x^4 + 7}$$

Solution

First, the only difference between these two is that one is going to positive infinity and the other is going to negative infinity. Sometimes this small difference will affect the value of the limit and at other times it won't.

Let's start with the first limit and as with our first set of examples it might be tempting to just "plug" in the infinity. Since both the numerator and denominator are polynomials we can use the above fact to determine the behavior of each. Doing this gives,

$$\lim_{x\to\infty} \frac{2x^4 - x^2 + 8x}{-5x^4 + 7} = \frac{\infty}{-\infty}$$

This is yet another indeterminate form. In this case we might be tempted to say that the limit is infinity (because of the infinity in the numerator), zero (because of the infinity in the denominator) or -1 (because something divided by itself is one). There are three separate arithmetic "rules" at work here and without work there is no way to know which "rule" will be correct and to make matters worse it's possible that none of them may work and we might get a completely different answer, say $-\frac{2}{5}$ to pick a number completely at random.

So, when we have a polynomial divided by a polynomial we're going to proceed much as we did with only polynomials. We first identify the largest power of x in the denominator (and yes, we only look at the denominator for this) and we then factor this out of both the numerator and denominator. Doing this for the first limit gives,

$$\lim_{x\to\infty} \frac{2x^4 - x^2 + 8x}{-5x^4 + 7} = \lim_{x\to\infty} \frac{x^4\left(2 - \frac{1}{x^2} + \frac{8}{x^3}\right)}{x^4\left(-5 + \frac{7}{x^4}\right)}$$

Once we've done this we can cancel the x^4 from both the numerator and the denominator and then use the Fact 1 above to take the limit of all the remaining terms. This gives,

$$\begin{aligned}\lim_{x\to\infty}\frac{2x^4-x^2+8x}{-5x^4+7} &= \lim_{x\to\infty}\frac{2-\frac{1}{x^2}+\frac{8}{x^3}}{-5+\frac{7}{x^4}} \\ &= \frac{2+0+0}{-5+0} \\ &= -\frac{2}{5}\end{aligned}$$

In this case the indeterminate form was neither of the "obvious" choices of infinity, zero, or -1 so be careful with make these kinds of assumptions with this kind of indeterminate forms.

The second limit is done in a similar fashion. Notice however, that nowhere in the work for the first limit did we actually use the fact that the limit was going to plus infinity. In this case it doesn't matter which infinity we are going towards we will get the same value for the limit.

$$\lim_{x\to-\infty}\frac{2x^4-x^2+8x}{-5x^4+7} = -\frac{2}{5}$$

In the previous example the infinity that we were using in the limit didn't change the answer. This will not always be the case so don't make the assumption that this will always be the case.

Let's take a look at an example where we get different answers for each limit.

Example 3

Evaluate each of the following limits.

$$\lim_{x\to\infty}\frac{\sqrt{3x^2+6}}{5-2x} \qquad \lim_{x\to-\infty}\frac{\sqrt{3x^2+6}}{5-2x}$$

Solution

The square root in this problem won't change our work, but it will make the work a little messier.

Let's start with the first limit. In this case the largest power of x in the denominator is just an x. So, we need to factor an x out of the numerator and the denominator. When we are done factoring the x out we will need an x in both of the numerator and the denominator. To get this in the numerator we will have to factor an x^2 out of the square root so that after we take the square root we will get an x.

This is probably not something you're used to doing, but just remember that when it comes out of the square root it needs to be an x and the only way have an x come out of a square root is to take the square root of x^2 and so that is what we'll need to factor out of the term under the radical. Here's the factoring work for this part,

$$\begin{aligned}\lim_{x\to\infty}\frac{\sqrt{3x^2+6}}{5-2x} &= \lim_{x\to\infty}\frac{\sqrt{x^2\left(3+\frac{6}{x^2}\right)}}{x\left(\frac{5}{x}-2\right)}\\ &= \lim_{x\to\infty}\frac{\sqrt{x^2}\sqrt{3+\frac{6}{x^2}}}{x\left(\frac{5}{x}-2\right)}\end{aligned}$$

This is where we need to be really careful with the square root in the problem. Don't forget that

$$\sqrt{x^2}=|x|$$

Square roots are ALWAYS positive and so we need the absolute value bars on the x to make sure that it will give a positive answer. This is not something that most people ever remember seeing in an Algebra class and in fact it's not always given in an Algebra class. However, at this point it becomes absolutely vital that we know and use this fact. Using this fact the limit becomes,

$$\lim_{x\to\infty}\frac{\sqrt{3x^2+6}}{5-2x} = \lim_{x\to\infty}\frac{|x|\sqrt{3+\frac{6}{x^2}}}{x\left(\frac{5}{x}-2\right)}$$

Now, we can't just cancel the x's. We first will need to get rid of the absolute value bars. To do this let's recall the definition of absolute value.

$$|x| = \begin{cases} x & \text{if } x\geq 0\\ -x & \text{if } x<0\end{cases}$$

In this case we are going out to plus infinity so we can safely assume that the x will be positive and so we can just drop the absolute value bars. The limit is then,

$$\begin{aligned}\lim_{x\to\infty}\frac{\sqrt{3x^2+6}}{5-2x} &= \lim_{x\to\infty}\frac{x\sqrt{3+\frac{6}{x^2}}}{x\left(\frac{5}{x}-2\right)}\\ &= \lim_{x\to\infty}\frac{\sqrt{3+\frac{6}{x^2}}}{\frac{5}{x}-2} = \frac{\sqrt{3+0}}{0-2} = -\frac{\sqrt{3}}{2}\end{aligned}$$

Let's now take a look at the second limit (the one with negative infinity). In this case we will need to pay attention to the limit that we are using. The initial work will be the same up until we reach the following step.

$$\lim_{x\to-\infty}\frac{\sqrt{3x^2+6}}{5-2x} = \lim_{x\to-\infty}\frac{|x|\sqrt{3+\frac{6}{x^2}}}{x\left(\frac{5}{x}-2\right)}$$

In this limit we are going to minus infinity so in this case we can assume that x is negative. So, in order to drop the absolute value bars in this case we will need to tack on a minus sign as well. The limit is then,

$$\begin{aligned}\lim_{x\to-\infty}\frac{\sqrt{3x^2+6}}{5-2x} &= \lim_{x\to-\infty}\frac{-x\sqrt{3+\frac{6}{x^2}}}{x\left(\frac{5}{x}-2\right)}\\ &= \lim_{x\to-\infty}\frac{-\sqrt{3+\frac{6}{x^2}}}{\frac{5}{x}-2}\\ &= \frac{\sqrt{3}}{2}\end{aligned}$$

So, as we saw in the last two examples sometimes the infinity in the limit will affect the answer and other times it won't. Note as well that it doesn't always just change the sign of the number. It can on occasion completely change the value. We'll see an example or two of this in the next section.

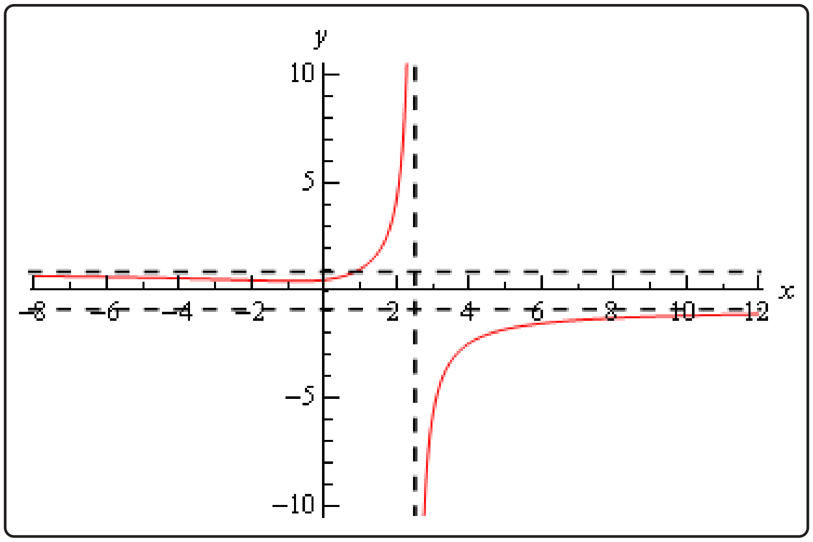
Before moving on to a couple of more examples let's revisit the idea of asymptotes that we first saw in the previous section. Just as we can have vertical asymptotes defined in terms of limits we can also have horizontal asymptotes defined in terms of limits.

Definition

The function $f(x)$ will have a horizontal asymptote at $y = L$ if either of the following are true.

$$\lim_{x\to\infty} f(x) = L \qquad\qquad \lim_{x\to-\infty} f(x) = L$$

We're not going to be doing much with asymptotes here, but it's an easy fact to give and we can use the previous example to illustrate all the asymptote ideas we've seen in the both this section and the previous section. The function in the last example will have two horizontal asymptotes. It will also have a vertical asymptote. Here is a graph of the function showing these.



Let's work another couple of examples involving rational expressions.

Example 4

Evaluate each of the following limits.

$$\lim_{z\to\infty}\frac{4z^2+z^6}{1-5z^3} \qquad \lim_{z\to-\infty}\frac{4z^2+z^6}{1-5z^3}$$

Solution

Let's do the first limit and in this case it looks like we will factor a z^3 out of both the numerator and denominator. Remember that we only look at the denominator when determining the largest power of z here. There is a larger power of z in the numerator but we ignore it. We ONLY look at the denominator when doing this! So, doing the factoring gives,

$$\begin{aligned}\lim_{z\to\infty}\frac{4z^2+z^6}{1-5z^3} &= \lim_{z\to\infty}\frac{z^3\left(\frac{4}{z}+z^3\right)}{z^3\left(\frac{1}{z^3}-5\right)}\\ &= \lim_{z\to\infty}\frac{\frac{4}{z}+z^3}{\frac{1}{z^3}-5}\end{aligned}$$

When we take the limit we'll need to be a little careful. The first term in the numerator and denominator will both be zero. However, the z^3 in the numerator will be going to plus infinity in the limit and so the limit is,

$$\lim_{z\to\infty}\frac{4z^2+z^6}{1-5z^3} = \frac{\infty}{-5} = -\infty$$

The final limit is negative because we have a quotient of positive quantity and a negative quantity.

Now, let's take a look at the second limit. Note that the only different in the work is at the final "evaluation" step and so we'll pick up the work there.

$$\lim_{z \to -\infty} \frac{4z^2 + z^6}{1 - 5z^3} = \lim_{z \to -\infty} \frac{\frac{4}{z} + z^3}{\frac{1}{z^3} - 5} = \frac{-\infty}{-5} = \infty$$

In this case the z^3 in the numerator gives negative infinity in the limit since we are going out to minus infinity and the power is odd. The answer is positive since we have a quotient of two negative numbers.

Example 5

Evaluate the following limit.

$$\lim_{t \to -\infty} \frac{t^2 - 5t - 9}{2t^4 + 3t^3}$$

Solution

In this case it looks like we will factor a t^4 out of both the numerator and denominator. Doing this gives,

$$\begin{aligned}
\lim_{t \to -\infty} \frac{t^2 - 5t - 9}{2t^4 + 3t^3} &= \lim_{t \to -\infty} \frac{t^4\left(\frac{1}{t^2} - \frac{5}{t^3} - \frac{9}{t^4}\right)}{t^4\left(2 + \frac{3}{t}\right)} \\
&= \lim_{t \to -\infty} \frac{\frac{1}{t^2} - \frac{5}{t^3} - \frac{9}{t^4}}{2 + \frac{3}{t}} \\
&= \frac{0}{2} \\
&= 0
\end{aligned}$$

In this case using Fact 1 we can see that the numerator is zero and so since the denominator is also not zero the fraction, and hence the limit, will be zero.

In this section we concentrated on limits at infinity with functions that only involved polynomials and/or rational expression involving polynomials. There are many more types of functions that we could use here. That is the subject of the next section.

To see a precise and mathematical definition of this kind of limit see the The Definition of the Limit section at the end of this chapter.

2.8 Limits at Infinity, Part II

In the previous section we looked at limits at infinity of polynomials and/or rational expression involving polynomials. In this section we want to take a look at some other types of functions that often show up in limits at infinity. The functions we'll be looking at here are exponentials, natural logarithms and inverse tangents.

Let's start by taking a look at a some of very basic examples involving exponential functions.

Example 1

Evaluate each of the following limits.

$$\lim_{x\to\infty} \mathbf{e}^{x} \qquad \lim_{x\to-\infty} \mathbf{e}^{x} \qquad \lim_{x\to\infty} \mathbf{e}^{-x} \qquad \lim_{x\to-\infty} \mathbf{e}^{-x}$$

Solution

There are really just restatements of facts given in the basic exponential section of the review so we'll leave it to you to go back and verify these.

$$\lim_{x\to\infty} \mathbf{e}^{x} = \infty \qquad \lim_{x\to-\infty} \mathbf{e}^{x} = 0 \qquad \lim_{x\to\infty} \mathbf{e}^{-x} = 0 \qquad \lim_{x\to-\infty} \mathbf{e}^{-x} = \infty$$

The main point of this example was to point out that if the exponent of an exponential goes to infinity in the limit then the exponential function will also go to infinity in the limit. Likewise, if the exponent goes to minus infinity in the limit then the exponential will go to zero in the limit.

Note as well, that in the last section the value of the limit did not depend on whether we went to plus or minus infinity. We've already seen in the above example that changing the sign on the infinity can change the answer so do not get locked into any assumptions you may have made from the work in the last section!

Here's a quick set of examples to illustrate these ideas.

Example 2

Evaluate each of the following limits.

(a) $\displaystyle\lim_{x\to\infty} \mathbf{e}^{2-4x-8x^2}$

(b) $\displaystyle\lim_{t\to-\infty} \mathbf{e}^{t^4-5t^2+1}$

(c) $\displaystyle\lim_{z\to 0^+} \mathbf{e}^{\frac{1}{z}}$

Solution

(a) $\lim\limits_{x\to\infty} \mathbf{e}^{2-4x-8x^2}$

In this part what we need to note (using Fact 2 above) is that in the limit the exponent of the exponential does the following,

$$\lim_{x\to\infty} \left(2-4x-8x^2\right) = -\infty$$

So, the exponent goes to minus infinity in the limit and so the exponential must go to zero in the limit using the ideas from the previous set of examples. So, the answer here is,

$$\lim_{x\to\infty} \mathbf{e}^{2-4x-8x^2} = 0$$

(b) $\lim\limits_{t\to-\infty} \mathbf{e}^{t^4-5t^2+1}$

Here let's first note that,

$$\lim_{t\to-\infty} \left(t^4-5t^2+1\right) = \infty$$

The exponent goes to infinity in the limit and so the exponential will also need to go to infinity in the limit. Or,

$$\lim_{t\to-\infty} \mathbf{e}^{t^4-5t^2+1} = \infty$$

(c) $\lim\limits_{z\to0^+} \mathbf{e}^{\frac{1}{z}}$

On the surface this part doesn't appear to belong in this section since it isn't a limit at infinity. However, it does fit into the ideas we're examining in this set of examples.

So, let's first note that using the idea from the previous section we have,

$$\lim_{z\to0^+} \frac{1}{z} = \infty$$

Remember that in order to do this limit here we do need to do a right-hand limit.

So, the exponent goes to infinity in the limit and so the exponential must also go to infinity.

Here's the answer to this part.

$$\lim_{z\to0^+} \mathbf{e}^{\frac{1}{z}} = \infty$$

Let's work some more complicated examples involving exponentials. In the following set of examples it won't be that the exponents are more complicated, but instead that there will be more than one exponential function to deal with.

Example 3

Evaluate each of the following limits.

(a) $\lim\limits_{x\to\infty}\left(\mathbf{e}^{10x} - 4\mathbf{e}^{6x} + 3\mathbf{e}^{x} + 2\mathbf{e}^{-2x} - 9\mathbf{e}^{-15x}\right)$

(b) $\lim\limits_{x\to-\infty}\left(\mathbf{e}^{10x} - 4\mathbf{e}^{6x} + 3\mathbf{e}^{x} + 2\mathbf{e}^{-2x} - 9\mathbf{e}^{-15x}\right)$

Solution

So, the only difference between these two limits is the fact that in the first we're taking the limit as we go to plus infinity and in the second we're going to minus infinity. To this point we've been able to "reuse" work from the first limit in the at least a portion of the second limit. With exponentials that will often not be the case, we're going to treat each of these as separate problems.

(a) $\lim\limits_{x\to\infty}\left(\mathbf{e}^{10x} - 4\mathbf{e}^{6x} + 3\mathbf{e}^{x} + 2\mathbf{e}^{-2x} - 9\mathbf{e}^{-15x}\right)$

Let's start by just taking the limit of each of the pieces and see what we get.

$$\lim_{x\to\infty}\left(\mathbf{e}^{10x} - 4\mathbf{e}^{6x} + 3\mathbf{e}^{x} + 2\mathbf{e}^{-2x} - 9\mathbf{e}^{-15x}\right) = \infty - \infty + \infty + 0 - 0$$

The last two terms aren't any problem (they will be in the next part however; do you see that?). The first three are a problem however as they present us with another indeterminate form.

When dealing with polynomials we factored out the term with the largest exponent in it. Let's do the same thing here. However, we now have to deal with both positive and negative exponents and just what do we mean by the "largest" exponent. When dealing with these here we look at the terms that are causing the problems and ask "what is the largest exponent in those terms?". So, since only the first three terms are causing us problems (*i.e.* they all evaluate to an infinity in the limit) we'll look only at those.

So, since $10x$ is the largest of the three exponents there we'll "factor" an $\mathbf{e}^{10x}$ out of the whole thing. Just as with polynomials we do the factoring by, in essence, dividing each term by $\mathbf{e}^{10x}$ and remembering that to simplify the division all we need to do is subtract the exponents. For example, let's just take a look at the last term,

$$\frac{-9\mathbf{e}^{-15x}}{\mathbf{e}^{10x}} = -9\mathbf{e}^{-15x-10x} = -9\mathbf{e}^{-25x}$$

Doing factoring on all terms then gives,

$$\lim_{x\to\infty}\left(\mathbf{e}^{10x}-4\mathbf{e}^{6x}+3\mathbf{e}^{x}+2\mathbf{e}^{-2x}-9\mathbf{e}^{-15x}\right)=\lim_{x\to\infty}\left[\mathbf{e}^{10x}\left(1-4\mathbf{e}^{-4x}+3\mathbf{e}^{-9x}+2\mathbf{e}^{-12x}-9\mathbf{e}^{-25x}\right)\right]$$

Notice that in doing this factoring all the remaining exponentials now have negative exponents and we know that for this limit (*i.e.* going out to positive infinity) these will all be zero in the limit and so will no longer cause problems.

We can now take the limit of the two factors. The first is clearly infinity and the second is clearly a finite number (one in this case) and so the Facts from the Infinite Limits section gives us the following limit,

$$\lim_{x\to\infty}\left(\mathbf{e}^{10x}-4\mathbf{e}^{6x}+3\mathbf{e}^{x}+2\mathbf{e}^{-2x}-9\mathbf{e}^{-15x}\right)=\infty$$

To simplify the work here a little all we really needed to do was factor the $\mathbf{e}^{10x}$ out of the "problem" terms (the first three in this case) as follows,

$$\begin{aligned}\lim_{x\to\infty}\left(\mathbf{e}^{10x}-4\mathbf{e}^{6x}+3\mathbf{e}^{x}\right)&=\lim_{x\to\infty}\left[\mathbf{e}^{10x}\left(1-4\mathbf{e}^{-4x}+3\mathbf{e}^{-9x}\right)+2\mathbf{e}^{-2x}-9\mathbf{e}^{-15x}\right]\\&=(\infty)(1)+0-0\\&=\infty\end{aligned}$$

We factored the $\mathbf{e}^{10x}$ out of all terms for the practice of doing the factoring and to avoid any issues with having the extra terms at the end. Note as well that while we wrote $(\infty)(1)$ for the limit of the first term we are really using the Facts from the Infinite Limit section to do that limit.

(b) $\displaystyle\lim_{x\to-\infty}\left(\mathbf{e}^{10x}-4\mathbf{e}^{6x}+3\mathbf{e}^{x}+2\mathbf{e}^{-2x}-9\mathbf{e}^{-15x}\right)$

Let's start this one off in the same manner as the first part. Let's take the limit of each of the pieces. This time note that because our limit is going to negative infinity the first three exponentials will in fact go to zero (because their exponents go to minus infinity in the limit). The final two exponentials will go to infinity in the limit (because their exponents go to plus infinity in the limit).

Taking the limits gives,

$$\lim_{x\to-\infty}\left(\mathbf{e}^{10x}-4\mathbf{e}^{6x}+3\mathbf{e}^{x}+2\mathbf{e}^{-2x}-9\mathbf{e}^{-15x}\right)=0-0+0+\infty-\infty$$

So, the last two terms are the problem here as they once again leave us with an indeterminate form. As with the first example we're going to factor out the "largest" exponent in the last two terms. This time however, "largest" doesn't refer to the bigger of the two numbers (-2 is bigger than -15). Instead we're going to use "largest" to refer to the exponent that is farther away from zero. Using this definition of "largest" means that we're going to factor an $\mathbf{e}^{-15x}$ out.

Again, remember that to factor this out all we really are doing is dividing each term by $\mathbf{e}^{-15x}$ and then subtracting exponents. Here's the work for the first term as an example,

$$\frac{\mathbf{e}^{10x}}{\mathbf{e}^{-15x}} = \mathbf{e}^{10x-(-15x)} = \mathbf{e}^{25x}$$

As with the first part we can either factor it out of only the "problem" terms (*i.e.* the last two terms), or all the terms. For the practice we'll factor it out of all the terms. Here is the factoring work for this limit,

$$\lim_{x\to-\infty}\left(\mathbf{e}^{10x} - 4\mathbf{e}^{6x} + 3\mathbf{e}^{x} + 2\mathbf{e}^{-2x} - 9\mathbf{e}^{-15x}\right) = \lim_{x\to-\infty}\left[\mathbf{e}^{-15x}\left(\mathbf{e}^{25x} - 4\mathbf{e}^{21x} + 3\mathbf{e}^{16x} + 2\mathbf{e}^{13x} - 9\right)\right]$$

Finally, after taking the limit of the two terms (the first is infinity and the second is a negative, finite number) and recalling the Facts from the Infinite Limit section we see that the limit is,

$$\lim_{x\to-\infty}\left(\mathbf{e}^{10x} - 4\mathbf{e}^{6x} + 3\mathbf{e}^{x} + 2\mathbf{e}^{-2x} - 9\mathbf{e}^{-15x}\right) = -\infty$$

So, when dealing with sums and/or differences of exponential functions we look for the exponential with the "largest" exponent and remember here that "largest" means the exponent farthest from zero. Also remember that if we're looking at a limit at plus infinity only the exponentials with positive exponents are going to cause problems so those are the only terms we look at in determining the largest exponent. Likewise, if we are looking at a limit at minus infinity then only exponentials with negative exponents are going to cause problems and so only those are looked at in determining the largest exponent.

Finally, as you might have been able to guess from the previous example when dealing with a sum and/or difference of exponentials all we need to do is look at the largest exponent to determine the behavior of the whole expression. Again, remembering that if the limit is at plus infinity we only look at exponentials with positive exponents and if we're looking at a limit at minus infinity we only look at exponentials with negative exponents.

Let's next take a look at some rational functions involving exponentials.

Example 4

Evaluate each of the following limits.

(a) $\displaystyle\lim_{x\to\infty}\frac{6\mathbf{e}^{4x}-\mathbf{e}^{-2x}}{8\mathbf{e}^{4x}-\mathbf{e}^{2x}+3\mathbf{e}^{-x}}$

(b) $\displaystyle\lim_{x\to-\infty}\frac{6\mathbf{e}^{4x}-\mathbf{e}^{-2x}}{8\mathbf{e}^{4x}-\mathbf{e}^{2x}+3\mathbf{e}^{-x}}$

(c) $\displaystyle\lim_{t\to-\infty}\frac{\mathbf{e}^{6t}-4\mathbf{e}^{-6t}}{2\mathbf{e}^{3t}-5\mathbf{e}^{-9t}+\mathbf{e}^{-3t}}$

Solution

As with the previous example, the only difference between the first two parts is that one of the limits is going to plus infinity and the other is going to minus infinity and just as with the previous example each will need to be worked differently.

(a) $\displaystyle\lim_{x\to\infty}\frac{6\mathbf{e}^{4x}-\mathbf{e}^{-2x}}{8\mathbf{e}^{4x}-\mathbf{e}^{2x}+3\mathbf{e}^{-x}}$

The basic concept involved in working this problem is the same as with rational expressions in the previous section. We look at the denominator and determine the exponential function with the "largest" exponent which we will then factor out from both numerator and denominator. We will use the same reasoning as we did with the previous example to determine the "largest" exponent. In the case since we are looking at a limit at plus infinity we only look at exponentials with positive exponents.

So, we'll factor an $\mathbf{e}^{4x}$ out of both then numerator and denominator. Once that is done we can cancel the $\mathbf{e}^{4x}$ and then take the limit of the remaining terms. Here is the work for this limit,

$$\begin{aligned}\lim_{x\to\infty}\frac{6\mathbf{e}^{4x}-\mathbf{e}^{-2x}}{8\mathbf{e}^{4x}-\mathbf{e}^{2x}+3\mathbf{e}^{-x}} &= \lim_{x\to\infty}\frac{\mathbf{e}^{4x}\left(6-\mathbf{e}^{-6x}\right)}{\mathbf{e}^{4x}\left(8-\mathbf{e}^{-2x}+3\mathbf{e}^{-5x}\right)}\\ &= \lim_{x\to\infty}\frac{6-\mathbf{e}^{-6x}}{8-\mathbf{e}^{-2x}+3\mathbf{e}^{-5x}}\\ &= \frac{6-0}{8-0+0}\\ &= \frac{3}{4}\end{aligned}$$

(b) $\displaystyle\lim_{x\to-\infty}\frac{6\mathbf{e}^{4x}-\mathbf{e}^{-2x}}{8\mathbf{e}^{4x}-\mathbf{e}^{2x}+3\mathbf{e}^{-x}}$

In this case we're going to minus infinity in the limit and so we'll look at exponentials in the denominator with negative exponents in determining the "largest" exponent. There's only one however in this problem so that is what we'll use.

Again, remember to only look at the denominator. Do NOT use the exponential from the numerator, even though that one is "larger" than the exponential in the denominator. We always look only at the denominator when determining what term to factor out regardless of what is going on in the numerator.

Here is the work for this part.

$$\begin{aligned}\lim_{x\to-\infty}\frac{6\mathbf{e}^{4x}-\mathbf{e}^{-2x}}{8\mathbf{e}^{4x}-\mathbf{e}^{2x}+3\mathbf{e}^{-x}} &= \lim_{x\to-\infty}\frac{\mathbf{e}^{-x}\left(6\mathbf{e}^{5x}-\mathbf{e}^{-x}\right)}{\mathbf{e}^{-x}\left(8\mathbf{e}^{5x}-\mathbf{e}^{3x}+3\right)}\\ &= \lim_{x\to-\infty}\frac{6\mathbf{e}^{5x}-\mathbf{e}^{-x}}{8\mathbf{e}^{5x}-\mathbf{e}^{3x}+3}\\ &= \frac{0-\infty}{0-0+3}\\ &= -\infty\end{aligned}$$

(c) $\displaystyle\lim_{t\to-\infty}\frac{\mathbf{e}^{6t}-4\mathbf{e}^{-6t}}{2\mathbf{e}^{3t}-5\mathbf{e}^{-9t}+\mathbf{e}^{-3t}}$

We'll do the work on this part with much less detail.

$$\begin{aligned}\lim_{t\to-\infty}\frac{\mathbf{e}^{6t}-4\mathbf{e}^{-6t}}{2\mathbf{e}^{3t}-5\mathbf{e}^{-9t}+\mathbf{e}^{-3t}} &= \lim_{t\to-\infty}\frac{\mathbf{e}^{-9t}\left(\mathbf{e}^{15t}-4\mathbf{e}^{3t}\right)}{\mathbf{e}^{-9t}\left(2\mathbf{e}^{12t}-5+\mathbf{e}^{6t}\right)}\\ &= \lim_{t\to-\infty}\frac{\mathbf{e}^{15t}-4\mathbf{e}^{3t}}{2\mathbf{e}^{12t}-5+\mathbf{e}^{6t}}\\ &= \frac{0-0}{0-5+0}\\ &= 0\end{aligned}$$

Next, let's take a quick look at some basic limits involving logarithms.

Example 5

Evaluate each of the following limits.

$$\lim_{x\to0^+}\ln x \qquad\qquad \lim_{x\to\infty}\ln x$$

Solution

As with the last example I'll leave it to you to verify these restatements from the basic logarithm section.

$$\lim_{x\to0^+}\ln x=-\infty \qquad\qquad \lim_{x\to\infty}\ln x=\infty$$

Note that we had to do a right-handed limit for the first one since we can't plug negative x's into a logarithm. This means that the normal limit won't exist since we must look at x's from both sides of the point in question and x's to the left of zero are negative.

From the previous example we can see that if the argument of a log (the stuff we're taking the log of) goes to zero from the right (*i.e.* always positive) then the log goes to negative infinity in the limit while if the argument goes to infinity then the log also goes to infinity in the limit.

Note as well that we can't look at a limit of a logarithm as x approaches minus infinity since we can't plug negative numbers into the logarithm.

Let's take a quick look at some logarithm examples.

Example 6

Evaluate each of the following limits.

(a) $\lim\limits_{x\to\infty} \ln\left(7x^3 - x^2 + 1\right)$

(b) $\lim\limits_{t\to-\infty} \ln\left(\dfrac{1}{t^2 - 5t}\right)$

Solution

(a) $\lim\limits_{x\to\infty} \ln\left(7x^3 - x^2 + 1\right)$

So, let's first look to see what the argument of the log is doing,

$$\lim_{x\to\infty} \left(7x^3 - x^2 + 1\right) = \infty$$

The argument of the log is going to infinity and so the log must also be going to infinity in the limit. The answer to this part is then,

$$\lim_{x\to\infty} \ln\left(7x^3 - x^2 + 1\right) = \infty$$

(b) $\lim\limits_{t\to-\infty} \ln\left(\dfrac{1}{t^2 - 5t}\right)$

First, note that the limit going to negative infinity here isn't a violation (necessarily) of the fact that we can't plug negative numbers into the logarithm. The real issue is whether or not the argument of the log will be negative or not.

Using the techniques from earlier in this section we can see that,

$$\lim_{t\to-\infty} \frac{1}{t^2 - 5t} = 0$$

and let's also note that for negative numbers (which we can assume we've got since we're going to minus infinity in the limit) the denominator will always be positive and so the quotient will also always be positive. Therefore, not only does the argument go to zero, it goes to zero from the right. This is exactly what we need to do this limit.

So, the answer here is,

$$\lim_{t \to -\infty} \ln\left(\frac{1}{t^2 - 5t}\right) = -\infty$$

As a final set of examples let's take a look at some limits involving inverse tangents.

Example 7

Evaluate each of the following limits.

(a) $\lim\limits_{x \to \infty} \tan^{-1} x$

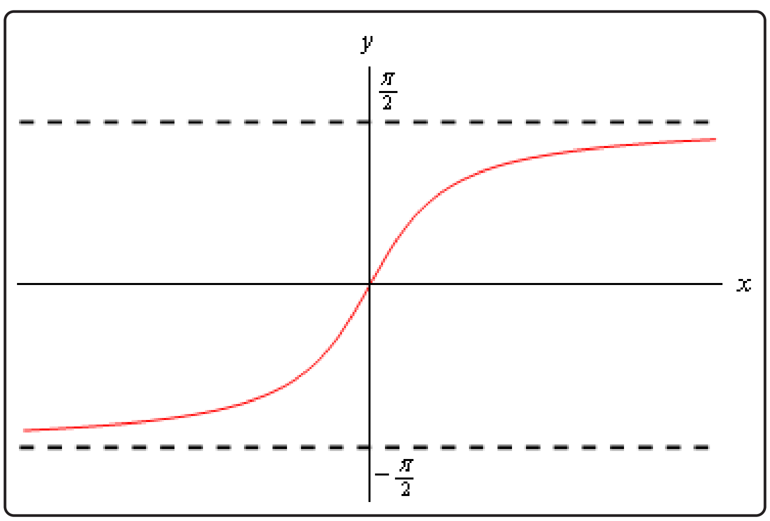
(b) $\lim\limits_{x \to -\infty} \tan^{-1} x$

(c) $\lim\limits_{x \to \infty} \tan^{-1}\left(x^3 - 5x + 6\right)$

(d) $\lim\limits_{x \to 0^-} \tan^{-1}\left(\frac{1}{x}\right)$

Solution

The first two parts here are really just the basic limits involving inverse tangents and can easily be found by examining the following sketch of inverse tangents. The remaining two parts are more involved but as with the exponential and logarithm limits really just refer back to the first two parts as we'll see.



(a) $\lim\limits_{x \to \infty} \tan^{-1} x$

As noted above all we really need to do here is look at the graph of the inverse tangent. Doing this shows us that we have the following value of the limit.

$$\lim_{x \to \infty} \tan^{-1} x = \frac{\pi}{2}$$

(b) $\lim\limits_{x \to -\infty} \tan^{-1} x$

Again, not much to do here other than examine the graph of the inverse tangent.

$$\lim_{x \to -\infty} \tan^{-1} x = -\frac{\pi}{2}$$

(c) $\lim\limits_{x \to \infty} \tan^{-1} \left(x^3 - 5x + 6\right)$

Okay, in part (a) above we saw that if the argument of the inverse tangent function (the stuff inside the parenthesis) goes to plus infinity then we know the value of the limit. In this case (using the techniques from the previous section) we have,

$$\lim_{x \to \infty} x^3 - 5x + 6 = \infty$$

So, this limit is,

$$\lim_{x \to \infty} \tan^{-1} \left(x^3 - 5x + 6\right) = \frac{\pi}{2}$$

(d) $\lim\limits_{x \to 0^-} \tan^{-1} \left(\frac{1}{x}\right)$

Even though this limit is not a limit at infinity we're still looking at the same basic idea here. We'll use part (b) from above as a guide for this limit. We know from the Infinite Limits section that we have the following limit for the argument of this inverse tangent,

$$\lim_{x \to 0^-} \frac{1}{x} = -\infty$$

So, since the argument goes to minus infinity in the limit we know that this limit must be,

$$\lim_{x \to 0^-} \tan^{-1} \left(\frac{1}{x}\right) = -\frac{\pi}{2}$$

To see a precise and mathematical definition of this kind of limit see the The Definition of the Limit section at the end of this chapter.

2.9 Continuity

Over the last few sections we've been using the term "nice enough" to define those functions that we could evaluate limits by just evaluating the function at the point in question. It's now time to formally define what we mean by "nice enough".

Definition

A function $f(x)$ is said to be **continuous** at $x = a$ if

$$\lim_{x \to a} f(x) = f(a)$$

A function is said to be continuous on the interval $[a, b]$ if it is continuous at each point in the interval.

Note that this definition is also implicitly assuming that both $f(a)$ and $\lim\limits_{x \to a} f(x)$ exist. If either of these do not exist the function will not be continuous at $x = a$.

This definition can be turned around into the following fact.

Fact 1

If $f(x)$ is continuous at $x = a$ then,

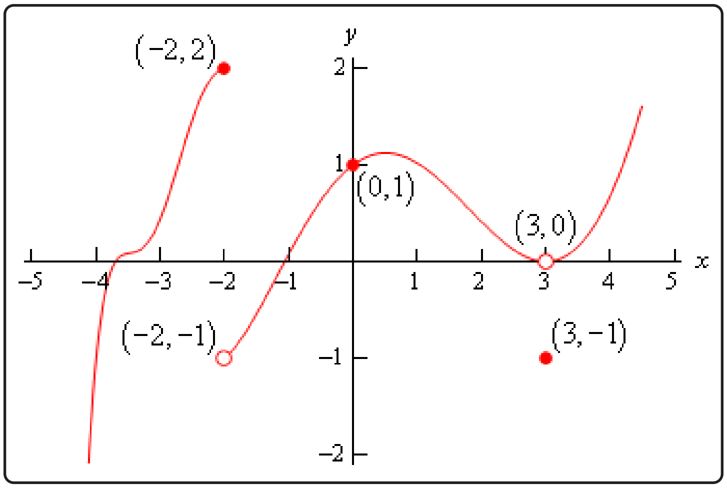
$$\lim_{x \to a} f(x) = f(a) \qquad \lim_{x \to a^-} f(x) = f(a) \qquad \lim_{x \to a^+} f(x) = f(a)$$

This is exactly the same fact that we first put down back when we started looking at limits with the exception that we have replaced the phrase "nice enough" with continuous.

It's nice to finally know what we mean by "nice enough", however, the definition doesn't really tell us just what it means for a function to be continuous. Let's take a look at an example to help us understand just what it means for a function to be continuous.

Example 1

Given the graph of $f(x)$, shown below, determine if $f(x)$ is continuous at $x = -2$, $x = 0$, and $x = 3$.



Solution

To answer the question for each point we'll need to get both the limit at that point and the function value at that point. If they are equal the function is continuous at that point and if they aren't equal the function isn't continuous at that point.

First $x = -2$.

$$f(-2) = 2 \qquad \lim_{x \to -2} f(x) \text{ doesn't exist}$$

The function value and the limit aren't the same and so the function is not continuous at this point. This kind of discontinuity in a graph is called a **jump discontinuity**. Jump discontinuities occur where the graph has a break in it as this graph does and the values of the function to either side of the break are finite (*i.e.* the function doesn't go to infinity).

Now $x = 0$.

$$f(0) = 1 \qquad \lim_{x \to 0} f(x) = 1$$

The function is continuous at this point since the function and limit have the same value.

Finally $x = 3$.

$$f(3) = -1 \qquad \lim_{x \to 3} f(x) = 0$$

The function is not continuous at this point. This kind of discontinuity is called a **removable discontinuity**. Removable discontinuities are those where there is a hole in the graph as there is in this case.

From this example we can get a quick "working" definition of continuity. A function is continuous on an interval if we can draw the graph from start to finish without ever once picking up our pencil. The graph in the last example has only two discontinuities since there are only two places where we would have to pick up our pencil in sketching it.

In other words, a function is continuous if its graph has no holes or breaks in it.

For many functions it's easy to determine where it won't be continuous. Functions won't be continuous where we have things like division by zero or logarithms of zero. Let's take a quick look at an example of determining where a function is not continuous.

Example 2

Determine where the function below is not continuous.

$$h(t) = \frac{4t + 10}{t^2 - 2t - 15}$$

Solution

Rational functions are continuous everywhere except where we have division by zero. So all that we need to is determine where the denominator is zero. That's easy enough to determine by setting the denominator equal to zero and solving.

$$t^2 - 2t - 15 = (t - 5)(t + 3) = 0$$

So, the function will not be continuous at $t = -3$ and $t = 5$.

A nice consequence of continuity is the following fact.

Fact 2

If $f(x)$ is continuous at $x = b$ and $\lim\limits_{x \to a} g(x) = b$ then,

$$\lim_{x \to a} f(g(x)) = f\left(\lim_{x \to a} g(x)\right)$$

To see a proof of this fact see the Proof of Various Limit Properties section in the Extras appendix. With this fact we can now do limits like the following example.

Example 3

Evaluate the following limit.

$$\lim_{x \to 0} \mathbf{e}^{\sin x}$$

Solution

Since we know that exponentials are continuous everywhere we can use the fact above.

$$\lim_{x \to 0} \mathbf{e}^{\sin x} = \mathbf{e}^{\lim\limits_{x \to 0} \sin x} = \mathbf{e}^{0} = 1$$

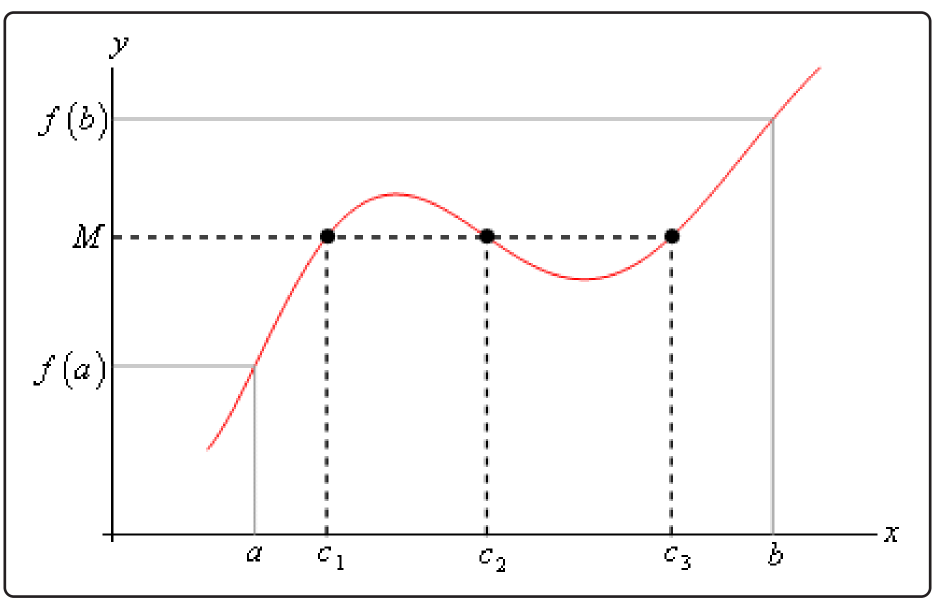
Another very nice consequence of continuity is the Intermediate Value Theorem.

Intermediate Value Theorem

Suppose that $f(x)$ is continuous on $[a, b]$ and let M be any number between $f(a)$ and $f(b)$. Then there exists a number c such that,

1. $a < c < b$
2. $f(c) = M$

All the Intermediate Value Theorem is really saying is that a continuous function will take on all values between $f(a)$ and $f(b)$. Below is a graph of a continuous function that illustrates the Intermediate Value Theorem.



As we can see from this image if we pick any value, M, that is between the value of $f(a)$ and the value of $f(b)$ and draw a line straight out from this point the line will hit the graph in at least one point. In other words, somewhere between a and b the function will take on the value of M. Also, as the figure shows the function may take on the value at more than one place.

It's also important to note that the Intermediate Value Theorem only says that the function will take on the value of M somewhere between a and b. It doesn't say just what that value will be. It only says that it exists.

So, the Intermediate Value Theorem tells us that a function will take the value of M somewhere between a and b but it doesn't tell us where it will take the value nor does it tell us how many times it will take the value. These are important ideas to remember about the Intermediate Value Theorem.

A nice use of the Intermediate Value Theorem is to prove the existence of roots of equations as the following example shows.

Example 4

Show that $p(x) = 2x^3 - 5x^2 - 10x + 5$ has a root somewhere in the interval $[-1, 2]$.

Solution

What we're really asking here is whether or not the function will take on the value

$$p(x) = 0$$

somewhere between -1 and 2. In other words, we want to show that there is a number c such that $-1 < c < 2$ and $p(c) = 0$. However if we define $M = 0$ and acknowledge that $a = -1$ and $b = 2$ we can see that these two condition on c are exactly the conclusions of the Intermediate Value Theorem.

So, this problem is set up to use the Intermediate Value Theorem and in fact, all we need to do is to show that the function is continuous and that $M = 0$ is between $p(-1)$ and $p(2)$ (*i.e.* $p(-1) < 0 < p(2)$ or $p(2) < 0 < p(-1)$ and we'll be done.

To do this all we need to do is compute,

$$p(-1) = 8 \qquad\qquad p(2) = -19$$

So, we have,

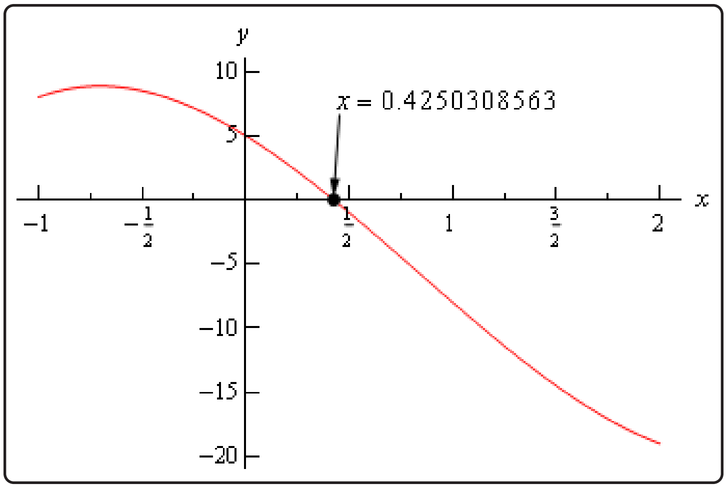
$$-19 = p(2) < 0 < p(-1) = 8$$

Therefore $M = 0$ is between $p(-1)$ and $p(2)$ and since $p(x)$ is a polynomial it's continuous everywhere and so in particular it's continuous on the interval $[-1, 2]$. So by the Intermediate Value Theorem there must be a number $-1 < c < 2$ so that,

$$p(c) = 0$$

Therefore, the polynomial does have a root between -1 and 2.

For the sake of completeness here is a graph showing the root that we just proved existed. Note that we used a computer program to actually find the root and that the Intermediate Value Theorem did not tell us what this value was.



Let's take a look at another example of the Intermediate Value Theorem.

Example 5

If possible, determine if $f(x) = 20\sin(x+3)\cos\left(\frac{x^2}{2}\right)$ takes the following values in the interval $[0,5]$.

(a) Does $f(x) = 10$?

(b) Does $f(x) = -10$?

Solution

Okay, so as with the previous example, we're being asked to determine, if possible, if the function takes on either of the two values above in the interval $[0,5]$. First, let's notice that this is a continuous function and so we know that we can use the Intermediate Value Theorem to do this problem.

Now, for each part we will let M be the given value for that part and then we'll need to show that M lives between $f(0)$ and $f(5)$. If it does, then we can use the Intermediate Value Theorem to prove that the function will take the given value.

So, since we'll need the two function evaluations for each part let's give them here,

$$f(0) = 2.8224 \qquad\qquad f(5) = 19.7436$$

Now, let's take a look at each part.

(a) Does $f(x) = 10$?

Okay, in this case we'll define $M = 10$ and we can see that,

$$f(0) = 2.8224 < 10 < 19.7436 = f(5)$$

So, by the Intermediate Value Theorem there must be a number $0 \le c \le 5$ such that

$$f(c) = 10$$

(b) Does $f(x) = -10$?

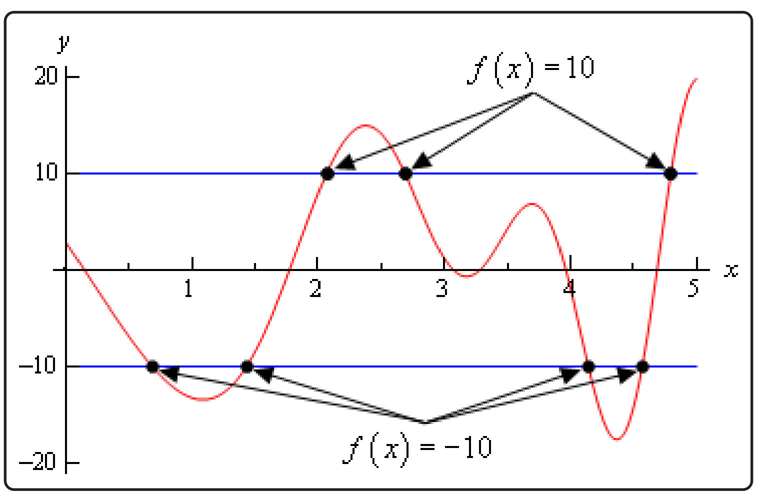
In this part we'll define $M = -10$. We now have a problem. In this part M does not live between $f(0)$ and $f(5)$. So, what does this mean for us? Does this mean that $f(x) \neq -10$ in $[0, 5]$?

Unfortunately for us, this doesn't mean anything. It is possible that $f(x) \neq -10$ in $[0, 5]$, but is it also possible that $f(x) = -10$ in $[0, 5]$. The Intermediate Value Theorem will only tell us that c's will exist. The theorem will NOT tell us that c's don't exist.

In this case it is not possible to determine if $f(x) = -10$ in $[0, 5]$ using the Intermediate Value Theorem.

Okay, as the previous example has shown, the Intermediate Value Theorem will not always be able to tell us what we want to know. Sometimes we can use it to verify that a function will take some value in a given interval and in other cases we won't be able to use it.

For completeness sake here is the graph of $f(x) = 20\sin(x+3)\cos\left(\frac{x^2}{2}\right)$ in the interval $[0, 5]$.



From this graph we can see that not only does $f(x) = -10$ in $[0, 5]$ it does so a total of 4 times!

Also note that as we verified in the first part of the previous example $f(x) = 10$ in $[0, 5]$ and in fact it does so a total of 3 times.

So, remember that the Intermediate Value Theorem will only verify that a function will take on a given value. It will never exclude a value from being taken by the function. Also, if we can use the Intermediate Value Theorem to verify that a function will take on a value it never tells us how many times the function will take on the value, it only tells us that it does take the value.

2.10 The Definition of the Limit

In this section we're going to be taking a look at the precise, mathematical definition of the three kinds of limits we looked at in this chapter. We'll be looking at the precise definition of limits at finite points that have finite values, limits that are infinity and limits at infinity. We'll also give the precise, mathematical definition of continuity.

Let's start this section out with the definition of a limit at a finite point that has a finite value.

Definition 1

Let $f(x)$ be a function defined on an interval that contains $x = a$, except possibly at $x = a$. Then we say that,

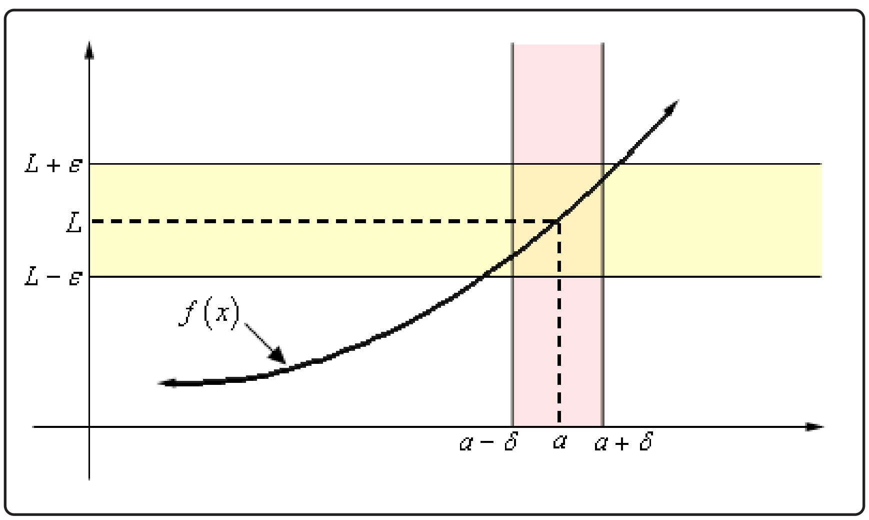
$$\lim_{x \to a} f(x) = L$$

if for every number $\varepsilon > 0$ there is some number $\delta > 0$ such that

$$|f(x) - L| < \varepsilon \qquad \text{whenever} \qquad 0 < |x - a| < \delta$$

Wow. That's a mouth full. Now that it's written down, just what does this mean?

Let's take a look at the following graph and let's also assume that the limit does exist.



What the definition is telling us is that for **any** number $\varepsilon > 0$ that we pick we can go to our graph and sketch two horizontal lines at $L + \varepsilon$ and $L - \varepsilon$ as shown on the graph above. Then somewhere out there in the world is another number $\delta > 0$, which we will need to determine, that will allow us to add in two vertical lines to our graph at $a + \delta$ and $a - \delta$.

If we take any x in the pink region, *i.e.* between $a + \delta$ and $a - \delta$, then this x will be closer to a than either of $a + \delta$ and $a - \delta$. Or,

$$|x - a| < \delta$$

If we now identify the point on the graph that our choice of x gives, then this point on the graph **will** lie in the intersection of the pink and yellow region. This means that this function value $f(x)$ will be closer to L than either of $L + \varepsilon$ and $L - \varepsilon$. Or,

$$|f(x) - L| < \varepsilon$$

If we take any value of x in the pink region then the graph for those values of x will lie in the yellow region.

Notice that there are actually an infinite number of possible δ's that we can choose. In fact, if we go back and look at the graph above it looks like we could have taken a slightly larger δ and still gotten the graph from that pink region to be completely contained in the yellow region.

Also, notice that as the definition points out we only need to make sure that the function is defined in some interval around $x = a$ but we don't really care if it is defined at $x = a$. Remember that limits do not care what is happening at the point, they only care what is happening around the point in question.

Okay, now that we've gotten the definition out of the way and made an attempt to understand it let's see how it's actually used in practice.

These are a little tricky sometimes and it can take a lot of practice to get good at these so don't feel too bad if you don't pick up on this stuff right away. We're going to be looking at a couple of examples that work out fairly easily.

Example 1

Use the definition of the limit to prove the following limit.

$$\lim_{x \to 0} x^2 = 0$$

Solution

In this case both L and a are zero. So, let $\varepsilon > 0$ be any number. Don't worry about what the number is, ε is just some arbitrary number. Now according to the definition of the limit, if this limit is to be true we will need to find some other number $\delta > 0$ so that the following will be true.

$$\left|x^2 - 0\right| < \varepsilon \qquad \text{whenever} \qquad 0 < |x - 0| < \delta$$

Or upon simplifying things we need,

$$\left|x^2\right| < \varepsilon \qquad \text{whenever} \qquad 0 < |x| < \delta$$

Often the way to go through these is to start with the left inequality and do a little simplification and see if that suggests a choice for δ. We'll start by bringing the exponent out of the absolute value bars and then taking the square root of both sides.

$$|x|^2 < \varepsilon \qquad \Rightarrow \qquad |x| < \sqrt{\varepsilon}$$

Now, the results of this simplification looks an awful lot like $0 < |x| < \delta$ with the exception of the "$0 <$" part. Missing that however isn't a problem, it is just telling us that we can't take $x = 0$. So, it looks like if we choose $\delta = \sqrt{\varepsilon}$ we should get what we want.

We'll next need to verify that our choice of δ will give us what we want, *i.e.*,

$$\left|x^2\right| < \varepsilon \qquad \text{whenever} \qquad 0 < |x| < \sqrt{\varepsilon}$$

Verification is in fact pretty much the same work that we did to get our guess. First, let's again let $\varepsilon > 0$ be any number and then choose $\delta = \sqrt{\varepsilon}$. Now, assume that $0 < |x| < \sqrt{\varepsilon}$. We need to show that by choosing x to satisfy this we will get,

$$\left|x^2\right| < \varepsilon$$

To start the verification process, we'll start with $\left|x^2\right|$ and then first strip out the exponent from the absolute values. Once this is done we'll use our assumption on x, namely that $|x| < \sqrt{\varepsilon}$. Doing all this gives,

$$\begin{aligned} \left|x^2\right| &= |x|^2 && \text{strip exponent out of absolute value bars} \\ &< \left(\sqrt{\varepsilon}\right)^2 && \text{use the assumption that } |x| < \sqrt{\varepsilon} \\ &= \varepsilon && \text{simplify} \end{aligned}$$

Or, upon taking the middle terms out, if we assume that $0 < |x| < \sqrt{\varepsilon}$ then we will get,

$$\left|x^2\right| < \varepsilon$$

and this is exactly what we needed to show.

So, just what have we done? We've shown that if we choose $\varepsilon > 0$ then we can find a $\delta > 0$ so that we have,

$$\left|x^2 - 0\right| < \varepsilon \qquad \text{whenever} \qquad 0 < |x - 0| < \sqrt{\varepsilon}$$

and according to our definition this means that,

$$\lim_{x \to 0} x^2 = 0$$

These can be a little tricky the first couple times through. Especially when it seems like we've got to do the work twice. In the previous example we did some simplification on the left-hand inequality to get our guess for δ and then seemingly went through exactly the same work to then prove that our guess was correct. This is often how these work, although we will see an example here in a bit where things don't work out quite so nicely.

So, having said that let's take a look at a slightly more complicated limit, although this one will still be fairly similar to the first example.

Example 2

Use the definition of the limit to prove the following limit.

$$\lim_{x \to 2} 5x - 4 = 6$$

Solution

We'll start this one out the same way that we did the first one. We won't be putting in quite the same amount of explanation however.

Let's start off by letting $\varepsilon > 0$ be any number then we need to find a number $\delta > 0$ so that the following will be true.

$$|(5x-4)-6| < \varepsilon \qquad \text{whenever} \qquad 0 < |x-2| < \delta$$

We'll start by simplifying the left inequality in an attempt to get a guess for δ. Doing this gives,

$$|(5x-4)-6| = |5x-10| = 5\,|x-2| < \varepsilon \qquad \Rightarrow \qquad |x-2| < \frac{\varepsilon}{5}$$

So, as with the first example it looks like if we do enough simplification on the left inequality we get something that looks an awful lot like the right inequality and this leads us to choose $\delta = \frac{\varepsilon}{5}$.

Let's now verify this guess. So, again let $\varepsilon > 0$ be any number and then choose $\delta = \frac{\varepsilon}{5}$. Next, assume that $0 < |x-2| < \delta = \frac{\varepsilon}{5}$ and we get the following,

$$\begin{aligned} |(5x-4)-6| &= |5x-10| && \text{simplify things a little} \\ &= 5\,|x-2| && \text{more simplification....} \\ &< 5\left(\frac{\varepsilon}{5}\right) && \text{use the assumption } \delta = \frac{\varepsilon}{5} \\ &= \varepsilon && \text{and some more simplification} \end{aligned}$$

So, we've shown that

$$|(5x-4)-6| < \varepsilon \qquad \text{whenever} \qquad 0 < |x-2| < \frac{\varepsilon}{5}$$

and so by our definition we have,

$$\lim_{x\to 2} 5x - 4 = 6$$

Okay, so again the process seems to suggest that we have to essentially redo all our work twice, once to make the guess for δ and then another time to prove our guess. Let's do an example that doesn't work out quite so nicely.

Example 3

Use the definition of the limit to prove the following limit.

$$\lim_{x\to 4} x^2 + x - 11 = 9$$

Solution

So, let's get started. Let $\varepsilon > 0$ be any number then we need to find a number $\delta > 0$ so that the following will be true.

$$\left|\left(x^2 + x - 11\right) - 9\right| < \varepsilon \qquad \text{whenever} \qquad 0 < |x-4| < \delta$$

We'll start the guess process in the same manner as the previous two examples.

$$\left|\left(x^2 + x - 11\right) - 9\right| = \left|x^2 + x - 20\right| = |(x+5)(x-4)| = |x+5|\,|x-4| < \varepsilon$$

Okay, we've managed to show that $\left|\left(x^2 + x - 11\right) - 9\right| < \varepsilon$ is equivalent to $|x+5|\,|x-4| < \varepsilon$. However, unlike the previous two examples, we've got an extra term in here that doesn't show up in the right inequality above. If we have any hope of proceeding here we're going to need to find some way to deal with the $|x+5|$.

To do this let's just note that if, by some chance, we can show that $|x+5| < K$ for some number K then, we'll have the following,

$$|x+5|\,|x-4| < K\,|x-4|$$

If we now assume that what we really want to show is $K\,|x-4| < \varepsilon$ instead of $|x+5|\,|x-4| < \varepsilon$ we get the following,

$$|x-4| < \frac{\varepsilon}{K}$$

This is starting to seem familiar isn't it?

All this work however, is based on the assumption that we can show that $|x+5| < K$ for some K. Without this assumption we can't do anything so let's see if we can do this.

Let's first remember that we are working on a limit here and let's also remember that limits are only really concerned with what is happening around the point in question, $x = 4$ in this case. So, it is safe to assume that whatever x is, it must be close to $x = 4$. This means we can safely assume that whatever x is, it is within a distance of, say one of $x = 4$. Or in terms of an inequality, we can assume that,

$$|x - 4| < 1$$

Why choose 1 here? There is no reason other than it's a nice number to work with. We could just have easily chosen 2, or 5, or $\frac{1}{3}$. The only difference our choice will make is on the actual value of K that we end up with. You might want to go through this process with another choice of K and see if you can do it.

So, let's start with $|x - 4| < 1$ and get rid of the absolute value bars and this solve the resulting inequality for x as follows,

$$-1 < x - 4 < 1 \qquad \Rightarrow \qquad 3 < x < 5$$

If we now add 5 to all parts of this inequality we get,

$$8 < x + 5 < 10$$

Now, since $x + 5 > 8 > 0$ (the positive part is important here) we can say that, provided $|x - 4| < 1$ we know that $x + 5 = |x + 5|$. Or, if take the double inequality above we have,

$$8 < x + 5 = |x + 5| < 10 \qquad \Rightarrow \qquad |x + 5| < 10 \qquad \Rightarrow \qquad K = 10$$

So, provided $|x - 4| < 1$ we can see that $|x + 5| < 10$ which in turn gives us,

$$|x - 4| < \frac{\varepsilon}{K} = \frac{\varepsilon}{10}$$

So, to this point we make two assumptions about $|x - 4|$ We've assumed that,

$$|x - 4| < \frac{\varepsilon}{10} \qquad \text{AND} \qquad |x - 4| < 1$$

It may not seem like it, but we're now ready to choose a δ. In the previous examples we had only a single assumption and we used that to give us δ. In this case we've got two and they BOTH need to be true. So, we'll let δ be the smaller of the two assumptions, 1 and $\frac{\varepsilon}{10}$. Mathematically, this is written as,

$$\delta = \min\left\{1, \frac{\varepsilon}{10}\right\}$$

By doing this we can guarantee that,

$$\delta \le \frac{\varepsilon}{10} \qquad \text{AND} \qquad \delta \le 1$$

Now that we've made our choice for δ we need to verify it. So, $\varepsilon > 0$ be any number and then choose $\delta = \min\left\{1, \frac{\varepsilon}{10}\right\}$. Assume that $0 < |x-4| < \delta = \min\left\{1, \frac{\varepsilon}{10}\right\}$. First, we get that,

$$0 < |x-4| < \delta \le \frac{\varepsilon}{10} \qquad \Rightarrow \qquad |x-4| < \frac{\varepsilon}{10}$$

We also get,

$$0 < |x-4| < \delta \le 1 \qquad \Rightarrow \qquad |x-4| < 1 \qquad \Rightarrow \qquad |x+5| < 10$$

Finally, all we need to do is,

$$\begin{aligned}
\left|\left(x^2+x-11\right)-9\right| &= \left|x^2+x-20\right| && \text{simplify things a little} \\
&= |x+5|\,|x-4| && \text{factor} \\
&< 10\,|x-4| && \text{use the assumption that } |x+5| < 10 \\
&< 10\left(\frac{\varepsilon}{10}\right) && \text{use the assumption that } |x-4| < \frac{\varepsilon}{10} \\
&< \varepsilon && \text{a little final simplification}
\end{aligned}$$

We've now managed to show that,

$$\left|\left(x^2+x-11\right)-9\right| < \varepsilon \qquad \text{whenever} \qquad 0 < |x-4| < \min\left\{1, \frac{\varepsilon}{10}\right\}$$

and so by our definition we have,

$$\lim_{x \to 4} x^2 + x - 11 = 9$$

Okay, that was a lot more work that the first two examples and unfortunately, it wasn't all that difficult of a problem. Well, maybe we should say that in comparison to some of the other limits we could have tried to prove it wasn't all that difficult. When first faced with these kinds of proofs using the precise definition of a limit they can all seem pretty difficult.

Do not feel bad if you don't get this stuff right away. It's very common to not understand this right away and to have to struggle a little to fully start to understand how these kinds of limit definition proofs work.

Next, let's give the precise definitions for the right- and left-handed limits.

Definition 2

For the right-hand limit we say that,

$$\lim_{x \to a^+} f(x) = L$$

if for every number $\varepsilon > 0$ there is some number $\delta > 0$ such that

$$|f(x) - L| < \varepsilon \qquad \text{whenever} \qquad 0 < x - a < \delta \quad (\text{or } a < x < a + \delta)$$

Definition 3

For the left-hand limit we say that,

$$\lim_{x \to a^-} f(x) = L$$

if for every number $\varepsilon > 0$ there is some number $\delta > 0$ such that

$$|f(x) - L| < \varepsilon \qquad \text{whenever} \qquad -\delta < x - a < 0 \quad (\text{or } a - \delta < x < a)$$

Note that with both of these definitions there are two ways to deal with the restriction on x and the one in parenthesis is probably the easier to use, although the main one given more closely matches the definition of the normal limit above.

Let's work a quick example of one of these, although as you'll see they work in much the same manner as the normal limit problems do.

Example 4

Use the definition of the limit to prove the following limit.

$$\lim_{x \to 0^+} \sqrt{x} = 0$$

Solution

Let $\varepsilon > 0$ be any number then we need to find a number $\delta > 0$ so that the following will be true.

$$\left|\sqrt{x} - 0\right| < \varepsilon \qquad \text{whenever} \qquad 0 < x - 0 < \delta$$

Or upon a little simplification we need to show,

$$\sqrt{x} < \varepsilon \qquad \text{whenever} \qquad 0 < x < \delta$$

As with the previous problems let's start with the left-hand inequality and see if we can't use that to get a guess for δ. The only simplification that we really need to do here is to square both sides.

$$\sqrt{x} < \varepsilon \qquad \Rightarrow \qquad x < \varepsilon^2$$

So, it looks like we can choose $\delta = \varepsilon^2$.

Let's verify this. Let $\varepsilon > 0$ be any number and chose $\delta = \varepsilon^2$. Next assume that $0 < x < \varepsilon^2$. This gives,

$$\begin{aligned} \left|\sqrt{x} - 0\right| &= \sqrt{x} && \text{some quick simplification} \\ &< \sqrt{\varepsilon^2} && \text{use the assumption that } x < \varepsilon^2 \\ &< \varepsilon && \text{one final simplification} \end{aligned}$$

We've now shown that,

$$\left|\sqrt{x} - 0\right| < \varepsilon \qquad \text{whenever} \qquad 0 < x - 0 < \varepsilon^2$$

and so by the definition of the right-hand limit we have,

$$\lim_{x \to 0^+} \sqrt{x} = 0$$

Let's now move onto the definition of infinite limits. Here are the two definitions that we need to cover both possibilities, limits that are positive infinity and limits that are negative infinity.

Definition 4

Let $f(x)$ be a function defined on an interval that contains $x = a$, except possibly at $x = a$. Then we say that,

$$\lim_{x \to a} f(x) = \infty$$

if for every number $M > 0$ there is some number $\delta > 0$ such that

$$f(x) > M \qquad \text{whenever} \qquad 0 < |x - a| < \delta$$

Definition 5

Let $f(x)$ be a function defined on an interval that contains $x = a$, except possibly at $x = a$. Then we say that,

$$\lim_{x \to a} f(x) = -\infty$$

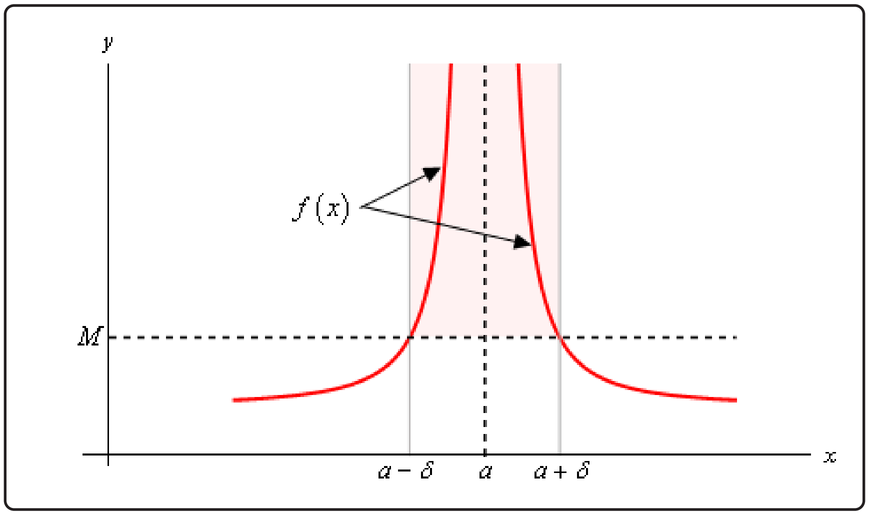
if for every number $N < 0$ there is some number $\delta > 0$ such that

$$f(x) < N \qquad \text{whenever} \qquad 0 < |x - a| < \delta$$

In these two definitions note that M must be a positive number and that N must be a negative number. That's an easy distinction to miss if you aren't paying close attention.

Also note that we could also write down definitions for one-sided limits that are infinity if we wanted to. We'll leave that to you to do if you'd like to.

Here is a quick sketch illustrating Definition 4.



What Definition 4 is telling us is that no matter how large we choose M to be we can always find an interval around $x = a$, given by $0 < |x - a| < \delta$ for some number δ, so that as long as we stay within that interval the graph of the function will be above the line $y = M$ as shown in the graph above. Also note that we don't need the function to actually exist at $x = a$ in order for the definition to hold. This is also illustrated in the graph above.

Note as well that the larger M is the smaller we're probably going to need to make δ.

To see an illustration of Definition 5 reflect the above graph about the x-axis and you'll see a sketch of Definition 5.

Let's work a quick example of one of these to see how these differ from the previous examples.

Example 5

Use the definition of the limit to prove the following limit.

$$\lim_{x \to 0} \frac{1}{x^2} = \infty$$

Solution

These work in pretty much the same manner as the previous set of examples do. The main difference is that we're working with an M now instead of an ε. So, let's get going.

Let $M > 0$ be any number and we'll need to choose a $\delta > 0$ so that,

$$\frac{1}{x^2} > M \qquad \text{whenever} \qquad 0 < |x - 0| = |x| < \delta$$

As with the all the previous problems we'll start with the left inequality and try to get something in the end that looks like the right inequality. To do this we'll basically solve the left inequality for x and we'll need to recall that $\sqrt{x^2} = |x|$. So, here's that work.

$$\frac{1}{x^2} > M \qquad \Rightarrow \qquad x^2 < \frac{1}{M} \qquad \Rightarrow \qquad |x| < \frac{1}{\sqrt{M}}$$

So, it looks like we can choose $\delta = \frac{1}{\sqrt{M}}$. All we need to do now is verify this guess.

Let $M > 0$ be any number, choose $\delta = \frac{1}{\sqrt{M}}$ and assume that $0 < |x| < \frac{1}{\sqrt{M}}$.

In the previous examples we tried to show that our assumptions satisfied the left inequality by working with it directly. However, in this, the function and our assumption on x that we've got actually will make this easier to start with the assumption on x and show that we can get the left inequality out of that. Note that this is being done this way mostly because of the function that we're working with and not because of the type of limit that we've got.

Doing this work gives,

$$\begin{aligned} |x| &< \frac{1}{\sqrt{M}} && \\ |x|^2 &< \frac{1}{M} && \text{square both sides} \\ x^2 &< \frac{1}{M} && \text{acknowledge that } |x|^2 = x^2 \\ \frac{1}{x^2} &> M && \text{solve for } M \end{aligned}$$

So, we've managed to show that,

$$\frac{1}{x^2} > M \qquad \text{whenever} \qquad 0 < |x - 0| < \frac{1}{\sqrt{M}}$$

and so by the definition of the limit we have,

$$\lim_{x \to 0} \frac{1}{x^2} = \infty$$

For our next set of limit definitions let's take a look at the two definitions for limits at infinity. Again, we need one for a limit at plus infinity and another for negative infinity.

Definition 6

Let $f(x)$ be a function defined on $x > K$ for some K. Then we say that,

$$\lim_{x \to \infty} f(x) = L$$

if for every number $\varepsilon > 0$ there is some number $M > 0$ such that

$$|f(x) - L| < \varepsilon \qquad \text{whenever} \qquad x > M$$

Definition 7

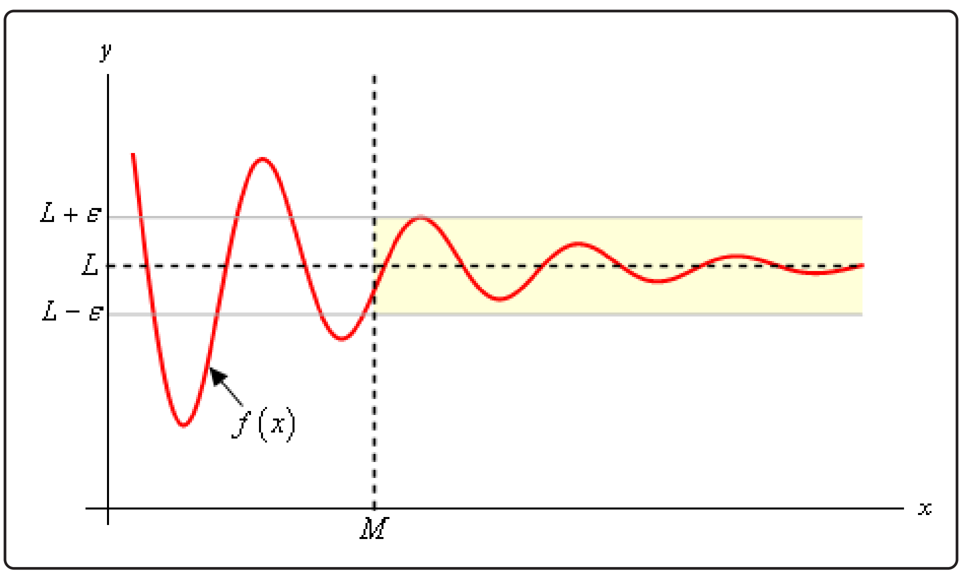
Let $f(x)$ be a function defined on $x < K$ for some K. Then we say that,

$$\lim_{x \to -\infty} f(x) = L$$

if for every number $\varepsilon > 0$ there is some number $N < 0$ such that

$$|f(x) - L| < \varepsilon \qquad \text{whenever} \qquad x < N$$

To see what these definitions are telling us here is a quick sketch illustrating Definition 6. Definition 6 tells us is that no matter how close to L we want to get, mathematically this is given by $|f(x) - L| < \varepsilon$ for any chosen ε, we can find another number M such that provided we take any x bigger than M, then the graph of the function for that x will be closer to L than either $L-\varepsilon$ and $L+\varepsilon$. Or, in other words, the graph will be in the shaded region as shown in the sketch below.



Finally, note that the smaller we make ε the larger we'll probably need to make M.

Here's a quick example of one of these limits.

Example 6

Use the definition of the limit to prove the following limit.

$$\lim_{x \to -\infty} \frac{1}{x} = 0$$

Solution

Let $\varepsilon > 0$ be any number and we'll need to choose a $N < 0$ so that,

$$\left|\frac{1}{x} - 0\right| = \frac{1}{|x|} < \varepsilon \qquad \text{whenever} \qquad x < N$$

Getting our guess for N isn't too bad here.

$$\frac{1}{|x|} < \varepsilon \qquad \Rightarrow \qquad |x| > \frac{1}{\varepsilon}$$

Since we're heading out towards negative infinity it looks like we can choose $N = -\frac{1}{\varepsilon}$. Note that we need the "$-$" to make sure that N is negative (recall that $\varepsilon > 0$).

Let's verify that our guess will work. Let $\varepsilon > 0$ and choose $N = -\frac{1}{\varepsilon}$ and assume that $x < -\frac{1}{\varepsilon}$. As with the previous example the function that we're working with here suggests that it will be easier to start with this assumption and show that we can get the left inequality out of that.

$$\begin{aligned} x &< -\frac{1}{\varepsilon} \\ |x| &> \left|-\frac{1}{\varepsilon}\right| && \text{take the absolute value} \\ |x| &> \frac{1}{\varepsilon} && \text{do a little simplification} \\ \frac{1}{|x|} &< \varepsilon && \text{solve for } |x| \\ \left|\frac{1}{x} - 0\right| &< \varepsilon && \text{rewrite things a little} \end{aligned}$$

Note that when we took the absolute value of both sides we changed both sides from negative numbers to positive numbers and so also had to change the direction of the inequality.

So, we've shown that,

$$\left|\frac{1}{x} - 0\right| = \frac{1}{|x|} < \varepsilon \qquad \text{whenever} \qquad x < -\frac{1}{\varepsilon}$$

and so by the definition of the limit we have,

$$\lim_{x \to -\infty} \frac{1}{x} = 0$$

For our final limit definition let's look at limits at infinity that are also infinite in value. There are four possible limits to define here. We'll do one of them and leave the other three to you to write down if you'd like to.

Definition 8

Let $f(x)$ be a function defined on $x > K$ for some K. Then we say that,

$$\lim_{x \to \infty} f(x) = \infty$$

if for every number $N > 0$ there is some number $M > 0$ such that

$$f(x) > N \qquad \text{whenever} \qquad x > M$$

The other three definitions are almost identical. The only differences are the signs of M and/or N and the corresponding inequality directions.

As a final definition in this section let's recall that we previously said that a function was continuous if,

$$\lim_{x \to a} f(x) = f(a)$$

So, since continuity, as we previously defined it, is defined in terms of a limit we can also now give a more precise definition of continuity. Here it is,

Definition 9

Let $f(x)$ be a function defined on an interval that contains $x = a$. Then we say that $f(x)$ is continuous at $x = a$ if for every number $\varepsilon > 0$ there is some number $\delta > 0$ such that

$$|f(x) - f(a)| < \varepsilon \qquad \text{whenever} \qquad 0 < |x - a| < \delta$$

This definition is very similar to the first definition in this section and of course that should make some sense since that is exactly the kind of limit that we're doing to show that a function is continuous. The only real difference is that here we need to make sure that the function is actually defined at $x = a$, while we didn't need to worry about that for the first definition since limits don't really care what is happening at the point.

We won't do any examples of proving a function is continuous at a point here mostly because we've already done some examples. Go back and look at the first three examples. In each of these examples the value of the limit was the value of the function evaluated at $x = a$ and so in each of these examples not only did we prove the value of the limit we also managed to prove that each of these functions are continuous at the point in question.

3 Derivatives

In this chapter we will start looking at the next major topic in a calculus class, derivatives. This chapter is devoted almost exclusively to finding/computing derivatives. We will, however, take a look at a single application of derivatives in this chapter. We will be leaving most of the applications of derivatives that we will be discussing to the next chapter.

This chapter will start out with defining just what a derivative is as well as look at a couple of the main interpretations. In the process we will start to understand just how interconnected the main topics of first Calculus course are. In particular, we will see that, in theory, we can't do derivatives unless we can also do limits.

However, having said that we'll also see that using limits to compute derivatives can be a fairly long process that is prone to inadvertent errors if we get in a hurry and, in some cases, will be all but impossible to do. Therefore, after discussing the definition of the derivative we'll move off to looking at some formulas for computing derivatives that will allow us to avoid having to use limits to compute derivatives. Note however that won't mean that we can just forget all about using limits to compute derivatives. That is still something that will, on occasion, come up so we can't forget about that.

We will discuss formulas for the following functions.

- Functions involving polynomials, roots and more generally, terms involving variables raised to a power.
- Trigonometric functions.
- Exponential and Logarithm functions.
- Inverse Trigonometric functions.
- Hyperbolic functions.

We'll also see very quickly that while the formulas for the functions above are nice they won't actually allow us to differentiate just any function that involved them. So, we will also discuss the Product and Quotient Rules allowing us to differentiate, oddly enough, products and quotients involving the functions listed above. We will also take a long look at something called the Chain Rule which will again greatly expand the number of functions we can differentiate. In fact, the Chain Rule may be the most important of the formulas we discuss as easily the majority of derivatives will be taking eventually will involve the Chain Rule at least partially.

In addition we will also take a look at implicit differentiation. This will, again, expand the number derivatives that we can find, including allowing us to find derivatives that we would not be able to find otherwise. Implicit differentiation will also allow us to look at the only application of derivatives that we will look at in this chapter, Related Rates. Related Rates problems will allow us to determine the rate of change of a quantity provided we know something about the rates of change for the other quantities in the problem.

We will also look at higher order derivatives. Or, in other words, we will take the derivative of a derivative and discuss an application of of at least one of the higher order derivatives.

We will then close out the chapter with a quick discussion of Logarithmic Differentiation. Logarithmic Differentiation is al alternative method of differentiation that can be used instead of the Product and Quotient Rule (sometimes easier sometimes not...). More importantly logarithmic differentiation will allow us to differentiate a class of functions that none of the formulas we will have discussed in this chapter up to this point would allow us differentiate.

3.1 The Definition of the Derivative

In the first section of the Limits chapter we saw that the computation of the slope of a tangent line, the instantaneous rate of change of a function, and the instantaneous velocity of an object at $x = a$ all required us to compute the following limit.

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

We also saw that with a small change of notation this limit could also be written as,

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \tag{3.1}$$

This is such an important limit and it arises in so many places that we give it a name. We call it a **derivative**. Here is the official definition of the derivative.

Definition of the Derivative

The **derivative of** $f(x)$ **with respect to x** is the function $f'(x)$ and is defined as,

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{3.2}$$

Note that we replaced all the *a*'s in Equation 3.1 with x's to acknowledge the fact that the derivative is really a function as well. We often "read" $f'(x)$ as "f prime of x".

Let's compute a couple of derivatives using the definition.

Example 1

Find the derivative of the following function using the definition of the derivative.

$$f(x) = 2x^2 - 16x + 35$$

Solution

So, all we really need to do is to plug this function into the definition of the derivative, Equation 3.2, and do some algebra. While, admittedly, the algebra will get somewhat unpleasant at times, but it's just algebra so don't get excited about the fact that we're now computing derivatives.

First plug the function into the definition of the derivative.

$$\begin{aligned} f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0}\frac{2(x+h)^2-16(x+h)+35-\left(2x^2-16x+35\right)}{h} \end{aligned}$$

Be careful and make sure that you properly deal with parenthesis when doing the subtracting.

Now, we know from the previous chapter that we can't just plug in $h=0$ since this will give us a division by zero error. So, we are going to have to do some work. In this case that means multiplying everything out and distributing the minus sign through on the second term. Doing this gives,

$$\begin{aligned} f'(x) &= \lim_{h\to 0}\frac{2x^2+4xh+2h^2-16x-16h+35-2x^2+16x-35}{h} \\ &= \lim_{h\to 0}\frac{4xh+2h^2-16h}{h} \end{aligned}$$

Notice that every term in the numerator that didn't have an h in it canceled out and we can now factor an h out of the numerator which will cancel against the h in the denominator. After that we can compute the limit.

$$\begin{aligned} f'(x) &= \lim_{h\to 0}\frac{h(4x+2h-16)}{h} \\ &= \lim_{h\to 0} 4x+2h-16 \\ &= 4x-16 \end{aligned}$$

So, the derivative is,

$$f'(x) = 4x-16$$

Example 2

Find the derivative of the following function using the definition of the derivative.

$$g(t) = \frac{t}{t+1}$$

Solution

This one is going to be a little messier as far as the algebra goes. However, outside of that it will work in exactly the same manner as the previous examples. First, we plug the function into the definition of the derivative,

$$\begin{aligned} g'(t) &= \lim_{h\to 0} \frac{g(t+h) - g(t)}{h} \\ &= \lim_{h\to 0} \frac{1}{h}\left(\frac{t+h}{t+h+1} - \frac{t}{t+1}\right) \end{aligned}$$

Note that we changed all the letters in the definition to match up with the given function. Also note that we wrote the fraction a much more compact manner to help us with the work.

As with the first problem we can't just plug in $h = 0$. So, we will need to simplify things a little. In this case we will need to combine the two terms in the numerator into a single rational expression as follows.

$$\begin{aligned} g'(t) &= \lim_{h\to 0} \frac{1}{h}\left(\frac{(t+h)(t+1) - t(t+h+1)}{(t+h+1)(t+1)}\right) \\ &= \lim_{h\to 0} \frac{1}{h}\left(\frac{t^2+t+th+h-\left(t^2+th+t\right)}{(t+h+1)(t+1)}\right) \\ &= \lim_{h\to 0} \frac{1}{h}\left(\frac{h}{(t+h+1)(t+1)}\right) \end{aligned}$$

Before finishing this let's note a couple of things. First, we didn't multiply out the denominator. Multiplying out the denominator will just overly complicate things so let's keep it simple. Next, as with the first example, after the simplification we only have terms with h's in them left in the numerator and so we can now cancel an h out.

So, upon canceling the h we can evaluate the limit and get the derivative.

$$\begin{aligned} g'(t) &= \lim_{h\to 0} \frac{1}{(t+h+1)(t+1)} \\ &= \frac{1}{(t+1)(t+1)} \\ &= \frac{1}{(t+1)^2} \end{aligned}$$

The derivative is then,

$$g'(t) = \frac{1}{(t+1)^2}$$

Example 3

Find the derivative of the following function using the definition of the derivative.

$$R(z) = \sqrt{5z - 8}$$

Solution

First plug into the definition of the derivative as we've done with the previous two examples.

$$\begin{aligned} R'(z) &= \lim_{h \to 0} \frac{R(z+h) - R(z)}{h} \\ &= \lim_{h \to 0} \frac{\sqrt{5(z+h) - 8} - \sqrt{5z - 8}}{h} \end{aligned}$$

In this problem we're going to have to rationalize the numerator. You do remember rationalization from an Algebra class right? In an Algebra class you probably only rationalized the denominator, but you can also rationalize numerators. Remember that in rationalizing the numerator (in this case) we multiply both the numerator and denominator by the numerator except we change the sign between the two terms. Here's the rationalizing work for this problem,

$$\begin{aligned} R'(z) &= \lim_{h \to 0} \frac{\left(\sqrt{5(z+h) - 8} - \sqrt{5z - 8}\right)}{h} \frac{\left(\sqrt{5(z+h) - 8} + \sqrt{5z - 8}\right)}{\left(\sqrt{5(z+h) - 8} + \sqrt{5z - 8}\right)} \\ &= \lim_{h \to 0} \frac{5z + 5h - 8 - (5z - 8)}{h\left(\sqrt{5(z+h) - 8} + \sqrt{5z - 8}\right)} \\ &= \lim_{h \to 0} \frac{5h}{h\left(\sqrt{5(z+h) - 8} + \sqrt{5z - 8}\right)} \end{aligned}$$

Again, after the simplification we have only h's left in the numerator. So, cancel the h and evaluate the limit.

$$\begin{aligned} R'(z) &= \lim_{h \to 0} \frac{5}{\sqrt{5(z+h) - 8} + \sqrt{5z - 8}} \\ &= \frac{5}{\sqrt{5z - 8} + \sqrt{5z - 8}} \\ &= \frac{5}{2\sqrt{5z - 8}} \end{aligned}$$

And so we get a derivative of,

$$R'(z) = \frac{5}{2\sqrt{5z - 8}}$$

Let's work one more example. This one will be a little different, but it's got a point that needs to be made.

Example 4

Determine $f'(0)$ for $f(x) = |x|$.

Solution

Since this problem is asking for the derivative at a specific point we'll go ahead and use that in our work. It will make our life easier and that's always a good thing.

So, plug into the definition and simplify.

$$\begin{aligned} f'(0) &= \lim_{h \to 0} \frac{f(0+h) - f(0)}{h} \\ &= \lim_{h \to 0} \frac{|0+h| - |0|}{h} \\ &= \lim_{h \to 0} \frac{|h|}{h} \end{aligned}$$

We saw a situation like this back when we were looking at limits at infinity. As in that section we can't just cancel the h's. We will have to look at the two one sided limits and recall that

$$|h| = \begin{cases} h & \text{if } h \geq 0 \\ -h & \text{if } h < 0 \end{cases}$$

$$\begin{aligned} \lim_{h \to 0^-} \frac{|h|}{h} &= \lim_{h \to 0^-} \frac{-h}{h} \qquad \text{because } h < 0 \text{ in a left-hand limit.} \\ &= \lim_{h \to 0^-} (-1) \\ &= -1 \end{aligned}$$

$$\begin{aligned} \lim_{h \to 0^+} \frac{|h|}{h} &= \lim_{h \to 0^+} \frac{h}{h} \qquad \text{because } h > 0 \text{ in a right-hand limit.} \\ &= \lim_{h \to 0^+} 1 \\ &= 1 \end{aligned}$$

The two one-sided limits are different and so

$$\lim_{h \to 0} \frac{|h|}{h}$$

doesn't exist. However, this is the limit that gives us the derivative that we're after.

If the limit doesn't exist then the derivative doesn't exist either.

In this example we have finally seen a function for which the derivative doesn't exist at a point. This is a fact of life that we've got to be aware of. Derivatives will not always exist. Note as well that this doesn't say anything about whether or not the derivative exists anywhere else. In fact, the derivative of the absolute value function exists at every point except the one we just looked at, $x = 0$.

The preceding discussion leads to the following definition.

Definition

A function $f(x)$ is called **differentiable** at $x = a$ if $f'(a)$ exists and $f(x)$ is called differentiable on an interval if the derivative exists for each point in that interval.

The next theorem shows us a very nice relationship between functions that are continuous and those that are differentiable.

Theorem

If $f(x)$ is differentiable at $x = a$ then $f(x)$ is continuous at $x = a$.

See the Proof of Various Derivative Formulas section of the Extras appendix to see the proof of this theorem.

Note that this theorem does not work in reverse. Consider $f(x) = |x|$ and take a look at,

$$\lim_{x \to 0} f(x) = \lim_{x \to 0} |x| = 0 = f(0)$$

So, $f(x) = |x|$ is continuous at $x = 0$ but we've just shown above in Example 4 that $f(x) = |x|$ is not differentiable at $x = 0$.

Alternate Notation

Next, we need to discuss some alternate notation for the derivative. The typical derivative notation is the "prime" notation. However, there is another notation that is used on occasion so let's cover that.

Given a function $y = f(x)$ all of the following are equivalent and represent the derivative of $f(x)$ with respect to *x*.

$$f'(x) = y' = \frac{df}{dx} = \frac{dy}{dx} = \frac{d}{dx}(f(x)) = \frac{d}{dx}(y)$$

Because we also need to evaluate derivatives on occasion we also need a notation for evaluating derivatives when using the fractional notation. So, if we want to evaluate the derivative at $x = a$ all of the following are equivalent.

$$f'(a) = y'\big|_{x=a} = \left.\frac{df}{dx}\right|_{x=a} = \left.\frac{dy}{dx}\right|_{x=a}$$

Note as well that on occasion we will drop the (x) part on the function to simplify the notation somewhat. In these cases the following are equivalent.

$$f'(x) = f'$$

As a final note in this section we'll acknowledge that computing most derivatives directly from the definition is a fairly complex (and sometimes painful) process filled with opportunities to make mistakes. In a couple of sections we'll start developing formulas and/or properties that will help us to take the derivative of many of the common functions so we won't need to resort to the definition of the derivative too often.

This does not mean however that it isn't important to know the definition of the derivative! It is an important definition that we should always know and keep in the back of our minds. It is just something that we're not going to be working with all that much.

3.2 Interpretation of the Derivative

Before moving on to the section where we learn how to compute derivatives by avoiding the limits we were evaluating in the previous section we need to take a quick look at some of the interpretations of the derivative. All of these interpretations arise from recalling how our definition of the derivative came about. The definition came about by noticing that all the problems that we worked in the first section in the Limits chapter required us to evaluate the same limit.

Rate of Change

The first interpretation of a derivative is rate of change. This was not the first problem that we looked at in the Limits chapter, but it is the most important interpretation of the derivative. If $f(x)$ represents a quantity at any x then the derivative $f'(a)$ represents the instantaneous rate of change of $f(x)$ at $x = a$.

Example 1

Suppose that the amount of water in a holding tank at t minutes is given by $V(t) = 2t^2 - 16t + 35$. Determine each of the following.

(a) Is the volume of water in the tank increasing or decreasing at $t = 1$ minute?

(b) Is the volume of water in the tank increasing or decreasing at $t = 5$ minutes?

(c) Is the volume of water in the tank changing faster at $t = 1$ or $t = 5$ minutes?

(d) Is the volume of water in the tank ever not changing? If so, when?

Solution

In the solution to this example we will use both notations for the derivative just to get you familiar with the different notations.

We are going to need the rate of change of the volume to answer these questions. This means that we will need the derivative of this function since that will give us a formula for the rate of change at any time t. Now, notice that the function giving the volume of water in the tank is the same function that we saw in Example 1 in the last section except the letters have changed. The change in letters between the function in this example versus the function in the example from the last section won't affect the work and so we can just use the answer from that example with an appropriate change in letters.

The derivative is.

$$V'(t) = 4t - 16 \qquad \text{OR} \qquad \frac{dV}{dt} = 4t - 16$$

Recall from our work in the first limits section that we determined that if the rate of change was positive then the quantity was increasing and if the rate of change was negative then the quantity was decreasing.

We can now work the problem.

(a) Is the volume of water in the tank increasing or decreasing at $t = 1$ minute?

In this case all that we need is the rate of change of the volume at $t = 1$ or,

$$V'(1) = -12 \qquad \text{OR} \qquad \left.\frac{dV}{dt}\right|_{t=1} = -12$$

So, at $t = 1$ the rate of change is negative and so the volume must be decreasing at this time.

(b) Is the volume of water in the tank increasing or decreasing at $t = 5$ minutes?

Again, we will need the rate of change at $t = 5$.

$$V'(5) = 4 \qquad \text{OR} \qquad \left.\frac{dV}{dt}\right|_{t=5} = 4$$

In this case the rate of change is positive and so the volume must be increasing at $t = 5$.

(c) Is the volume of water in the tank changing faster at $t = 1$ or $t = 5$ minutes?

To answer this question all that we look at is the size of the rate of change and we don't worry about the sign of the rate of change. All that we need to know here is that the larger the number the faster the rate of change. So, in this case the volume is changing faster at $t = 1$ than at $t = 5$.

(d) Is the volume of water in the tank ever not changing? If so, when?

The volume will not be changing if it has a rate of change of zero. In order to have a rate of change of zero this means that the derivative must be zero. So, to answer this question we will then need to solve

$$V'(t) = 0 \qquad \text{OR} \qquad \frac{dV}{dt} = 0$$

This is easy enough to do.

$$4t - 16 = 0 \qquad \Rightarrow \qquad t = 4$$

So at $t = 4$ the volume isn't changing. Note that all this is saying is that for a brief instant the volume isn't changing. It doesn't say that at this point the volume will quit changing permanently.

If we go back to our answers from parts (a) and (b) we can get an idea about what is going on. At $t = 1$ the volume is decreasing and at $t = 5$ the volume is increasing. So, at some point in time the volume needs to switch from decreasing to increasing. That time is $t = 4$.

This is the time in which the volume goes from decreasing to increasing and so for the briefest instant in time the volume will quit changing as it changes from decreasing to increasing.

Note that one of the more common mistakes that students make in these kinds of problems is to try and determine increasing/decreasing from the function values rather than the derivatives. In this case if we took the function values at $t = 0$, $t = 1$ and $t = 5$ we would get,

$$V(0) = 35 \qquad V(1) = 21 \qquad V(5) = 5$$

Clearly as we go from $t = 0$ to $t = 1$ the volume has decreased. This might lead us to decide that AT $t = 1$ the volume is decreasing. However, we just can't say that. All we can say is that between $t = 0$ and $t = 1$ the volume has decreased at some point in time. The only way to know what is happening right at $t = 1$ is to compute $V'(1)$ and look at its sign to determine increasing/decreasing. In this case $V'(1)$ is negative and so the volume really is decreasing at $t = 1$.

Now, if we'd plugged into the function rather than the derivative we would have gotten the correct answer for $t = 1$ even though our reasoning would have been wrong. It's important to not let this give you the idea that this will always be the case. It just happened to work out in the case of $t = 1$.

To see that this won't always work let's now look at $t = 5$. If we plug $t = 1$ and $t = 5$ into the volume we can see that again as we go from $t = 1$ to $t = 5$ the volume has decreased. Again, however all this says is that the volume HAS decreased somewhere between $t = 1$ and $t = 5$. It does NOT say that the volume is decreasing at $t = 5$. The only way to know what is going on right at $t = 5$ is to compute $V'(5)$ and in this case $V'(5)$ is positive and so the volume is actually increasing at $t = 5$.

So, be careful. When asked to determine if a function is increasing or decreasing at a point make sure and look at the derivative. It is the only sure way to get the correct answer. We are not looking to determine is the function has increased/decreased by the time we reach a particular point. We are looking to determine if the function is increasing/decreasing at that point in question.

Slope of Tangent Line

This is the next major interpretation of the derivative. The slope of the tangent line to $f(x)$ at $x = a$ is $f'(a)$. The tangent line then is given by,

$$y = f(a) + f'(a)(x - a)$$

Example 2

Find the tangent line to the following function at $z = 3$.

$$R(z) = \sqrt{5z - 8}$$

Solution

We first need the derivative of the function and we found that in Example 3 in the last section. The derivative is,

$$R'(z) = \frac{5}{2\sqrt{5z - 8}}$$

Now all that we need is the function value and derivative (for the slope) at $z = 3$.

$$R(3) = \sqrt{7} \qquad m = R'(3) = \frac{5}{2\sqrt{7}}$$

The tangent line is then,

$$y = \sqrt{7} + \frac{5}{2\sqrt{7}}(z - 3)$$

Velocity

Recall that this can be thought of as a special case of the rate of change interpretation. If the position of an object is given by $f(t)$ after t units of time the velocity of the object at $t = a$ is given by $f'(a)$.

Example 3

Suppose that the position of an object after t hours is given by,

$$g(t) = \frac{t}{t+1}$$

Answer both of the following about this object.

(a) Is the object moving to the right or the left at $t = 10$ hours?

(b) Does the object ever stop moving?

Solution

Once again, we need the derivative and we found that in Example 2 in the last section. The derivative is,

$$g'(t) = \frac{1}{(t+1)^2}$$

(a) Is the object moving to the right or the left at $t = 10$ hours?

To determine if the object is moving to the right (velocity is positive) or left (velocity is negative) we need the derivative at $t = 10$.

$$g'(10) = \frac{1}{121}$$

So, the velocity at $t = 10$ is positive and so the object is moving to the right at $t = 10$.

(b) Does the object ever stop moving?

The object will stop moving if the velocity is ever zero. However, note that the only way a rational expression will ever be zero is if the numerator is zero. Since the numerator of the derivative (and hence the speed) is a constant it can't be zero.

Therefore, the object will never stop moving.

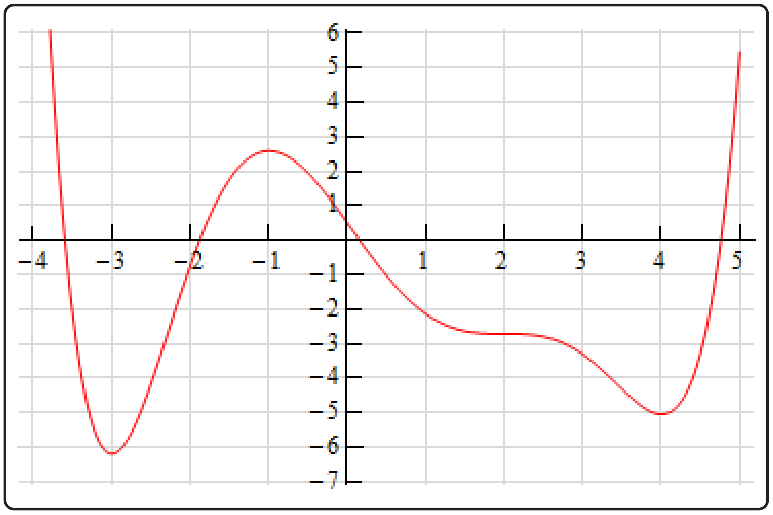
In fact, we can say a little more here. The object will always be moving to the right since the velocity is always positive.

We've seen three major interpretations of the derivative here. You will need to remember these, especially the rate of change, as they will show up continually throughout this course.

Before we leave this section let's work one more example that encompasses some of the ideas discussed here and is just a nice example to work.

Example 4

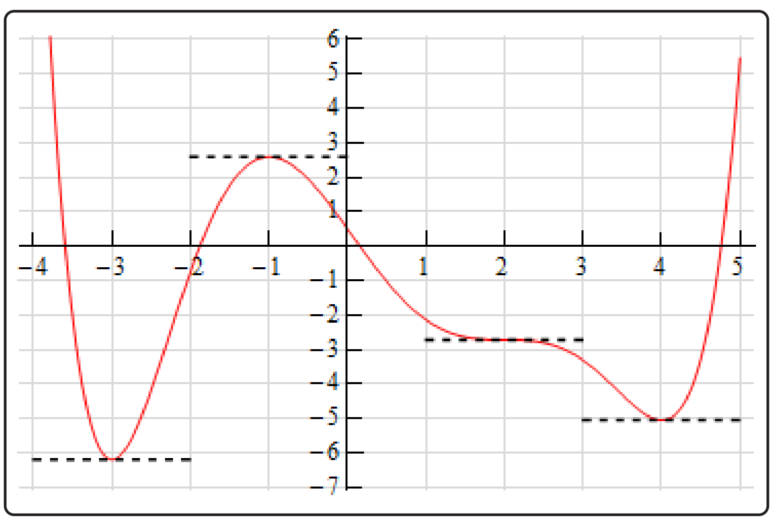
Below is the sketch of a function $f(x)$. Sketch the graph of the derivative of this function, $f'(x)$.



Solution

At first glance this seems to an all but impossible task. However, if you have some basic knowledge of the interpretations of the derivative you can get a sketch of the derivative. It will not be a perfect sketch for the most part, but you should be able to get most of the basic features of the derivative in the sketch.

Let's start off with the following sketch of the function with a couple of additions.



Notice that at $x = -3$, $x = -1$, $x = 2$ and $x = 4$ the tangent line to the function is horizontal. This means that the slope of the tangent line must be zero. Now, we know that the slope of the tangent line at a particular point is also the value of the derivative of the function at that point. Therefore, we now know that,

$$f'(-3) = 0 \qquad f'(-1) = 0 \qquad f'(2) = 0 \qquad f'(4) = 0$$

This is a good starting point for us. It gives us a few points on the graph of the derivative. It also breaks the domain of the function up into regions where the function is increasing and decreasing. We know, from our discussions above, that if the function is increasing at a point then the derivative must be positive at that point. Likewise, we know that if the function is decreasing at a point then the derivative must be negative at that point.

We can now give the following information about the derivative.

$$\begin{array}{lr} x < -3 & f'(x) < 0 \\ -3 < x < -1 & f'(x) > 0 \\ -1 < x < 2 & f'(x) < 0 \\ 2 < x < 4 & f'(x) < 0 \\ x > 4 & f'(x) > 0 \end{array}$$

Remember that we are giving the signs of the derivatives here and these are solely a function of whether the function is increasing or decreasing. The sign of the function itself is completely immaterial here and will not in any way effect the sign of the derivative.

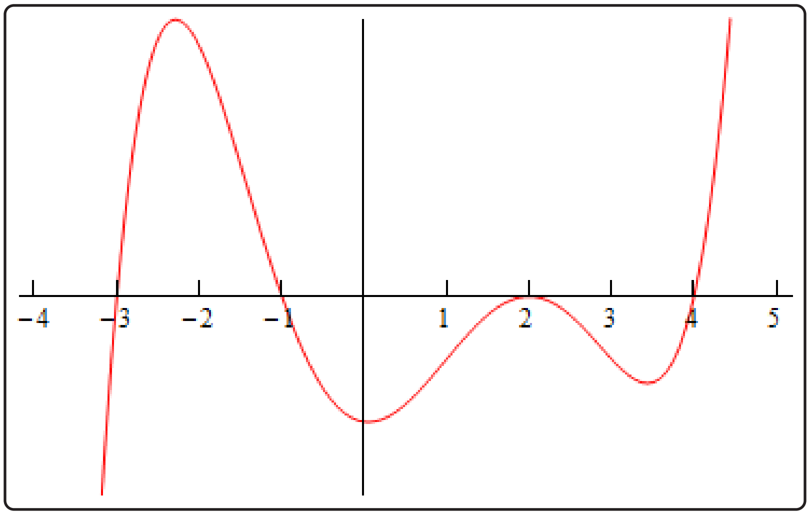
This may still seem like we don't have enough information to get a sketch, but we can get a little bit more information about the derivative from the graph of the function. In the range $x < -3$ we know that the derivative must be negative, however we can also see that the derivative needs to be increasing in this range. It is negative here until we reach $x = -3$ and at this point the derivative must be zero. The only way for the derivative to be negative to the left of $x = -3$ and zero at $x = -3$ is for the derivative to increase as we increase x towards $x = -3$.

Now, in the range $-3 < x < -1$ we know that the derivative must be zero at the endpoints and positive in between the two endpoints. Directly to the right of $x = -3$ the derivative must also be increasing (because it starts at zero and then goes positive - therefore it must be increasing). So, the derivative in this range must start out increasing and must eventually get back to zero at $x = -1$. So, at some point in this interval the derivative must start decreasing before it reaches $x = -1$. Now, we have to be careful here because this is just general behavior here at the two endpoints. We won't know where the derivative goes from increasing to decreasing and it may well change between increasing and decreasing several times before we reach $x = -1$. All we can really say is that immediately to the right of $x = -3$ the derivative will be increasing and immediately to the left of $x = -1$ the derivative will be decreasing.

Next, for the ranges $-1 < x < 2$ and $2 < x < 4$ we know the derivative will be zero at the endpoints and negative in between. Also, following the type of reasoning given above we can see in each of these ranges that the derivative will be decreasing just to the right of the left-hand endpoint and increasing just to the left of the right hand endpoint.

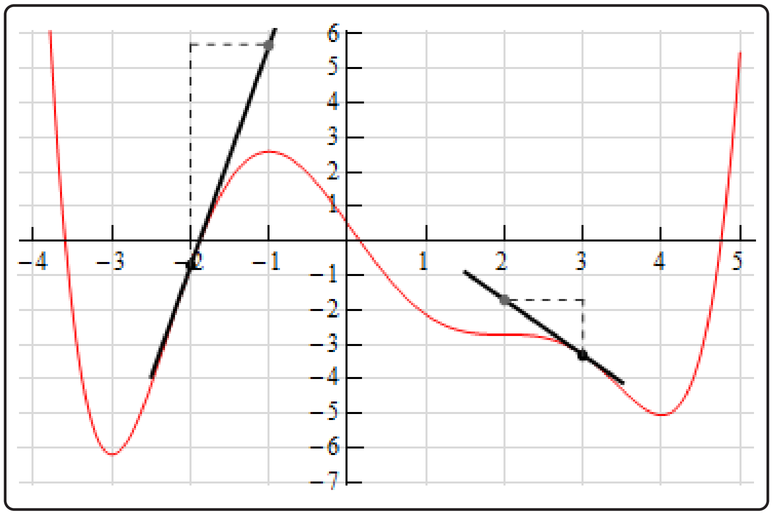
Finally, in the last region $x > 4$ we know that the derivative is zero at $x = 4$ and positive to the right of $x = 4$. Once again, following the reasoning above, the derivative must also be increasing in this range.

Putting all of this material together (and always taking the simplest choices for increasing and/or decreasing information) gives us the following sketch for the derivative.



Note that this was done with the actual derivative and so is in fact accurate. Any sketch you do will probably not look quite the same. The "humps" in each of the regions may be at different places and/or different heights for example. Also, note that we left off the vertical scale because given the information that we've got at this point there was no real way to know this information.

That doesn't mean however that we can't get some ideas of specific points on the derivative other than where we know the derivative to be zero. To see this let's check out the following graph of the function (not the derivative, but the function).

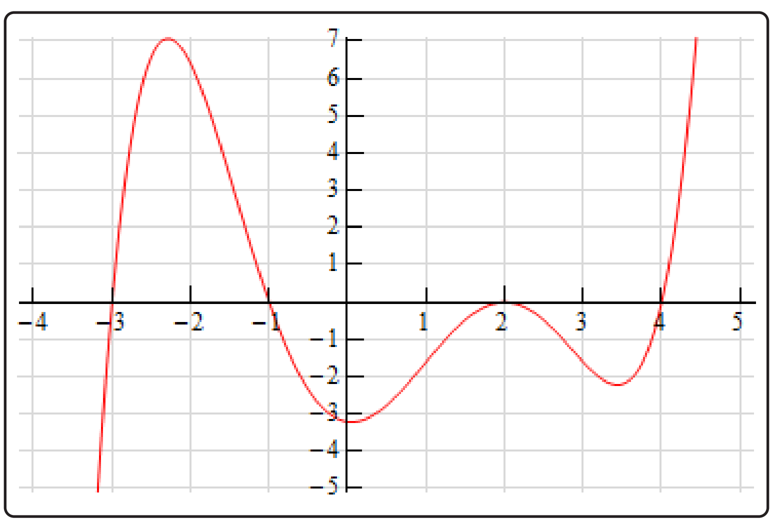


At $x = -2$ and $x = 3$ we've sketched in a couple of tangent lines. We can use the basic rise/run slope concept to estimate the value of the derivative at these points.

Let's start at $x = 3$. We've got two points on the line here. We can see that each seem to be about one-quarter of the way off the grid line. So, taking that into account and the fact that we go through one complete grid we can see that the slope of the tangent line, and hence the derivative, is approximately -1.5.

At $x = -2$ it looks like (with some heavy estimation) that the second point is about 6.5 grids above the first point and so the slope of the tangent line here, and hence the derivative, is approximately 6.5.

Here is the sketch of the derivative with the vertical scale included and from this we can see that in fact our estimates are pretty close to reality.



Note that this idea of estimating values of derivatives can be a tricky process and does require a fair amount of (possible bad) approximations so while it can be used, you need to be careful with it.

We'll close out this section by noting that while we're not going to include an example here we could also use the graph of the derivative to give us a sketch of the function itself. In fact, in the next chapter where we discuss some applications of the derivative we will be looking using information the derivative gives us to sketch the graph of a function.

3.3 Differentiation Formulas

In the first section of this chapter we saw the definition of the derivative and we computed a couple of derivatives using the definition. As we saw in those examples there was a fair amount of work involved in computing the limits and the functions that we worked with were not terribly complicated.

For more complex functions using the definition of the derivative would be an almost impossible task. Luckily for us we won't have to use the definition terribly often. We will have to use it on occasion, however we have a large collection of formulas and properties that we can use to simplify our life considerably and will allow us to avoid using the definition whenever possible.

We will introduce most of these formulas over the course of the next several sections. We will start in this section with some of the basic properties and formulas. We will give the properties and formulas in this section in both "prime" notation and "fraction" notation.

Properties

1. $(f(x) \pm g(x))' = f'(x) \pm g'(x)$ OR $\dfrac{d}{dx}(f(x) \pm g(x)) = \dfrac{df}{dx} \pm \dfrac{dg}{dx}$

   In other words, to differentiate a sum or difference all we need to do is differentiate the individual terms and then put them back together with the appropriate signs. Note as well that this property is not limited to two functions.

   See the Proof of Various Derivative Formulas section of the Extras appendix to see the proof of this property. It's a very simple proof using the definition of the derivative.

2. $(cf(x))' = cf'(x)$ OR $\dfrac{d}{dx}(cf(x)) = c\dfrac{df}{dx}$, c is any number

   In other words, we can "factor" a multiplicative constant out of a derivative if we need to. See the Proof of Various Derivative Formulas section of the Extras appendix to see the proof of this property.

Note that we have not included formulas for the derivative of products or quotients of two functions here. The derivative of a product or quotient of two functions is not the product or quotient of the derivatives of the individual pieces. We will take a look at these in the next section.

Next, let's take a quick look at a couple of basic "computation" formulas that will allow us to actually compute some derivatives.

Formulas

1. If $f(x) = c$ then $f'(x) = 0$ OR $\frac{d}{dx}(c) = 0$

   The derivative of a constant is zero. See the Proof of Various Derivative Formulas section of the Extras appendix to see the proof of this formula.

2. If $f(x) = x^n$ then $f'(x) = nx^{n-1}$ OR $\frac{d}{dx}(x^n) = nx^{n-1}$, n is any number.

   This formula is sometimes called the **power rule**. All we are doing here is bringing the original exponent down in front and multiplying and then subtracting one from the original exponent.

   Note as well that in order to use this formula n must be a number, it can't be a variable. Also note that the base, the x, must be a variable, it can't be a number. It will be tempting in some later sections to misuse the Power Rule when we run in some functions where the exponent isn't a number and/or the base isn't a variable.

   See the Proof of Various Derivative Formulas section of the Extras appendix to see the proof of this formula. There are actually three different proofs in this section. The first two restrict the formula to n being an integer because at this point that is all that we can do at this point. The third proof is for the general rule but does suppose that you've read most of this chapter.

These are the only properties and formulas that we'll give in this section. Let's compute some derivatives using these properties.

Example 1

Differentiate each of the following functions.

(a) $f(x) = 15x^{100} - 3x^{12} + 5x - 46$

(b) $g(t) = 2t^6 + 7t^{-6}$

(c) $y = 8z^3 - \frac{1}{3z^5} + z - 23$

(d) $T(x) = \sqrt{x} + 9\sqrt[3]{x^7} - \frac{2}{\sqrt[5]{x^2}}$

(e) $h(x) = x^{\pi} - x^{\sqrt{2}}$

Solution

(a) $f(x) = 15x^{100} - 3x^{12} + 5x - 46$

In this case we have the sum and difference of four terms and so we will differentiate each of the terms using the first property from above and then put them back together with the proper sign. Also, for each term with a multiplicative constant remember that all we need to do is "factor" the constant out (using the second property) and then do the derivative.

$$\begin{aligned} f'(x) &= 15(100)x^{99} - 3(12)x^{11} + 5(1)x^0 - 0 \\ &= 1500x^{99} - 36x^{11} + 5 \end{aligned}$$

Notice that in the third term the exponent was a one and so upon subtracting 1 from the original exponent we get a new exponent of zero. Now recall that $x^0 = 1$. Don't forget to do any basic arithmetic that needs to be done such as any multiplication and/or division in the coefficients.

(b) $g(t) = 2t^6 + 7t^{-6}$

The point of this problem is to make sure that you deal with negative exponents correctly. Here is the derivative.

$$\begin{aligned} g'(t) &= 2(6)t^5 + 7(-6)t^{-7} \\ &= 12t^5 - 42t^{-7} \end{aligned}$$

Make sure that you correctly deal with the exponents in these cases, especially the negative exponents. It is an easy mistake to "go the other way" when subtracting one off from a negative exponent and get $-6t^{-5}$ instead of the correct $-6t^{-7}$.

(c) $y = 8z^3 - \dfrac{1}{3z^5} + z - 23$

Now in this function the second term is not correctly set up for us to use the power rule. The power rule requires that the term be a variable to a power only and the term must be in the numerator. So, prior to differentiating we first need to rewrite the second term into a form that we can deal with.

$$y = 8z^3 - \frac{1}{3}z^{-5} + z - 23$$

Note that we left the 3 in the denominator and only moved the variable up to the numerator. Remember that the only thing that gets an exponent is the term that is immediately to the left of the exponent. If we'd wanted the three to come up as well we'd have written,

$$\frac{1}{(3z)^5}$$

so be careful with this! It's a very common mistake to bring the 3 up into the numerator as well at this stage.

Now that we've gotten the function rewritten into a proper form that allows us to use the Power Rule we can differentiate the function. Here is the derivative for this part.

$$y' = 24z^2 + \frac{5}{3}z^{-6} + 1$$

(d) $T(x) = \sqrt{x} + 9\sqrt[3]{x^7} - \frac{2}{\sqrt[5]{x^2}}$

All of the terms in this function have roots in them. In order to use the power rule we need to first convert all the roots to fractional exponents. Again, remember that the Power Rule requires us to have a variable to a number and that it must be in the numerator of the term. Here is the function written in "proper" form.

$$\begin{aligned} T(x) &= x^{\frac{1}{2}} + 9\left(x^7\right)^{\frac{1}{3}} - \frac{2}{\left(x^2\right)^{\frac{1}{5}}} \\ &= x^{\frac{1}{2}} + 9x^{\frac{7}{3}} - \frac{2}{x^{\frac{2}{5}}} \\ &= x^{\frac{1}{2}} + 9x^{\frac{7}{3}} - 2x^{-\frac{2}{5}} \end{aligned}$$

In the last two terms we combined the exponents. You should always do this with this kind of term. In a later section we will learn of a technique that would allow us to differentiate this term without combining exponents, however it will take significantly more work to do. Also, don't forget to move the term in the denominator of the third term up to the numerator. We can now differentiate the function.

$$\begin{aligned} T'(x) &= \frac{1}{2}x^{-\frac{1}{2}} + 9\left(\frac{7}{3}\right)x^{\frac{4}{3}} - 2\left(-\frac{2}{5}\right)x^{-\frac{7}{5}} \\ &= \frac{1}{2}x^{-\frac{1}{2}} + \frac{63}{3}x^{\frac{4}{3}} + \frac{4}{5}x^{-\frac{7}{5}} \end{aligned}$$

Make sure that you can deal with fractional exponents. You will see a lot of them in this class.

(e) $h(x) = x^{\pi} - x^{\sqrt{2}}$

In all of the previous examples the exponents have been nice integers or fractions. That is usually what we'll see in this class. However, the exponent only needs to be a number so don't get excited about problems like this one. They work exactly the same.

$$h'(x) = \pi x^{\pi - 1} - \sqrt{2}x^{\sqrt{2} - 1}$$

The answer is a little messy and we won't reduce the exponents down to decimals. However, this problem is not terribly difficult it just looks that way initially.

There is a general rule about derivatives in this class that you will need to get into the habit of using. When you see radicals you should always first convert the radical to a fractional exponent and then simplify exponents as much as possible. Following this rule will save you a lot of grief in the future.

Back when we first put down the properties we noted that we hadn't included a property for products and quotients. That doesn't mean that we can't differentiate any product or quotient at this point. There are some that we can do.

Example 2

Differentiate each of the following functions.

(a) $y = \sqrt[3]{x^2}\left(2x - x^2\right)$

(b) $h\left(t\right) = \dfrac{2t^5 + t^2 - 5}{t^2}$

Solution

(a) $y = \sqrt[3]{x^2}\left(2x - x^2\right)$

In this function we can't just differentiate the first term, differentiate the second term and then multiply the two back together. That just won't work. We will discuss this in detail in the next section so if you're not sure you believe that hold on for a bit and we'll be looking at that soon as well as showing you an example of why it won't work.

It is still possible to do this derivative however. All that we need to do is convert the radical to fractional exponents (as we should anyway) and then multiply this through the parenthesis.

$$y = x^{\frac{2}{3}}\left(2x - x^2\right) = 2x^{\frac{5}{3}} - x^{\frac{8}{3}}$$

Now we can differentiate the function.

$$y' = \frac{10}{3}x^{\frac{2}{3}} - \frac{8}{3}x^{\frac{5}{3}}$$

(b) $h\left(t\right) = \dfrac{2t^5 + t^2 - 5}{t^2}$

As with the first part we can't just differentiate the numerator and the denominator and the put it back together as a fraction. Again, if you're not sure you believe this hold on until the next section and we'll take a more detailed look at this.

We can simplify this rational expression however as follows.

$$h\left(t\right) = \frac{2t^5}{t^2} + \frac{t^2}{t^2} - \frac{5}{t^2} = 2t^3 + 1 - 5t^{-2}$$

This is a function that we can differentiate.

$$h'(t) = 6t^2 + 10t^{-3}$$

So, as we saw in this example there are a few products and quotients that we can differentiate. If we can first do some simplification the functions will sometimes simplify into a form that can be differentiated using the properties and formulas in this section.

Before moving on to the next section let's work a couple of examples to remind us once again of some of the interpretations of the derivative.

Example 3

Is $f(x) = 2x^3 + \dfrac{300}{x^3} + 4$ increasing, decreasing or not changing at $x = -2$?

Solution

We know that the rate of change of a function is given by the functions derivative so all we need to do is it rewrite the function (to deal with the second term) and then take the derivative.

$$f(x) = 2x^3 + 300x^{-3} + 4 \qquad \Rightarrow \qquad f'(x) = 6x^2 - 900x^{-4} = 6x^2 - \frac{900}{x^4}$$

Note that we rewrote the last term in the derivative back as a fraction. This is not something we've done to this point and is only being done here to help with the evaluation in the next step. It's often easier to do the evaluation with positive exponents.

So, upon evaluating the derivative we get

$$f'(-2) = 6(4) - \frac{900}{16} = -\frac{129}{4} = -32.25$$

So, at $x = -2$ the derivative is negative and so the function is decreasing at $x = -2$.

Example 4

Find the equation of the tangent line to $f(x) = 4x - 8\sqrt{x}$ at $x = 16$.

Solution

We know that the equation of a tangent line is given by,

$$y = f(a) + f'(a)(x - a)$$

So, we will need the derivative of the function (don't forget to get rid of the radical).

$$f(x) = 4x - 8x^{\frac{1}{2}} \qquad \Rightarrow \qquad f'(x) = 4 - 4x^{-\frac{1}{2}} = 4 - \frac{4}{x^{\frac{1}{2}}}$$

Again, notice that we eliminated the negative exponent in the derivative solely for the sake of the evaluation. All we need to do then is evaluate the function and the derivative at the point in question, $x = 16$.

$$f(16) = 64 - 8(4) = 32 \qquad f'(16) = 4 - \frac{4}{4} = 3$$

The tangent line is then,

$$y = 32 + 3(x - 16) = 3x - 16$$

Example 5

The position of an object at any time t (in hours) is given by,

$$s(t) = 2t^3 - 21t^2 + 60t - 10$$

Determine when the object is moving to the right and when the object is moving to the left.

Solution

The only way that we'll know for sure which direction the object is moving is to have the velocity in hand. Recall that if the velocity is positive the object is moving off to the right and if the velocity is negative then the object is moving to the left.

We need the derivative in order to get the velocity of the object. The derivative, and hence the velocity, is,

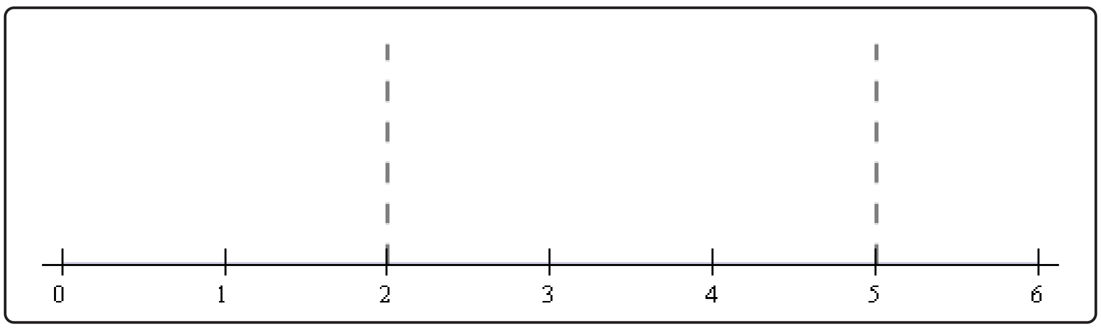
$$s'(t) = 6t^2 - 42t + 60 = 6\left(t^2 - 7t + 10\right) = 6(t-2)(t-5)$$

The reason for factoring the derivative will be apparent shortly.

Now, we need to determine where the derivative is positive and where the derivative is negative. There are several ways to do this. The method that we tend to prefer is the following.

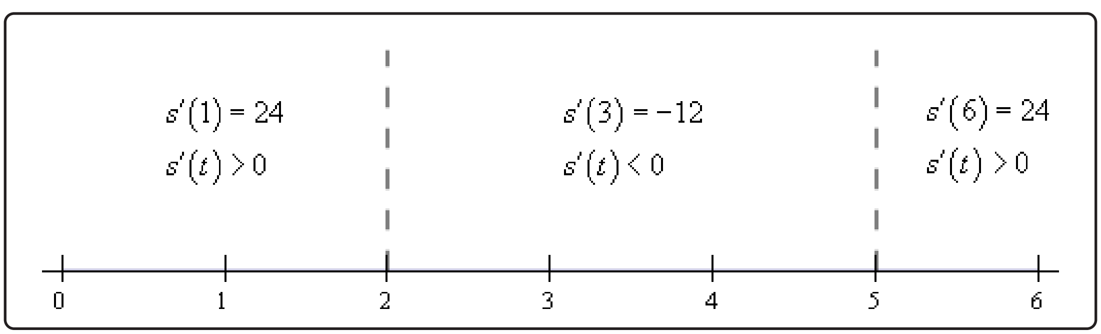
Since polynomials are continuous we know from the Intermediate Value Theorem that if the polynomial ever changes sign then it must have first gone through zero. So, if we knew where the derivative was zero we would know the only points where the derivative *might* change sign.

We can see from the factored form of the derivative that the derivative will be zero at $t = 2$ and $t = 5$. Let's graph these points on a number line.



Now, we can see that these two points divide the number line into three distinct regions. In each of these regions we **know** that the derivative will be the same sign. Recall the derivative can only change sign at the two points that are used to divide the number line up into the regions.

Therefore, all that we need to do is to check the derivative at a test point in each region and the derivative in that region will have the same sign as the test point. Here is the number line with the test points and results shown.



Here are the intervals in which the derivative is positive and negative.

$$\begin{aligned} \text{positive :} \quad & -\infty < t < 2 \ \ \& \ \ 5 < t < \infty \\ \text{negative :} \quad & 2 < t < 5 \end{aligned}$$

We included negative t's here because we could even though they may not make much sense for this problem. Once we know this we also can answer the question. The object is moving to the right and left in the following intervals.

$$\begin{aligned} \text{moving to the right :} \quad & -\infty < t < 2 \ \ \& \ \ 5 < t < \infty \\ \text{moving to the left :} \quad & 2 < t < 5 \end{aligned}$$

Make sure that you can do the kind of work that we just did in this example. You will be asked numerous times over the course of the next two chapters to determine where functions are positive and/or negative. If you need some review or want to practice these kinds of problems you should check out the Solving Inequalities section of the Algebra/Trig Review.

3.4 Product and Quotient Rule

In the previous section we noted that we had to be careful when differentiating products or quotients. It's now time to look at products and quotients and see why.

First let's take a look at why we have to be careful with products and quotients. Suppose that we have the two functions $f(x) = x^3$ and $g(x) = x^6$. Let's start by computing the derivative of the product of these two functions. This is easy enough to do directly.

$$(f\,g)' = \left(x^3 x^6\right)' = \left(x^9\right)' = 9x^8$$

Remember that on occasion we will drop the (x) part on the functions to simplify notation somewhat. We've done that in the work above.

Now, let's try the following.

$$f'(x)\,g'(x) = \left(3x^2\right)\left(6x^5\right) = 18x^7$$

So, we can very quickly see that.

$$(f\,g)' \neq f'\,g'$$

In other words, the derivative of a product is not the product of the derivatives.

Using the same functions we can do the same thing for quotients.

$$\left(\frac{f}{g}\right)' = \left(\frac{x^3}{x^6}\right)' = \left(\frac{1}{x^3}\right)' = \left(x^{-3}\right)' = -3x^{-4} = -\frac{3}{x^4}$$

$$\frac{f'(x)}{g'(x)} = \frac{3x^2}{6x^5} = \frac{1}{2x^3}$$

So, again we can see that,

$$\left(\frac{f}{g}\right)' \neq \frac{f'}{g'}$$

To differentiate products and quotients we have the **Product Rule** and the **Quotient Rule**.

Product Rule

If the two functions $f(x)$ and $g(x)$ are differentiable (*i.e.* the derivative exist) then the product is differentiable and,

$$\big(f(x)\,g(x)\big)' = f'(x)\,g(x) + f(x)\,g'(x)$$

The proof of the Product Rule is shown in the Proof of Various Derivative Formulas section of the Extras appendix.

Quotient Rule

If the two functions $f(x)$ and $g(x)$ are differentiable (*i.e.* the derivative exist) then the quotient is differentiable and,

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)\,g(x) - f(x)\,g'(x)}{\left[g(x)\right]^2}$$

Note that the numerator of the quotient rule is very similar to the product rule so be careful to not mix the two up!

The proof of the Quotient Rule is shown in the Proof of Various Derivative Formulas section of the Extras appendix.

Let's do a couple of examples of the product rule.

Example 1

Differentiate each of the following functions.

(a) $y = \sqrt[3]{x^2}\left(2x - x^2\right)$

(b) $f(x) = \left(6x^3 - x\right)(10 - 20x)$

Solution

At this point there really aren't a lot of reasons to use the product rule. As we noted in the previous section all we would need to do for either of these is to just multiply out the product and then differentiate.

With that said we will use the product rule on these so we can see an example or two. As we add more functions to our repertoire and as the functions become more complicated the product rule will become more useful and in many cases required.

(a) $y = \sqrt[3]{x^2}\left(2x - x^2\right)$

Note that we took the derivative of this function in the previous section and didn't use the product rule at that point. We should however get the same result here as we did then.

Now let's do the problem here. There's not really a lot to do here other than use the product rule. However, before doing that we should convert the radical to a fractional exponent as always.

$$y = x^{\frac{2}{3}}\left(2x - x^2\right)$$

Now let's take the derivative. So, we take the derivative of the first function times the second then add on to that the first function times the derivative of the second function.

$$y' = \frac{2}{3}x^{-\frac{1}{3}}\left(2x - x^2\right) + x^{\frac{2}{3}}\left(2 - 2x\right)$$

This is NOT what we got in the previous section for this derivative. However, with some simplification we can arrive at the same answer.

$$y' = \frac{4}{3}x^{\frac{2}{3}} - \frac{2}{3}x^{\frac{5}{3}} + 2x^{\frac{2}{3}} - 2x^{\frac{5}{3}} = \frac{10}{3}x^{\frac{2}{3}} - \frac{8}{3}x^{\frac{5}{3}}$$

This is what we got for an answer in the previous section so that is a good check of the product rule.

(b) $f\left(x\right) = \left(6x^3 - x\right)\left(10 - 20x\right)$

This one is actually easier than the previous one. Let's just run it through the product rule.

$$\begin{aligned} f'\left(x\right) &= \left(18x^2 - 1\right)\left(10 - 20x\right) + \left(6x^3 - x\right)\left(-20\right) \\ &= -480x^3 + 180x^2 + 40x - 10 \end{aligned}$$

Since it was easy to do we went ahead and simplified the results a little.

Let's now work an example or two with the quotient rule. In this case, unlike the product rule examples, a couple of these functions will require the quotient rule in order to get the derivative. For the last two however, we can avoid the quotient rule if we'd like to as we'll see.

Example 2

Differentiate each of the following functions.

(a) $W\left(z\right) = \dfrac{3z + 9}{2 - z}$

(b) $h\left(x\right) = \dfrac{4\sqrt{x}}{x^2 - 2}$

(c) $f\left(x\right) = \dfrac{4}{x^6}$

(d) $y = \dfrac{w^6}{5}$

Solution

(a) $W(z) = \dfrac{3z+9}{2-z}$

There isn't a lot to do here other than to use the quotient rule. Here is the work for this function.

$$\begin{aligned} W'(z) &= \frac{3(2-z)-(3z+9)(-1)}{(2-z)^2} \\ &= \frac{15}{(2-z)^2} \end{aligned}$$

(b) $h(x) = \dfrac{4\sqrt{x}}{x^2-2}$

Again, not much to do here other than use the quotient rule. Don't forget to convert the square root into a fractional exponent.

$$\begin{aligned} h'(x) &= \frac{4\left(\frac{1}{2}\right)x^{-\frac{1}{2}}\left(x^2-2\right)-4x^{\frac{1}{2}}(2x)}{\left(x^2-2\right)^2} \\ &= \frac{2x^{\frac{3}{2}}-4x^{-\frac{1}{2}}-8x^{\frac{3}{2}}}{\left(x^2-2\right)^2} \\ &= \frac{-6x^{\frac{3}{2}}-4x^{-\frac{1}{2}}}{\left(x^2-2\right)^2} \end{aligned}$$

(c) $f(x) = \dfrac{4}{x^6}$

It seems strange to have this one here rather than being the first part of this example given that it definitely appears to be easier than any of the previous two. In fact, it is easier. There is a point to doing it here rather than first. In this case there are two ways to do compute this derivative. There is an easy way and a hard way and in this case the hard way is the quotient rule. That's the point of this example.

Let's do the quotient rule and see what we get.

$$f'(x) = \frac{(0)\left(x^6\right)-4\left(6x^5\right)}{\left(x^6\right)^2} = \frac{-24x^5}{x^{12}} = -\frac{24}{x^7}$$

Now, that was the "hard" way. So, what was so hard about it? Well actually it wasn't that hard, there is just an easier way to do it that's all. However, having said that, a common mistake here is to do the derivative of the numerator (a constant) incorrectly. For some reason many people will give the derivative of the numerator in these kinds of problems as a 1 instead of 0! Also, there is some simplification that needs to be done in these kinds of problems if you do the quotient rule.

The easy way is to do what we did in the previous section.

$$f(x) = 4x^{-6} \qquad f'(x) = -24x^{-7} = -\frac{24}{x^7}$$

Either way will work, but I'd rather take the easier route if I had the choice.

(d) $y = \dfrac{w^6}{5}$

This problem also seems a little out of place. However, it is here again to make a point. Do not confuse this with a quotient rule problem. While you can do the quotient rule on this function there is no reason to use the quotient rule on this. Simply rewrite the function as

$$y = \frac{1}{5}w^6$$

and differentiate as always.

$$y' = \frac{6}{5}w^5$$

Finally, let's not forget about our applications of derivatives.

Example 3

Suppose that the amount of air in a balloon at any time t is given by

$$V(t) = \frac{6\sqrt[3]{t}}{4t+1}$$

Determine if the balloon is being filled with air or being drained of air at $t = 8$.

Solution

If the balloon is being filled with air then the volume is increasing and if it's being drained of air then the volume will be decreasing. In other words, we need to get the derivative so that we can determine the rate of change of the volume at $t = 8$.

This will require the quotient rule.

$$\begin{aligned} V'(t) &= \frac{2t^{-\frac{2}{3}}(4t+1) - 6t^{\frac{1}{3}}(4)}{(4t+1)^2} \\ &= \frac{-16t^{\frac{1}{3}} + 2t^{-\frac{2}{3}}}{(4t+1)^2} \\ &= \frac{-16t^{\frac{1}{3}} + \frac{2}{t^{\frac{2}{3}}}}{(4t+1)^2} \end{aligned}$$

Note that we simplified the numerator more than usual here. This was only done to make the derivative easier to evaluate.

The rate of change of the volume at $t = 8$ is then,

$$\begin{aligned} V'(8) &= \frac{-16(2) + \frac{2}{4}}{(33)^2} \qquad\qquad (8)^{\frac{1}{3}} = 2 \qquad (8)^{\frac{2}{3}} = \left((8)^{\frac{1}{3}}\right)^2 = (2)^2 = 4 \\ &= -\frac{63}{2178} = -\frac{7}{242} \end{aligned}$$

So, the rate of change of the volume at $t = 8$ is negative and so the volume must be decreasing. Therefore, air is being drained out of the balloon at $t = 8$.

As a final topic let's note that the product rule can be extended to more than two functions, for instance.

$$\begin{aligned} (f\,g\,h)' &= f'\,g\,h + f\,g'\,h + f\,g\,h' \\ (f\,g\,h\,w)' &= f'\,g\,h\,w + f\,g'\,h\,w + f\,g\,h'\,w + f\,g\,h\,w' \end{aligned}$$

Deriving these products of more than two functions is actually pretty simple. For example, let's take a look at the three function product rule.

First, we don't think of it as a product of three functions but instead of the product rule of the two functions $f\,g$ and h which we can then use the two function product rule on. Doing this gives,

$$(f\,g\,h)' = ([f\,g]\;h)' = [f\,g]'h + [f\,g]\,h'$$

Note that we put brackets on the $f\,g$ part to make it clear we are thinking of that term as a single function. Now all we need to do is use the two function product rule on the $[f\,g]'$ term and then do a little simplification.

$$(f\,g\,h)' = \left[f'\,g + f\,g'\right]h + [f\,g]\,h' = f'\,g\,h + f\,g'\,h + f\,g\,h'$$

Any product rule with more functions can be derived in a similar fashion.

With this section and the previous section we are now able to differentiate powers of x as well as sums, differences, products and quotients of these kinds of functions. However, there are many more functions out there in the world that are not in this form. The next few sections give many of these functions as well as give their derivatives.

3.5 Derivatives of Trig Functions

With this section we're going to start looking at the derivatives of functions other than polynomials or roots of polynomials. We'll start this process off by taking a look at the derivatives of the six trig functions. Two of the derivatives will be derived. The remaining four are left to you and will follow similar proofs for the two given here.

Before we actually get into the derivatives of the trig functions we need to give a couple of limits that will show up in the derivation of two of the derivatives.

Fact

$$\lim_{\theta \to 0} \frac{\sin \theta}{\theta} = 1 \qquad\qquad \lim_{\theta \to 0} \frac{\cos \theta - 1}{\theta} = 0$$

See the Proof of Trig Limits section of the Extras appendix to see the proof of these two limits.

Before proceeding a quick note. Students often ask why we always use radians in a Calculus class. This is the reason why! The proof of the formula involving sine above requires the angles to be in radians. If the angles are in degrees the limit involving sine is not 1 and so the formulas we will derive below would also change. The formulas below would pick up an extra constant that would just get in the way of our work and so we use radians to avoid that. So, remember to always use radians in a Calculus class!

Before we start differentiating trig functions let's work a quick set of limit problems that this fact now allows us to do.

Example 1

Evaluate each of the following limits.

(a) $\displaystyle\lim_{\theta \to 0} \frac{\sin \theta}{6\theta}$

(b) $\displaystyle\lim_{x \to 0} \frac{\sin(6x)}{x}$

(c) $\displaystyle\lim_{x \to 0} \frac{x}{\sin(7x)}$

(d) $\displaystyle\lim_{t \to 0} \frac{\sin(3t)}{\sin(8t)}$

(e) $\displaystyle\lim_{x \to 4} \frac{\sin(x-4)}{x-4}$

(f) $\displaystyle\lim_{z\to 0} \frac{\cos(2z)-1}{z}$

Solution

(a) $\displaystyle\lim_{\theta\to 0} \frac{\sin\theta}{6\theta}$

There really isn't a whole lot to this limit. In fact, it's only here to contrast with the next example so you can see the difference in how these work. In this case since there is only a 6 in the denominator we'll just factor this out and then use the fact.

$$\lim_{\theta\to 0} \frac{\sin\theta}{6\theta} = \frac{1}{6}\lim_{\theta\to 0} \frac{\sin\theta}{\theta} = \frac{1}{6}(1) = \frac{1}{6}$$

(b) $\displaystyle\lim_{x\to 0} \frac{\sin(6x)}{x}$

Now, in this case we can't factor the 6 out of the sine so we're stuck with it there and we'll need to figure out a way to deal with it. To do this problem we need to notice that in the fact the argument of the sine is the same as the denominator (*i.e.* both θ's). So we need to get both of the argument of the sine and the denominator to be the same. We can do this by multiplying the numerator and the denominator by 6 as follows.

$$\lim_{x\to 0} \frac{\sin(6x)}{x} = \lim_{x\to 0} \frac{6\sin(6x)}{6x} = 6\lim_{x\to 0} \frac{\sin(6x)}{6x}$$

Note that we factored the 6 in the numerator out of the limit. At this point, while it may not look like it, we can use the fact above to finish the limit.

To see that we can use the fact on this limit let's do a **change of variables**. A change of variables is really just a renaming of portions of the problem to make something look more like something we know how to deal with. They can't always be done, but sometimes, such as this case, they can simplify the problem. The change of variables here is to let $\theta = 6x$ and then notice that as $x \to 0$ we also have $\theta \to 6(0) = 0$. When doing a change of variables in a limit we need to change all the x's into θ's and that includes the one in the limit.

Doing the change of variables on this limit gives,

$$\begin{aligned} \lim_{x\to 0} \frac{\sin(6x)}{x} &= 6\lim_{x\to 0} \frac{\sin(6x)}{6x} \qquad\qquad \text{let}\theta = 6x \\ &= 6\lim_{\theta\to 0} \frac{\sin(\theta)}{\theta} \\ &= 6(1) \\ &= 6 \end{aligned}$$

And there we are. Note that we didn't really need to do a change of variables here. All we really need to notice is that the argument of the sine is the same as the denominator and then we can use the fact. A change of variables, in this case, is really only needed to make it clear that the fact does work.

(c) $\displaystyle\lim_{x\to 0}\frac{x}{\sin(7x)}$

In this case we appear to have a small problem in that the function we're taking the limit of here is upside down compared to that in the fact. This is not the problem it appears to be once we notice that,

$$\frac{x}{\sin(7x)}=\frac{1}{\frac{\sin(7x)}{x}}$$

and then all we need to do is recall a nice property of limits that allows us to do ,

$$\begin{aligned}\lim_{x\to 0}\frac{x}{\sin(7x)}&=\lim_{x\to 0}\frac{1}{\frac{\sin(7x)}{x}}\\&=\frac{\lim\limits_{x\to 0}1}{\lim\limits_{x\to 0}\frac{\sin(7x)}{x}}\\&=\frac{1}{\lim\limits_{x\to 0}\frac{\sin(7x)}{x}}\end{aligned}$$

With a little rewriting we can see that we do in fact end up needing to do a limit like the one we did in the previous part. So, let's do the limit here and this time we won't bother with a change of variable to help us out. All we need to do is multiply the numerator and denominator of the fraction in the denominator by 7 to get things set up to use the fact. Here is the work for this limit.

$$\begin{aligned}\lim_{x\to 0}\frac{x}{\sin(7x)}&=\frac{1}{\lim\limits_{x\to 0}\frac{7\sin(7x)}{7x}}\\&=\frac{1}{7\lim\limits_{x\to 0}\frac{\sin(7x)}{7x}}\\&=\frac{1}{(7)(1)}\\&=\frac{1}{7}\end{aligned}$$

(d) $\displaystyle\lim_{t\to 0}\frac{\sin(3t)}{\sin(8t)}$

This limit looks nothing like the limit in the fact, however it can be thought of as a combination of the previous two parts by doing a little rewriting. First, we'll split the fraction up as follows,

$$\lim_{t\to 0}\frac{\sin(3t)}{\sin(8t)} = \lim_{t\to 0}\frac{\sin(3t)}{1}\frac{1}{\sin(8t)}$$

Now, the fact wants a t in the denominator of the first and in the numerator of the second. This is easy enough to do if we multiply the whole thing by $\frac{t}{t}$ (which is just one after all and so won't change the problem) and then do a little rearranging as follows,

$$\begin{aligned}\lim_{t\to 0}\frac{\sin(3t)}{\sin(8t)} &= \lim_{t\to 0}\frac{\sin(3t)}{1}\frac{1}{\sin(8t)}\frac{t}{t}\\ &= \lim_{t\to 0}\frac{\sin(3t)}{t}\frac{t}{\sin(8t)}\\ &= \left(\lim_{t\to 0}\frac{\sin(3t)}{t}\right)\left(\lim_{t\to 0}\frac{t}{\sin(8t)}\right)\end{aligned}$$

At this point we can see that this really is two limits that we've seen before. Here is the work for each of these and notice on the second limit that we're going to work it a little differently than we did in the previous part. This time we're going to notice that it doesn't really matter whether the sine is in the numerator or the denominator as long as the argument of the sine is the same as what's in the numerator the limit is still one.

Here is the work for this limit.

$$\begin{aligned}\lim_{t\to 0}\frac{\sin(3t)}{\sin(8t)} &= \left(\lim_{t\to 0}\frac{3\sin(3t)}{3t}\right)\left(\lim_{t\to 0}\frac{8t}{8\sin(8t)}\right)\\ &= \left(3\lim_{t\to 0}\frac{\sin(3t)}{3t}\right)\left(\frac{1}{8}\lim_{t\to 0}\frac{8t}{\sin(8t)}\right)\\ &= (3)\left(\frac{1}{8}\right)\\ &= \frac{3}{8}\end{aligned}$$

(e) $\displaystyle\lim_{x\to 4}\frac{\sin(x-4)}{x-4}$

This limit almost looks the same as that in the fact in the sense that the argument of the sine is the same as what is in the denominator. However, notice that, in the limit, x is going to 4 and not 0 as the fact requires. However, with a change of variables we can see that this limit is in fact set to use the fact above regardless.

So, let $\theta = x-4$ and then notice that as $x\to 4$ we have $\theta\to 0$. Therefore, after doing the change of variable the limit becomes,

$$\lim_{x\to 4}\frac{\sin(x-4)}{x-4} = \lim_{\theta\to 0}\frac{\sin\theta}{\theta} = 1$$

(f) $\displaystyle\lim_{z \to 0} \frac{\cos(2z) - 1}{z}$

The previous parts of this example all used the sine portion of the fact. However, we could just have easily used the cosine portion so here is a quick example using the cosine portion to illustrate this. We'll not put in much explanation here as this really does work in the same manner as the sine portion.

$$\begin{aligned}\lim_{z \to 0} \frac{\cos(2z) - 1}{z} &= \lim_{z \to 0} \frac{2(\cos(2z) - 1)}{2z} \\ &= 2 \lim_{z \to 0} \frac{\cos(2z) - 1}{2z} \\ &= 2(0) \\ 0 \end{aligned}$$

All that is required to use the fact is that the argument of the cosine is the same as the denominator.

Okay, now that we've gotten this set of limit examples out of the way let's get back to the main point of this section, differentiating trig functions.

We'll start with finding the derivative of the sine function. To do this we will need to use the definition of the derivative. It's been a while since we've had to use this, but sometimes there just isn't anything we can do about it. Here is the definition of the derivative for the sine function.

$$\frac{d}{dx}(\sin(x)) = \lim_{h \to 0} \frac{\sin(x + h) - \sin(x)}{h}$$

Since we can't just plug in $h = 0$ to evaluate the limit we will need to use the following trig formula on the first sine in the numerator.

$$\sin(x + h) = \sin(x)\cos(h) + \cos(x)\sin(h)$$

Doing this gives us,

$$\begin{aligned}\frac{d}{dx}(\sin(x)) &= \lim_{h \to 0} \frac{\sin(x)\cos(h) + \cos(x)\sin(h) - \sin(x)}{h} \\ &= \lim_{h \to 0} \frac{\sin(x)(\cos(h) - 1) + \cos(x)\sin(h)}{h} \\ &= \lim_{h \to 0} \sin(x)\frac{\cos(h) - 1}{h} + \lim_{h \to 0} \cos(x)\frac{\sin(h)}{h}\end{aligned}$$

As you can see upon using the trig formula we can combine the first and third term and then factor a sine out of that. We can then break up the fraction into two pieces, both of which can be dealt with separately.

Now, both of the limits here are limits as h approaches zero. In the first limit we have a $\sin(x)$ and in the second limit we have a $\cos(x)$. Both of these are only functions of x only and as h moves in towards zero this has no effect on the value of x. Therefore, as far as the limits are concerned, these two functions are constants and can be factored out of their respective limits. Doing this gives,

$$\frac{d}{dx}(\sin(x)) = \sin(x)\lim_{h\to 0}\frac{\cos(h)-1}{h} + \cos(x)\lim_{h\to 0}\frac{\sin(h)}{h}$$

At this point all we need to do is use the limits in the fact above to finish out this problem.

$$\frac{d}{dx}(\sin(x)) = \sin(x)(0) + \cos(x)(1) = \cos(x)$$

Differentiating cosine is done in a similar fashion. It will require a different trig formula, but other than that is an almost identical proof. The details will be left to you. When done with the proof you should get,

$$\frac{d}{dx}(\cos(x)) = -\sin(x)$$

With these two out of the way the remaining four are fairly simple to get. All the remaining four trig functions can be defined in terms of sine and cosine and these definitions, along with appropriate derivative rules, can be used to get their derivatives.

Let's take a look at tangent. Tangent is defined as,

$$\tan(x) = \frac{\sin(x)}{\cos(x)}$$

Now that we have the derivatives of sine and cosine all that we need to do is use the quotient rule on this. Let's do that.

$$\begin{aligned}\frac{d}{dx}(\tan(x)) &= \frac{d}{dx}\left(\frac{\sin(x)}{\cos(x)}\right)\\ &= \frac{\cos(x)\cos(x) - \sin(x)(-\sin(x))}{(\cos(x))^2}\\ &= \frac{\cos^2(x)+\sin^2(x)}{\cos^2(x)}\end{aligned}$$

Now, recall that $\cos^2(x)+\sin^2(x) = 1$ and if we also recall the definition of secant in terms of cosine we arrive at,

$$\begin{aligned}\frac{d}{dx}(\tan(x)) &= \frac{\cos^2(x)+\sin^2(x)}{\cos^2(x)}\\ &= \frac{1}{\cos^2(x)}\\ &= \sec^2(x)\end{aligned}$$

The remaining three trig functions are also quotients involving sine and/or cosine and so can be differentiated in a similar manner. We'll leave the details to you. Here are the derivatives of all six of the trig functions.

Derivatives of the six trig functions

$$\frac{d}{dx}\left(\sin(x)\right) = \cos(x) \qquad \frac{d}{dx}\left(\cos(x)\right) = -\sin(x)$$

$$\frac{d}{dx}\left(\tan(x)\right) = \sec^2(x) \qquad \frac{d}{dx}\left(\cot(x)\right) = -\csc^2(x)$$

$$\frac{d}{dx}\left(\sec(x)\right) = \sec(x)\tan(x) \qquad \frac{d}{dx}\left(\csc(x)\right) = -\csc(x)\cot(x)$$

At this point we should work some examples.

Example 2

Differentiate each of the following functions.

(a) $g(x) = 3\sec(x) - 10\cot(x)$

(b) $h(w) = 3w^{-4} - w^2\tan(w)$

(c) $y = 5\sin(x)\cos(x) + 4\csc(x)$

(d) $P(t) = \dfrac{\sin(t)}{3 - 2\cos(t)}$

Solution

(a) $g(x) = 3\sec(x) - 10\cot(x)$

There really isn't a whole lot to this problem. We'll just differentiate each term using the formulas from above.

$$\begin{aligned} g'(x) &= 3\sec(x)\tan(x) - 10\left(-\csc^2(x)\right) \\ &= 3\sec(x)\tan(x) + 10\csc^2(x) \end{aligned}$$

(b) $h(w) = 3w^{-4} - w^2\tan(w)$

In this part we will need to use the product rule on the second term and note that we really will need the product rule here. There is no other way to do this derivative unlike what we saw when we first looked at the product rule. When we first looked at the product rule the only functions we knew how to differentiate were polynomials and in those cases all we really needed to do was multiply them out and we could take the derivative without the product rule. We are now getting into the point where we will be forced to do the product rule at times regardless of whether or not we want to.

We will also need to be careful with the minus sign in front of the second term and make sure that it gets dealt with properly. There are two ways to deal with this. One way it to make sure that you use a set of parentheses as follows,

$$\begin{aligned} h'(w) &= -12w^{-5} - \left(2w\tan(w) + w^2\sec^2(w)\right) \\ &= -12w^{-5} - 2w\tan(w) - w^2\sec^2(w) \end{aligned}$$

Because the second term is being subtracted off of the first term then the whole derivative of the second term must also be subtracted off of the derivative of the first term. The parenthesis make this idea clear.

A potentially easier way to do this is to think of the minus sign as part of the first function in the product. Or, in other words the two functions in the product, using this idea, are $-w^2$ and $\tan(w)$. Doing this gives,

$$h'(w) = -12w^{-5} - 2w\tan(w) - w^2\sec^2(w)$$

So, regardless of how you approach this problem you will get the same derivative.

(c) $y = 5\sin(x)\cos(x) + 4\csc(x)$

As with the previous part we'll need to use the product rule on the first term. We will also think of the 5 as part of the first function in the product to make sure we deal with it correctly. Alternatively, you could make use of a set of parentheses to make sure the 5 gets dealt with properly. Either way will work, but we'll stick with thinking of the 5 as part of the first term in the product. Here's the derivative of this function.

$$\begin{aligned} y' &= 5\cos(x)\cos(x) + 5\sin(x)(-\sin(x)) - 4\csc(x)\cot(x) \\ &= 5\cos^2(x) - 5\sin^2(x) - 4\csc(x)\cot(x) \end{aligned}$$

(d) $P(t) = \dfrac{\sin(t)}{3 - 2\cos(t)}$

In this part we'll need to use the quotient rule to take the derivative.

$$\begin{aligned} P'(t) &= \frac{\cos(t)(3 - 2\cos(t)) - \sin(t)(2\sin(t))}{(3 - 2\cos(t))^2} \\ &= \frac{3\cos(t) - 2\cos^2(t) - 2\sin^2(t)}{(3 - 2\cos(t))^2} \end{aligned}$$

Be careful with the signs when differentiating the denominator. The negative sign we get from differentiating the cosine will cancel against the negative sign that is already there.

This appears to be done, but there is actually a fair amount of simplification that can yet be done. To do this we need to factor out a "-2" from the last two terms in the numerator and the make use of the fact that $\cos^2(\theta) + \sin^2(\theta) = 1$.

$$\begin{aligned} P'(t) &= \frac{3\cos(t) - 2\left(\cos^2(t) + \sin^2(t)\right)}{(3 - 2\cos(t))^2} \\ &= \frac{3\cos(t) - 2}{(3 - 2\cos(t))^2} \end{aligned}$$

As a final problem here let's not forget that we still have our standard interpretations to derivatives.

Example 3

Suppose that the amount of money in a bank account is given by

$$P(t) = 500 + 100\cos(t) - 150\sin(t)$$

where t is in years. During the first 10 years in which the account is open when is the amount of money in the account increasing?

Solution

To determine when the amount of money is increasing we need to determine when the rate of change is positive. Since we know that the rate of change is given by the derivative that is the first thing that we need to find.

$$P'(t) = -100\sin(t) - 150\cos(t)$$

Now, we need to determine where in the first 10 years this will be positive. This is equivalent to asking where in the interval $[0, 10]$ is the derivative positive. Recall that both sine and cosine are continuous functions and so the derivative is also a continuous function. The Intermediate Value Theorem then tells us that the derivative can only change sign if it first goes through zero.

So, we need to solve the following equation.

$$\begin{aligned} -100\sin(t) - 150\cos(t) &= 0 \\ 100\sin(t) &= -150\cos(t) \\ \frac{\sin(t)}{\cos(t)} &= -1.5 \\ \tan(t) &= -1.5 \end{aligned}$$

The solution to this equation is,

$$\begin{aligned} t &= 2.1588 + 2\pi n, \qquad n = 0, \pm 1, \pm 2, \ldots \\ t &= 5.3004 + 2\pi n, \qquad n = 0, \pm 1, \pm 2, \ldots \end{aligned}$$

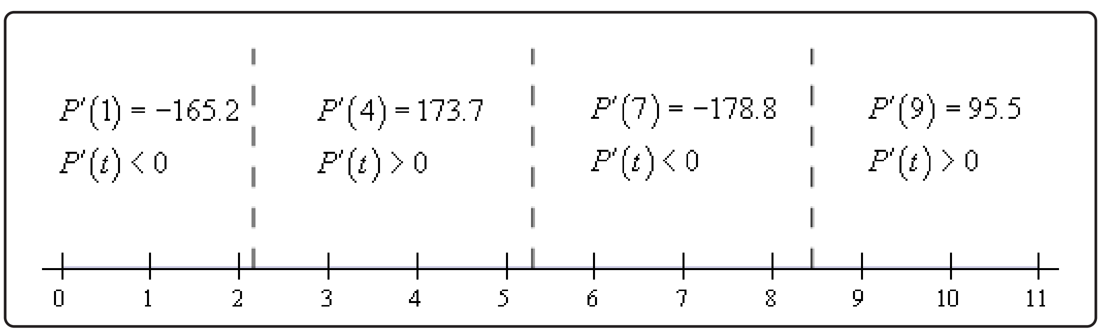
If you don't recall how to solve trig equations go back and take a look at the sections on solving trig equations in the Review chapter.

We are only interested in those solutions that fall in the range $[0, 10]$. Plugging in values of n into the solutions above we see that the values we need are,

$$\begin{aligned} t &= 2.1588 \qquad t = 2.1588 + 2\pi = 8.4420 \\ t &= 5.3004 \end{aligned}$$

So, much like solving polynomial inequalities all that we need to do is sketch in a number line and add in these points. These points will divide the number line into regions in which the derivative must always be the same sign. All that we need to do then is choose a test point from each region to determine the sign of the derivative in that region.

Here is the number line with all the information on it.



So, it looks like the amount of money in the bank account will be increasing during the following intervals.

$$2.1588 < t < 5.3004 \qquad 8.4420 < t < 10$$

Note that we can't say anything about what is happening after $t = 10$ since we haven't done any work for t's after that point.

In this section we saw how to differentiate trig functions. We also saw in the last example that our interpretations of the derivative are still valid so we can't forget those.

Also, it is important that we be able to solve trig equations as this is something that will arise off and on in this course. It is also important that we can do the kinds of number lines that we used in the last example to determine where a function is positive and where a function is negative. This is something that we will be doing on occasion in both this chapter and the next.

3.6 Derivatives of Exponential and Logarithm Functions

The next set of functions that we want to take a look at are exponential and logarithm functions. The most common exponential and logarithm functions in a calculus course are the natural exponential function, $\mathbf{e}^x$, and the natural logarithm function, $\ln(x)$. We will take a more general approach however and look at the general exponential and logarithm function.

Exponential Functions

We'll start off by looking at the exponential function,

$$f(x) = a^x$$

We want to differentiate this. The power rule that we looked at a couple of sections ago won't work as that required the exponent to be a fixed number and the base to be a variable. That is exactly the opposite from what we've got with this function. So, we're going to have to start with the definition of the derivative.

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \to 0} \frac{a^{x+h} - a^x}{h} \\ &= \lim_{h \to 0} \frac{a^x a^h - a^x}{h} \\ &= \lim_{h \to 0} \frac{a^x\left(a^h - 1\right)}{h} \end{aligned}$$

Now, the a^x is not affected by the limit since it doesn't have any h's in it and so is a constant as far as the limit is concerned. We can therefore factor this out of the limit. This gives,

$$f'(x) = a^x \lim_{h \to 0} \frac{a^h - 1}{h}$$

Now let's notice that the limit we've got above is exactly the definition of the derivative of $f(x) = a^x$ at $x = 0$, *i.e.* $f'(0)$. Therefore, the derivative becomes,

$$f'(x) = f'(0)\, a^x$$

So, we are kind of stuck. We need to know the derivative in order to get the derivative!

There is one value of a that we can deal with at this point. Back in the Exponential Functions section of the Review chapter we stated that $\mathbf{e} = 2.718281828459 05\ldots$ What we didn't do however is actually define where **e** comes from. There are in fact a variety of ways to define **e**. Here are three of them.

Some Definitions of e

1. $\mathbf{e} = \lim\limits_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$
2. **e** is the unique positive number for which $\lim\limits_{h \to 0} \frac{\mathbf{e}^h - 1}{h} = 1$
3. $\mathbf{e} = \sum\limits_{n=0}^{\infty} \frac{1}{n!}$

The second one is the important one for us because that limit is exactly the limit that we're working with above. So, this definition leads to the following fact,

Fact 1

For the natural exponential function, $f(x) = \mathbf{e}^x$ we have $f'(0) = \lim\limits_{h \to 0} \frac{\mathbf{e}^h - 1}{h} = 1$.

So, provided we are using the natural exponential function we get the following.

$$f(x) = \mathbf{e}^x \qquad \Rightarrow \qquad f'(x) = \mathbf{e}^x$$

At this point we're missing some knowledge that will allow us to easily get the derivative for a general function. Eventually we will be able to show that for a general exponential function we have,

$$f(x) = a^x \qquad \Rightarrow \qquad f'(x) = a^x \ln(a)$$

Logarithm Functions

Let's now briefly get the derivatives for logarithms. In this case we will need to start with the following fact about functions that are inverses of each other.

Fact 2

If $f(x)$ and $g(x)$ are inverses of each other then,

$$g'(x) = \frac{1}{f'(g(x))}$$

So, how is this fact useful to us? Well recall that the natural exponential function and the natural logarithm function are inverses of each other and we know what the derivative of the natural exponential function is!

So, if we have $f(x) = \mathbf{e}^x$ and $g(x) = \ln x$ then,

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{\mathbf{e}^{g(x)}} = \frac{1}{\mathbf{e}^{\ln x}} = \frac{1}{x}$$

The last step just uses the fact that the two functions are inverses of each other.

Putting this all together gives,

$$\frac{d}{dx}\Big(\ln(x)\Big) = \frac{1}{x} \qquad x > 0$$

Note that we need to require that $x > 0$ since this is required for the logarithm and so must also be required for its derivative. It can also be shown that,

$$\frac{d}{dx}\Big(\ln|x|\Big) = \frac{1}{x} \qquad x \neq 0$$

Using this all we need to avoid is $x = 0$.

In this case, unlike the exponential function case, we can actually find the derivative of the general logarithm function. All that we need is the derivative of the natural logarithm, which we just found, and the change of base formula. Using the change of base formula we can write a general logarithm as,

$$\log_a x = \frac{\ln x}{\ln a}$$

Differentiation is then fairly simple.

$$\begin{aligned}\frac{d}{dx}\Big(\log_a(x)\Big) &= \frac{d}{dx}\left(\frac{\ln(x)}{\ln(a)}\right)\\ &= \frac{1}{\ln(a)}\frac{d}{dx}\Big(\ln(x)\Big)\\ &= \frac{1}{x\ln(a)}\end{aligned}$$

We took advantage of the fact that a was a constant and so $\ln a$ is also a constant and can be factored out of the derivative. Putting all this together gives,

$$\frac{d}{dx}\Big(\log_a(x)\Big) = \frac{1}{x\ln(a)}$$

Here is a summary of the derivatives in this section.

Derivative of Exponential and Logarithm Functions

$$\frac{d}{dx}\Big(\mathbf{e}^x\Big) = \mathbf{e}^x \qquad \frac{d}{dx}\Big(a^x\Big) = a^x\ln(a)$$

$$\frac{d}{dx}\Big(\ln(x)\Big) = \frac{1}{x} \qquad \frac{d}{dx}\Big(\log_a(x)\Big) = \frac{1}{x\ln(a)}$$

Okay, now that we have the derivations of the formulas out of the way let's compute a couple of derivatives.

Example 1

Differentiate each of the following functions.

(a) $R(w) = 4^w - 5\log_9(w)$

(b) $f(x) = 3\mathbf{e}^x + 10x^3 \ln(x)$

(c) $y = \dfrac{5\mathbf{e}^x}{3\mathbf{e}^x + 1}$

Solution

(a) $R(w) = 4^w - 5\log_9$

This will be the only example that doesn't involve the natural exponential and natural logarithm functions.

$$R'(w) = 4^w \ln(4) - \frac{5}{w\ln(9)}$$

(b) $f(x) = 3\mathbf{e}^x + 10x^3 \ln(x)$

Not much to this one. Just remember to use the product rule on the second term.

$$\begin{aligned} f'(x) &= 3\mathbf{e}^x + 30x^2\ln(x) + 10x^3\left(\frac{1}{x}\right) \\ &= 3\mathbf{e}^x + 30x^2\ln(x) + 10x^2 \end{aligned}$$

(c) $y = \dfrac{5\mathbf{e}^x}{3\mathbf{e}^x + 1}$

We'll need to use the quotient rule on this one.

$$\begin{aligned} y' &= \frac{5\mathbf{e}^x(3\mathbf{e}^x + 1) - (5\mathbf{e}^x)(3\mathbf{e}^x)}{(3\mathbf{e}^x + 1)^2} \\ &= \frac{15\mathbf{e}^{2x} + 5\mathbf{e}^x - 15\mathbf{e}^{2x}}{(3\mathbf{e}^x + 1)^2} \\ &= \frac{5\mathbf{e}^x}{(3\mathbf{e}^x + 1)^2} \end{aligned}$$

There's really not a lot to differentiating natural logarithms and natural exponential functions at this point as long as you remember the formulas. In later sections as we get more formulas under our belt they will become more complicated.

Next, we need to do our obligatory application/interpretation problem so we don't forget about them.

Example 2

Suppose that the position of an object is given by

$$s(t) = t\mathbf{e}^{t}$$

Does the object ever stop moving?

Solution

First, we will need the derivative. We need this to determine if the object ever stops moving since at that point (provided there is one) the velocity will be zero and recall that the derivative of the position function is the velocity of the object.

The derivative is,

$$s'(t) = \mathbf{e}^{t} + t\mathbf{e}^{t} = (1+t)\,\mathbf{e}^{t}$$

So, we need to determine if the derivative is ever zero. To do this we will need to solve,

$$(1+t)\,\mathbf{e}^{t} = 0$$

Now, we know that exponential functions are never zero and so this will only be zero at $t = -1$. So, if we are going to allow negative values of t then the object will stop moving once at $t = -1$. If we aren't going to allow negative values of t then the object will never stop moving.

Before moving on to the next section we need to go back over a couple of derivatives to make sure that we don't confuse the two. The two derivatives are,

$$\frac{d}{dx}\left(x^{n}\right) = nx^{n-1} \qquad \text{Power Rule}$$

$$\frac{d}{dx}\left(a^{x}\right) = a^{x}\ln(a) \qquad \text{Derivative of an exponential function}$$

It is important to note that with the Power rule the exponent MUST be a constant and the base MUST be a variable while we need exactly the opposite for the derivative of an exponential function. For an exponential function the exponent MUST be a variable and the base MUST be a constant.

It is easy to get locked into one of these formulas and just use it for both of these. We also haven't even talked about what to do if both the exponent and the base involve variables. We'll see this situation in a later section.

3.7 Derivatives of Inverse Trig Functions

In this section we are going to look at the derivatives of the inverse trig functions. In order to derive the derivatives of inverse trig functions we'll need the formula from the last section relating the derivatives of inverse functions. If $f(x)$ and $g(x)$ are inverse functions then,

$$g'(x) = \frac{1}{f'(g(x))}$$

Recall as well that two functions are inverses if $f(g(x)) = x$ and $g(f(x)) = x$.

We'll go through inverse sine, inverse cosine and inverse tangent in detail here and leave the other three to you to derive if you'd like to.

Inverse Sine
Let's start with inverse sine. Here is the definition of the inverse sine.

$$y = \sin^{-1}(x) \qquad \Leftrightarrow \qquad \sin(y) = x \quad \text{for} \quad -\frac{\pi}{2} \le y \le \frac{\pi}{2}$$

So, evaluating an inverse trig function is the same as asking what angle (*i.e.* y) did we plug into the sine function to get x. The restrictions on y given above are there to make sure that we get a consistent answer out of the inverse sine. We know that there are in fact an infinite number of angles that will work and we want a consistent value when we work with inverse sine. Using the range of angles above gives all possible values of the sine function exactly once. If you're not sure of that sketch out a unit circle and you'll see that that range of angles (the y's) will cover all possible values of sine.

Note as well that since $-1 \le \sin(y) \le 1$ we also have $-1 \le x \le 1$.

Let's work a quick example.

Example 1

Evaluate $\sin^{-1}\left(\frac{1}{2}\right)$

Solution

So, we are really asking what angle y solves the following equation.

$$\sin(y) = \frac{1}{2}$$

and we are restricted to the values of y above.

From a unit circle we can quickly see that $y = \frac{\pi}{6}$.

We have the following relationship between the inverse sine function and the sine function.

$$\sin\left(\sin^{-1}(x)\right) = x \qquad\qquad \sin^{-1}\left(\sin(x)\right) = x$$

In other words they are inverses of each other. This means that we can use the fact above to find the derivative of inverse sine. Let's start with,

$$f(x) = \sin(x) \qquad\qquad g(x) = \sin^{-1}(x)$$

Then,

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{\cos\left(\sin^{-1}x\right)}$$

This is not a very useful formula. Let's see if we can get a better formula. Let's start by recalling the definition of the inverse sine function.

$$y = \sin^{-1}(x) \qquad\qquad \Rightarrow \qquad\qquad x = \sin(y)$$

Using the first part of this definition the denominator in the derivative becomes,

$$\cos\left(\sin^{-1}(x)\right) = \cos(y)$$

Now, recall that

$$\cos^2(y) + \sin^2(y) = 1 \qquad\qquad \Rightarrow \qquad\qquad \cos(y) = \sqrt{1 - \sin^2(y)}$$

Using this, the denominator is now,

$$\cos\left(\sin^{-1}(x)\right) = \cos(y) = \sqrt{1 - \sin^2(y)}$$

Now, use the second part of the definition of the inverse sine function. The denominator is then,

$$\cos\left(\sin^{-1}(x)\right) = \sqrt{1 - \sin^2(y)} = \sqrt{1 - x^2}$$

Putting all of this together gives the following derivative.

$$\frac{d}{dx}\left(\sin^{-1}(x)\right) = \frac{1}{\sqrt{1 - x^2}}$$

Inverse Cosine

Now let's take a look at the inverse cosine. Here is the definition for the inverse cosine.

$$y = \cos^{-1}(x) \qquad\qquad \Leftrightarrow \qquad\qquad \cos(y) = x \quad \text{for} \quad 0 \le y \le \pi$$

As with the inverse sine we've got a restriction on the angles, y, that we get out of the inverse cosine function. Again, if you'd like to verify this a quick sketch of a unit circle should convince you that this range will cover all possible values of cosine exactly once. Also, we also have $-1 \le x \le 1$ because $-1 \le \cos(y) \le 1$.

Example 2

Evaluate $\cos^{-1}\left(-\frac{\sqrt{2}}{2}\right)$.

Solution

As with the inverse sine we are really just asking the following.

$$\cos(y) = -\frac{\sqrt{2}}{2}$$

where y must meet the requirements given above. From a unit circle we can see that we must have $y = \frac{3\pi}{4}$.

The inverse cosine and cosine functions are also inverses of each other and so we have,

$$\cos\left(\cos^{-1}(x)\right) = x \qquad \cos^{-1}\left(\cos(x)\right) = x$$

To find the derivative we'll do the same kind of work that we did with the inverse sine above. If we start with

$$f(x) = \cos(x) \qquad g(x) = \cos^{-1}(x)$$

then,

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{-\sin\left(\cos^{-1}(x)\right)}$$

Simplifying the denominator here is almost identical to the work we did for the inverse sine and so isn't shown here. Upon simplifying we get the following derivative.

$$\frac{d}{dx}\left(\cos^{-1}(x)\right) = -\frac{1}{\sqrt{1-x^2}}$$

So, the derivative of the inverse cosine is nearly identical to the derivative of the inverse sine. The only difference is the negative sign.

Inverse Tangent

Here is the definition of the inverse tangent.

$$y = \tan^{-1}(x) \qquad \Leftrightarrow \qquad \tan(y) = x \quad \text{for} \quad -\frac{\pi}{2} < y < \frac{\pi}{2}$$

Again, we have a restriction on y, but notice that we can't let y be either of the two endpoints in the restriction above since tangent isn't even defined at those two points. To convince yourself that this range will cover all possible values of tangent do a quick sketch of the tangent function and we can see that in this range we do indeed cover all possible values of tangent. Also, in this case there are no restrictions on x because tangent can take on all possible values.

Example 3

Evaluate $\tan^{-1} 1$.

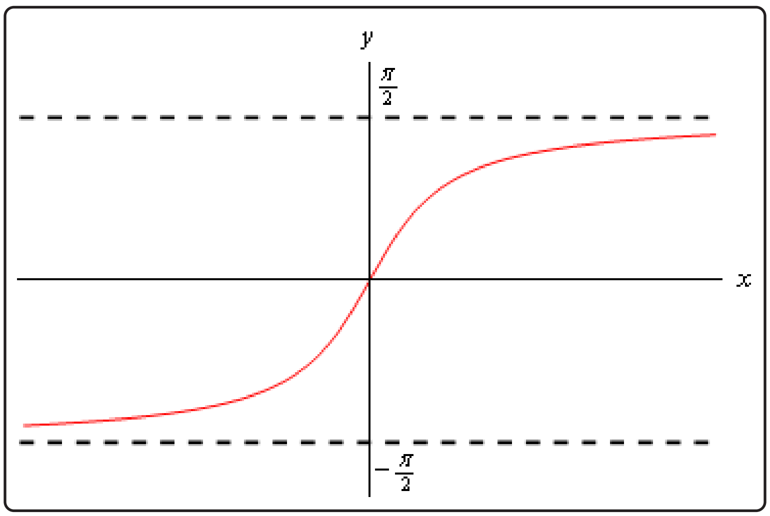
Solution

Here we are asking,

$$\tan y = 1$$

where y satisfies the restrictions given above. From a unit circle we can see that $y = \frac{\pi}{4}$.

Because there is no restriction on x we can ask for the limits of the inverse tangent function as x goes to plus or minus infinity. To do this we'll need the graph of the inverse tangent function. This is shown below.



From this graph we can see that

$$\lim_{x \to \infty} \tan^{-1}(x) = \frac{\pi}{2} \qquad \lim_{x \to -\infty} \tan^{-1}(x) = -\frac{\pi}{2}$$

The tangent and inverse tangent functions are inverse functions so,

$$\tan\left(\tan^{-1}(x)\right) = x \qquad \tan^{-1}\left(\tan(x)\right) = x$$

Therefore, to find the derivative of the inverse tangent function we can start with

$$f(x) = \tan(x) \qquad g(x) = \tan^{-1}(x)$$

We then have,

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{\sec^2\left(\tan^{-1}(x)\right)}$$

Simplifying the denominator is similar to the inverse sine, but different enough to warrant showing the details. We'll start with the definition of the inverse tangent.

$$y = \tan^{-1}(x) \qquad \Rightarrow \qquad \tan(y) = x$$

The denominator is then,

$$\sec^2\left(\tan^{-1}(x)\right) = \sec^2(y)$$

Now, if we start with the fact that

$$\cos^2(y) + \sin^2(y) = 1$$

and divide every term by $\cos^2(y)$ we will get,

$$1 + \tan^2(y) = \sec^2(y)$$

The denominator is then,

$$\sec^2\left(\tan^{-1}(x)\right) = \sec^2(y) = 1 + \tan^2(y)$$

Finally using the second portion of the definition of the inverse tangent function gives us,

$$\sec^2\left(\tan^{-1}(x)\right) = 1 + \tan^2(y) = 1 + x^2$$

The derivative of the inverse tangent is then,

$$\frac{d}{dx}\left(\tan^{-1}(x)\right) = \frac{1}{1+x^2}$$

There are three more inverse trig functions but the three shown here the most common ones. Formulas for the remaining three could be derived by a similar process as we did those above. Here are the derivatives of all six inverse trig functions.

Derivatives of Inverse Trig Functions

$$\frac{d}{dx}\left(\sin^{-1}(x)\right) = \frac{1}{\sqrt{1-x^2}} \qquad \frac{d}{dx}\left(\cos^{-1}(x)\right) = -\frac{1}{\sqrt{1-x^2}}$$

$$\frac{d}{dx}\left(\tan^{-1}(x)\right) = \frac{1}{1+x^2} \qquad \frac{d}{dx}\left(\cot^{-1}(x)\right) = -\frac{1}{1+x^2}$$

$$\frac{d}{dx}\left(\sec^{-1}(x)\right) = \frac{1}{|x|\sqrt{x^2-1}} \qquad \frac{d}{dx}\left(\csc^{-1}(x)\right) = -\frac{1}{|x|\sqrt{x^2-1}}$$

We should probably now do a couple of quick derivatives here before moving on to the next section.

Example 4

Differentiate the following functions.

(a) $f(t) = 4\cos^{-1}(t) - 10\tan^{-1}(t)$

(b) $y = \sqrt{z}\sin^{-1}(z)$

Solution

(a) $f(t) = 4\cos^{-1}(t) - 10\tan^{-1}(t)$

Not much to do with this one other than differentiate each term.

$$f'(t) = -\frac{4}{\sqrt{1-t^2}} - \frac{10}{1+t^2}$$

(b) $y = \sqrt{z}\sin^{-1}(z)$

Don't forget to convert the radical to fractional exponents before using the product rule.

$$y' = \frac{1}{2}z^{-\frac{1}{2}}\sin^{-1}(z) + \frac{\sqrt{z}}{\sqrt{1-z^2}}$$

Alternate Notation

There is some alternate notation that is used on occasion to denote the inverse trig functions. This notation is,

$$\begin{aligned} \sin^{-1}(x) &= \arcsin(x) & \cos^{-1}(x) &= \arccos(x) \\ \tan^{-1}(x) &= \arctan(x) & \cot^{-1}(x) &= \operatorname{arccot}(x) \\ \sec^{-1}(x) &= \operatorname{arcsec}(x) & \csc^{-1}(x) &= \operatorname{arccsc}(x) \end{aligned}$$

3.8 Derivatives of Hyperbolic Functions

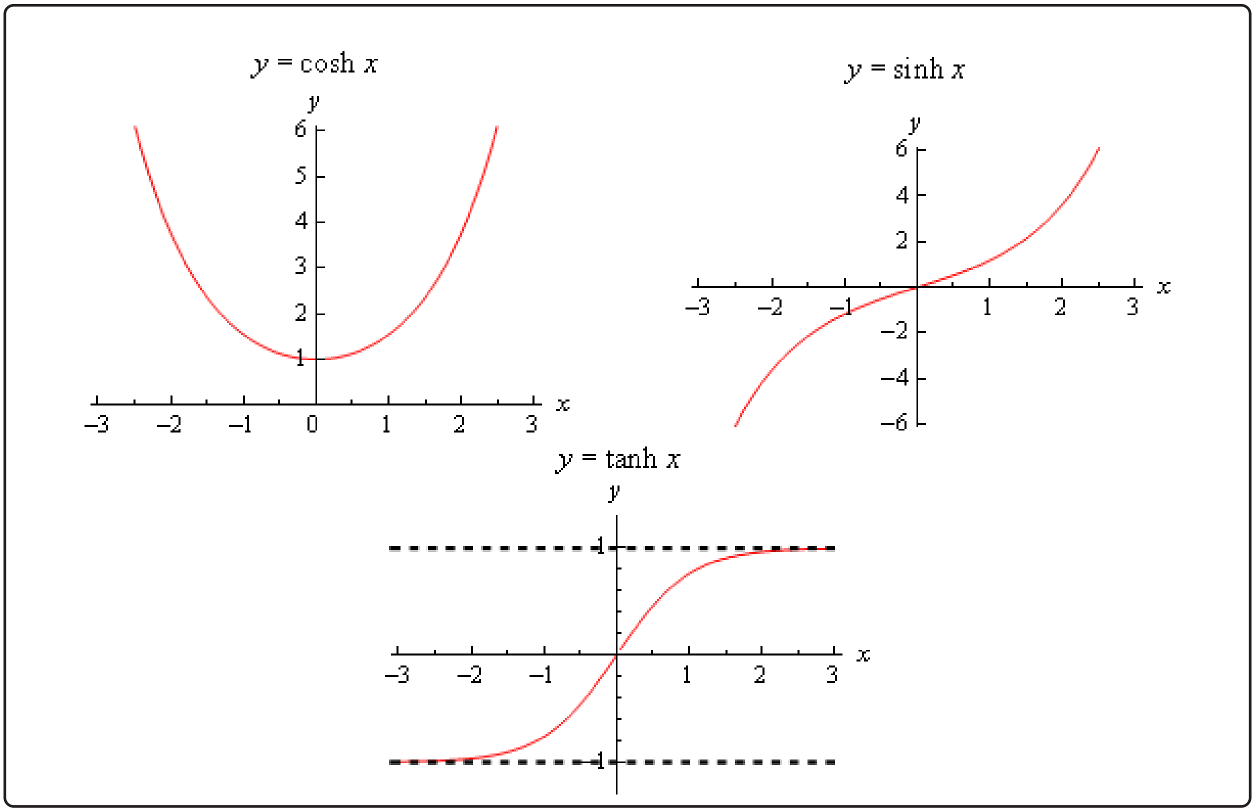
The last set of functions that we're going to be looking in this chapter at are the hyperbolic functions. In many physical situations combinations of $\mathbf{e}^x$ and $\mathbf{e}^{-x}$ arise fairly often. Because of this these combinations are given names. There are six hyperbolic functions and they are defined as follows.

$$\sinh(x) = \frac{\mathbf{e}^x - \mathbf{e}^{-x}}{2} \qquad \cosh(x) = \frac{\mathbf{e}^x + \mathbf{e}^{-x}}{2}$$

$$\tanh(x) = \frac{\sinh x}{\cosh(x)} \qquad \coth(x) = \frac{\cosh(x)}{\sinh(x)} = \frac{1}{\tanh(x)}$$

$$\operatorname{sech}(x) = \frac{1}{\cosh(x)} \qquad \operatorname{csch}(x) = \frac{1}{\sinh(x)}$$

Here are the graphs of the three main hyperbolic functions.



We also have the following facts about the hyperbolic functions.

$$\sinh(-x) = -\sinh(x) \qquad \cosh(-x) = \cosh(x)$$

$$\cosh^2(x) - \sinh^2(x) = 1 \qquad 1 - \tanh^2(x) = \operatorname{sech}^2(x)$$

You'll note that these are similar, but not quite the same, to some of the more common trig identities so be careful to not confuse the identities here with those of the standard trig functions.

Because the hyperbolic functions are defined in terms of exponential functions finding their derivatives is fairly simple provided you've already read through the next section. We haven't however so we'll need the following formula that can be easily proved after we've covered the next section.

$$\frac{d}{dx}\left(\mathbf{e}^{-x}\right) = -\mathbf{e}^{-x}$$

With this formula we'll do the derivative for hyperbolic sine and leave the rest to you as an exercise.

$$\frac{d}{dx}\left(\sinh(x)\right) = \frac{d}{dx}\left(\frac{\mathbf{e}^{x} - \mathbf{e}^{-x}}{2}\right) = \frac{\mathbf{e}^{x} - \left(-\mathbf{e}^{-x}\right)}{2} = \frac{\mathbf{e}^{x} + \mathbf{e}^{-x}}{2} = \cosh(x)$$

For the rest we can either use the definition of the hyperbolic function and/or the quotient rule. Here are all six derivatives.

Derivative of Hyperbolic Functions

$$\frac{d}{dx}\left(\sinh(x)\right) = \cosh(x) \qquad \frac{d}{dx}\left(\cosh(x)\right) = \sinh(x)$$

$$\frac{d}{dx}\left(\tanh(x)\right) = \operatorname{sech}^2(x) \qquad \frac{d}{dx}\left(\coth(x)\right) = -\operatorname{csch}^2(x)$$

$$\frac{d}{dx}\left(\operatorname{sech}(x)\right) = -\operatorname{sech}(x)\tanh(x) \qquad \frac{d}{dx}\left(\operatorname{csch}(x)\right) = -\operatorname{csch}(x)\coth(x)$$

Here are a couple of quick derivatives using hyperbolic functions.

Example 1

Differentiate each of the following functions.

(a) $f(x) = 2x^5\cosh(x)$

(b) $h(t) = \dfrac{\sinh(t)}{t+1}$

Solution

(a) $f(x) = 2x^5\cosh(x)$

$$f'(x) = 10x^4\cosh(x) + 2x^5\sinh(x)$$

(b) $h(t) = \dfrac{\sinh(t)}{t+1}$

$$h'(t) = \frac{(t+1)\cosh(t) - \sinh(t)}{(t+1)^2}$$

3.9 Chain Rule

We've taken a lot of derivatives over the course of the last few sections. However, if you look back they have all been functions similar to the following kinds of functions.

$$R(z) = \sqrt{z} \qquad f(t) = t^{50} \qquad y = \tan(x) \qquad h(w) = \mathbf{e}^{w} \qquad g(x) = \ln x$$

These are all fairly simple functions in that wherever the variable appears it is by itself. What about functions like the following,

$$R(z) = \sqrt{5z-8} \qquad f(t) = \left(2t^3 + \cos(t)\right)^{50} \qquad y = \tan\left(\sqrt[3]{3x^2} + \tan(5x)\right)$$
$$h(w) = \mathbf{e}^{w^4 - 3w^2 + 9} \qquad g(x) = \ln\left(x^{-4} + x^4\right)$$

None of our rules will work on these functions and yet some of these functions are closer to the derivatives that we're liable to run into than the functions in the first set.

Let's take the first one for example. Back in the section on the definition of the derivative we actually used the definition to compute this derivative. In that section we found that,

$$R'(z) = \frac{5}{2\sqrt{5z-8}}$$

If we were to just use the power rule on this we would get,

$$\frac{1}{2}(5z-8)^{-\frac{1}{2}} = \frac{1}{2\sqrt{5z-8}}$$

which is not the derivative that we computed using the definition. It is close, but it's not the same. So, the power rule alone simply won't work to get the derivative here.

Let's keep looking at this function and note that if we define,

$$f(z) = \sqrt{z} \qquad g(z) = 5z - 8$$

then we can write the function as a composition.

$$R(z) = (f \circ g)(z) = f(g(z)) = \sqrt{5z-8}$$

and it turns out that it's actually fairly simple to differentiate a function composition using the **Chain Rule**. There are two forms of the chain rule. Here they are.

Chain Rule

Suppose that we have two functions $f(x)$ and $g(x)$ and they are both differentiable.

1. If we define $F(x) = (f \circ g)(x)$ then the derivative of $F(x)$ is,

$$F'(x) = f'(g(x)) \;\; g'(x)$$

2. If we have $y = f(u)$ and $u = g(x)$ then the derivative of y is,

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

Each of these forms have their uses, however we will work mostly with the first form in this class. To see the proof of the Chain Rule see the Proof of Various Derivative Formulas section of the Extras appendix.

Now, let's go back and use the Chain Rule on the function that we used when we opened this section.

Example 1

Use the Chain Rule to differentiate $R(z) = \sqrt{5z-8}$.

Solution

We've already identified the two functions that we needed for the composition, but let's write them back down anyway and take their derivatives.

$$f(z) = \sqrt{z} \qquad\qquad g(z) = 5z - 8$$
$$f'(z) = \frac{1}{2\sqrt{z}} \qquad\qquad g'(z) = 5$$

So, using the chain rule we get,

$$\begin{aligned} R'(z) &= f'\big(g(z)\big) \;\; g'(z) \\ &= f'\big(5z-8\big) \;\; g'(z) \\ &= \frac{1}{2}(5z-8)^{-\frac{1}{2}} \;\; (5) \\ &= \frac{1}{2\sqrt{5z-8}} \;\; (5) \\ &= \frac{5}{2\sqrt{5z-8}} \end{aligned}$$

And this is what we got using the definition of the derivative.

In general, we don't really do all the composition stuff in using the Chain Rule. That can get a little complicated and in fact obscures the fact that there is a quick and easy way of remembering the chain rule that doesn't require us to think in terms of function composition.

Let's take the function from the previous example and rewrite it slightly.

$$R(z) = \underbrace{(5z-8)}_{\text{inside function}}{}^{\overbrace{\tfrac{1}{2}}^{}}\!\!\!\!\!\!\underbrace{\phantom{\tfrac{1}{2}}}_{\substack{\text{outside}\\\text{function}}}$$

This function has an "inside function" and an "outside function". The outside function is the square root or the exponent of $\frac{1}{2}$ depending on how you want to think of it and the inside function is the stuff that we're taking the square root of or raising to the $\frac{1}{2}$, again depending on how you want to look at it.

The derivative is then,

$$R'(z) = \overbrace{\frac{1}{2}\underbrace{(5z-8)}_{\substack{\text{inside function}\\\text{left alone}}}{}^{-\frac{1}{2}}}^{\substack{\text{derivative of}\\\text{outside function}}}\underbrace{(5)}_{\substack{\text{derivative of}\\\text{inside function}}}$$

In general, this is how we think of the chain rule. We identify the "inside function" and the "outside function". We then differentiate the outside function leaving the inside function alone and multiply all of this by the derivative of the inside function. In its general form this is,

$$F'(x) = \underbrace{f'}_{\substack{\text{derivative of}\\\text{outside function}}}\underbrace{(g(x))}_{\substack{\text{inside function}\\\text{left alone}}}\underbrace{g'(x)}_{\substack{\text{derivative of}\\\text{inside function}}}$$

We can always identify the "outside function" in the examples below by asking ourselves how we would evaluate the function. For instance in the $R(z)$ case if we were to ask ourselves what $R(2)$ is we would first evaluate the stuff under the radical and then finally take the square root of this result. The square root is the last operation that we perform in the evaluation and this is also the outside function. The outside function will always be the last operation you would perform if you were going to evaluate the function.

Let's take a look at some examples of the Chain Rule.

Example 2

Differentiate each of the following.

(a) $f(x) = \sin\left(3x^2 + x\right)$

(b) $f(t) = \left(2t^3 + \cos(t)\right)^{50}$

(c) $h(w) = \mathbf{e}^{w^4 - 3w^2 + 9}$

(d) $g(x) = \ln\left(x^{-4} + x^4\right)$

(e) $y = \sec(1 - 5x)$

(f) $P(t) = \cos^4(t) + \cos\left(t^4\right)$

Solution

(a) $f(x) = \sin\left(3x^2 + x\right)$

It looks like the outside function is the sine and the inside function is $3x^2 + x$. The derivative is then.

$$f'(x) = \underbrace{\cos}_{\substack{\text{derivative of} \\ \text{outside function}}} \quad \underbrace{\left(3x^2 + x\right)}_{\substack{\text{leave inside} \\ \text{function alone}}} \quad \underbrace{(6x + 1)}_{\substack{\text{times derivative} \\ \text{of inside function}}}$$

Or with a little rewriting,

$$f'(x) = (6x + 1)\cos\left(3x^2 + x\right)$$

(b) $f(t) = \left(2t^3 + \cos(t)\right)^{50}$

In this case the outside function is the exponent of 50 and the inside function is all the stuff on the inside of the parenthesis. The derivative is then.

$$\begin{aligned} f'(t) &= 50\left(2t^3 + \cos(t)\right)^{49}\left(6t^2 - \sin(t)\right) \\ &= 50\left(6t^2 - \sin(t)\right)\left(2t^3 + \cos(t)\right)^{49} \end{aligned}$$

(c) $h(w) = \mathbf{e}^{w^4 - 3w^2 + 9}$

Identifying the outside function in the previous two was fairly simple since it really was the “outside” function in some sense. In this case we need to be a little careful. Recall that the outside function is the last operation that we would perform in an evaluation. In this case if we were to evaluate this function the last operation would be the exponential. Therefore, the outside function is the exponential function and the inside function is its exponent.

Here's the derivative.

$$\begin{aligned} h'(w) &= \mathbf{e}^{w^4-3w^2+9}\left(4w^3-6w\right) \\ &= \left(4w^3-6w\right)\mathbf{e}^{w^4-3w^2+9} \end{aligned}$$

Remember, we leave the inside function alone when we differentiate the outside function. So, the derivative of the exponential function (with the inside left alone) is just the original function.

(d) $g(x) = \ln\left(x^{-4}+x^4\right)$

Here the outside function is the natural logarithm and the inside function is stuff on the inside of the logarithm.

$$g'(x) = \frac{1}{x^{-4}+x^4}\left(-4x^{-5}+4x^3\right) = \frac{-4x^{-5}+4x^3}{x^{-4}+x^4}$$

Again remember to leave the inside function alone when differentiating the outside function. So, upon differentiating the logarithm we end up not with 1/x but instead with 1/(inside function).

(e) $y = \sec(1-5x)$

In this case the outside function is the secant and the inside is the $1-5x$.

$$\begin{aligned} y' &= \sec(1-5x)\tan(1-5x)(-5) \\ &= -5\sec(1-5x)\tan(1-5x) \end{aligned}$$

In this case the derivative of the outside function is $\sec(x)\tan(x)$. However, since we leave the inside function alone we don't get x's in both. Instead we get $1-5x$ in both.

(f) $P(t) = \cos^4(t) + \cos\left(t^4\right)$

There are two points to this problem. First, there are two terms and each will require a different application of the chain rule. That will often be the case so don't expect just a single chain rule when doing these problems. Second, we need to be very careful in choosing the outside and inside function for each term.

Recall that the first term can actually be written as,

$$\cos^4(t) = \left(\cos(t)\right)^4$$

So, in the first term the outside function is the exponent of 4 and the inside function is the cosine. In the second term it's exactly the opposite. In the second term the outside function is the cosine and the inside function is t^4. Here's the derivative for this function.

$$\begin{aligned} P'(t) &= 4\cos^3(t)(-\sin(t)) - \sin(t^4)(4t^3) \\ &= -4\sin(t)\cos^3(t) - 4t^3\sin(t^4) \end{aligned}$$

There are a couple of general formulas that we can get for some special cases of the chain rule. Let's take a quick look at those.

Example 3

Differentiate each of the following.

(a) $f(x) = [g(x)]^n$

(b) $f(x) = \mathbf{e}^{g(x)}$

(c) $f(x) = \ln(g(x))$

Solution

(a) $f(x) = [g(x)]^n$

The outside function is the exponent and the inside is $g(x)$.

$$f'(x) = n[g(x)]^{n-1} g'(x)$$

(b) $f(x) = \mathbf{e}^{g(x)}$

The outside function is the exponential function and the inside is $g(x)$.

$$f'(x) = g'(x)\,\mathbf{e}^{g(x)}$$

(c) $f(x) = \ln(g(x))$

The outside function is the logarithm and the inside is $g(x)$.

$$f'(x) = \frac{1}{g(x)} g'(x) = \frac{g'(x)}{g(x)}$$

The formulas in this example are really just special cases of the Chain Rule but may be useful to remember in order to quickly do some of these derivatives.

Now, let's also not forget the other rules that we've got for doing derivatives. For the most part we'll not be explicitly identifying the inside and outside functions for the remainder of the problems in this section. We will be assuming that you can see our choices based on the previous examples and the work that we have shown.

Example 4

Differentiate each of the following.

(a) $T(x) = \tan^{-1}(2x)\ \sqrt[3]{1-3x^2}$

(b) $f(z) = \sin\left(z\mathbf{e}^z\right)$

(c) $y = \dfrac{\left(x^3+4\right)^5}{\left(1-2x^2\right)^3}$

(d) $h(t) = \left(\dfrac{2t+3}{6-t^2}\right)^3$

Solution

(a) $T(x) = \tan^{-1}(2x)\ \sqrt[3]{1-3x^2}$

Let's first notice that this problem is first and foremost a product rule problem. This is a product of two functions, the inverse tangent and the root and so the first thing we'll need to do in taking the derivative is use the product rule. However, in using the product rule and each derivative will require a chain rule application as well.

$$\begin{aligned} T'(x) &= \frac{1}{1+(2x)^2}(2)\ \left(1-3x^2\right)^{\frac{1}{3}} + \tan^{-1}(2x)\left(\frac{1}{3}\right)\left(1-3x^2\right)^{-\frac{2}{3}}(-6x) \\ &= \frac{2\left(1-3x^2\right)^{\frac{1}{3}}}{1+(2x)^2} - 2x\left(1-3x^2\right)^{-\frac{2}{3}}\tan^{-1}(2x) \end{aligned}$$

In this part be careful with the inverse tangent. We know that,

$$\frac{d}{dx}\left(\tan^{-1}x\right) = \frac{1}{1+x^2}$$

When doing the chain rule with this we remember that we've got to leave the inside function alone. That means that where we have the x^2 in the derivative of $\tan^{-1}(x)$ we will need to have (inside function)^2.

(b) $f(z) = \sin(z\mathbf{e}^z)$

Now contrast this with the previous problem. In the previous problem we had a product that required us to use the chain rule in applying the product rule. In this problem we will first need to apply the chain rule and when we go to differentiate the inside function we'll need to use the product rule.

Here is the chain rule portion of the problem.

$$f'(z) = \cos(z\mathbf{e}^z)\frac{d}{dz}\left[z\mathbf{e}^z\right]$$

In this case we did not actually do the derivative of the inside yet. We just left it in the derivative notation to make it clear that in order to do the derivative of the inside function we now have a product rule.

Here is the rest of the work for this problem.

$$f'(z) = \cos(z\mathbf{e}^z)(\mathbf{e}^z + z\mathbf{e}^z)$$

(c) $y = \dfrac{(x^3+4)^5}{(1-2x^2)^3}$

For this problem we clearly have a rational expression and so the first thing that we'll need to do is apply the quotient rule. In the process of using the quotient rule we'll need to use the chain rule when differentiating the numerator and denominator.

$$y' = \frac{5(x^3+4)^4(3x^2)(1-2x^2)^3 - (x^3+4)^5(3)(1-2x^2)^2(-4x)}{\left((1-2x^2)^3\right)^2}$$

These tend to be a little messy. Notice that when we go to simplify that we'll be able to a fair amount of factoring in the numerator and this will often greatly simplify the derivative.

$$\begin{aligned} y' &= \frac{(x^3+4)^4(1-2x^2)^2\left(5(3x^2)(1-2x^2) - (x^3+4)(3)(-4x)\right)}{(1-2x^2)^6} \\ &= \frac{3x(x^3+4)^4(5x-6x^3+16)}{(1-2x^2)^4} \end{aligned}$$

After factoring we were able to cancel some of the terms in the numerator against the denominator. So even though the initial chain rule was fairly messy the final answer is significantly simpler because of the factoring.

(d) $h(t) = \left(\dfrac{2t+3}{6-t^2}\right)^3$

Unlike the previous problem the first step for derivative is to use the chain rule and then once we go to differentiate the inside function we'll need to do the quotient rule.

Here is the work for this problem.

$$\begin{aligned} h'(t) &= 3\left(\frac{2t+3}{6-t^2}\right)^2 \frac{d}{dt}\left[\frac{2t+3}{6-t^2}\right] \\ &= 3\left(\frac{2t+3}{6-t^2}\right)^2 \left[\frac{2\left(6-t^2\right)-(2t+3)(-2t)}{\left(6-t^2\right)^2}\right] \\ &= 3\left(\frac{2t+3}{6-t^2}\right)^2 \left[\frac{2t^2+6t+12}{\left(6-t^2\right)^2}\right] \end{aligned}$$

As with the second part above we did not initially differentiate the inside function in the first step to make it clear that it would be quotient rule from that point on.

There were several points in the last example. First is to not forget that we've still got other derivatives rules that are still needed on occasion. Just because we now have the chain rule does not mean that the product and quotient rule will no longer be needed.

In addition, as the last example illustrated, the order in which they are done will vary as well. Some problems will be product or quotient rule problems that involve the chain rule. Other problems however, will first require the use the chain rule and in the process of doing that we'll need to use the product and/or quotient rule.

Most of the examples in this section won't involve the product or quotient rule to make the problems a little shorter. However, in practice they will often be in the same problem so you need to be prepared for these kinds of problems.

Now, let's take a look at some more complicated examples.

Example 5

Differentiate each of the following.

(a) $h(z) = \dfrac{2}{\left(4z + \mathbf{e}^{-9z}\right)^{10}}$

(b) $f(y) = \sqrt{2y + \left(3y + 4y^2\right)^3}$

(c) $y = \tan\left(\sqrt[3]{3x^2} + \ln\left(5x^4\right)\right)$

(d) $g(t) = \sin^3\left(\mathbf{e}^{1-t} + 3\sin(6t)\right)$

Solution

We're going to be a little more careful in these problems than we were in the previous ones. The reason will be quickly apparent.

(a) $h(z) = \dfrac{2}{\left(4z + \mathbf{e}^{-9z}\right)^{10}}$

In this case let's first rewrite the function in a form that will be a little easier to deal with.

$$h(z) = 2\left(4z + \mathbf{e}^{-9z}\right)^{-10}$$

Now, let's start the derivative.

$$h'(z) = -20\left(4z + \mathbf{e}^{-9z}\right)^{-11} \frac{d}{dz}\left[4z + \mathbf{e}^{-9z}\right]$$

Notice that we didn't actually do the derivative of the inside function yet. This is to allow us to notice that when we do differentiate the second term we will require the chain rule again. Notice as well that we will only need the chain rule on the exponential and not the first term. In many functions we will be using the chain rule more than once so don't get excited about this when it happens.

Let's go ahead and finish this example out.

$$h'(z) = -20\left(4z + \mathbf{e}^{-9z}\right)^{-11}\left(4 - 9\mathbf{e}^{-9z}\right)$$

Be careful with the second application of the chain rule. Only the exponential gets multiplied by the "-9" since that's the derivative of the inside function for that term only. One of the more common mistakes in these kinds of problems is to multiply the whole thing by the "-9" and not just the second term.

(b) $f(y) = \sqrt{2y + (3y + 4y^2)^3}$

We'll not put as many words into this example, but we're still going to be careful with this derivative so make sure you can follow each of the steps here.

$$\begin{aligned} f'(y) &= \frac{1}{2}\left(2y + \left(3y + 4y^2\right)^3\right)^{-\frac{1}{2}} \frac{d}{dy}\left[2y + \left(3y + 4y^2\right)^3\right] \\ &= \frac{1}{2}\left(2y + \left(3y + 4y^2\right)^3\right)^{-\frac{1}{2}} \left(2 + 3\left(3y + 4y^2\right)^2 (3 + 8y)\right) \\ &= \frac{1}{2}\left(2y + \left(3y + 4y^2\right)^3\right)^{-\frac{1}{2}} \left(2 + (9 + 24y)\left(3y + 4y^2\right)^2\right) \end{aligned}$$

As with the first example the second term of the inside function required the chain rule to differentiate it. Also note that again we need to be careful when multiplying by the derivative of the inside function when doing the chain rule on the second term.

(c) $y = \tan\left(\sqrt[3]{3x^2} + \ln\left(5x^4\right)\right)$

Let's jump right into this one.

$$\begin{aligned} y' &= \sec^2\left(\sqrt[3]{3x^2} + \ln\left(5x^4\right)\right) \frac{d}{dx}\left[\left(3x^2\right)^{\frac{1}{3}} + \ln\left(5x^4\right)\right] \\ &= \sec^2\left(\sqrt[3]{3x^2} + \ln\left(5x^4\right)\right) \left(\frac{1}{3}\left(3x^2\right)^{-\frac{2}{3}}(6x) + \frac{20x^3}{5x^4}\right) \\ &= \left(2x\left(3x^2\right)^{-\frac{2}{3}} + \frac{4}{x}\right) \sec^2\left(\sqrt[3]{3x^2} + \ln\left(5x^4\right)\right) \end{aligned}$$

In this example both of the terms in the inside function required a separate application of the chain rule.

(d) $g(t) = \sin^3\left(\mathbf{e}^{1-t} + 3\sin(6t)\right)$

We'll need to be a little careful with this one.

$$\begin{aligned} g'(t) &= 3\sin^2\left(\mathbf{e}^{1-t} + 3\sin(6t)\right) \frac{d}{dt}\left[\sin\left(\mathbf{e}^{1-t} + 3\sin(6t)\right)\right] \\ &= 3\sin^2\left(\mathbf{e}^{1-t} + 3\sin(6t)\right)\cos\left(\mathbf{e}^{1-t} + 3\sin(6t)\right) \frac{d}{dt}\left[\mathbf{e}^{1-t} + 3\sin(6t)\right] \\ &= 3\sin^2\left(\mathbf{e}^{1-t} + 3\sin(6t)\right)\cos\left(\mathbf{e}^{1-t} + 3\sin(6t)\right)\left[\mathbf{e}^{1-t}(-1) + 3\cos(6t)(6)\right) \\ &= 3\left(-\mathbf{e}^{1-t} + 18\cos(6t)\right)\sin^2\left(\mathbf{e}^{1-t} + 3\sin(6t)\right)\cos\left(\mathbf{e}^{1-t} + 3\sin(6t)\right) \end{aligned}$$

This problem required a total of 4 chain rules to complete.

Sometimes these can get quite unpleasant and require many applications of the chain rule. Initially, in these cases it's usually best to be careful as we did in this previous set of examples and write out a couple of extra steps rather than trying to do it all in one step in your head. Once you get better at the chain rule you'll find that you can do these fairly quickly in your head.

Finally, before we move onto the next section there is one more issue that we need to address. In the Derivatives of Exponential and Logarithm Functions section we claimed that,

$$f(x) = a^x \qquad \Rightarrow \qquad f'(x) = a^x \ln(a)$$

but at the time we didn't have the knowledge to do this. We now do. What we needed was the chain rule.

First, notice that using a property of logarithms we can write a as,

$$a = \mathbf{e}^{\ln(a)}$$

This may seem kind of silly, but it is needed to compute the derivative. Now, using this we can write the function as,

$$\begin{aligned} f(x) &= a^x \\ &= (a)^x \\ &= \left(\mathbf{e}^{\ln(a)}\right)^x \\ &= \mathbf{e}^{\left(\ln(a)\right)x} \\ &= \mathbf{e}^{x \ \ln(a)} \end{aligned}$$

Okay, now that we've gotten that taken care of all we need to remember is that a is a constant and so $\ln a$ is also a constant. Now, differentiating the final version of this function is a (hopefully) fairly simple Chain Rule problem.

$$f'(x) = \mathbf{e}^{x \ \ln(a)}\left(\ln(a)\right)$$

Now, all we need to do is rewrite the first term back as a^x to get,

$$f'(x) = a^x \ln(a)$$

So, not too bad if you can see the trick to rewriting the a and with using the Chain Rule.

3.10 Implicit Differentiation

To this point we've done quite a few derivatives, but they have all been derivatives of functions of the form $y = f(x)$. Unfortunately, not all the functions that we're going to look at will fall into this form.

Let's take a look at an example of a function like this.

Example 1

Find y' for $xy = 1$.

Solution

There are actually two solution methods for this problem.

Solution 1 :
This is the simple way of doing the problem. Just solve for y to get the function in the form that we're used to dealing with and then differentiate.

$$y = \frac{1}{x} \qquad \Rightarrow \qquad y' = -\frac{1}{x^2}$$

So, that's easy enough to do. However, there are some functions for which this can't be done. That's where the second solution technique comes into play.

Solution 2 :
In this case we're going to leave the function in the form that we were given and work with it in that form. However, let's recall from the first part of this solution that if we could solve for y then we will get y as a function of x. In other words, if we could solve for y (as we could in this case but won't always be able to do) we get $y = y(x)$. Let's rewrite the equation to note this.

$$xy = x\,y(x) = 1$$

Be careful here and note that when we write $y(x)$ we don't mean y times x. What we are noting here is that y is some (probably unknown) function of x. This is important to recall when doing this solution technique.

The next step in this solution is to differentiate both sides with respect to x as follows,

$$\frac{d}{dx}\Big(x\,y(x)\Big) = \frac{d}{dx}(1)$$

The right side is easy. It's just the derivative of a constant. The left side is also easy, but we've got to recognize that we've actually got a product here, the x and the $y(x)$. So, to do the derivative of the left side we'll need to do the product rule. Doing this gives,

$$(1)\,y(x) + x\frac{d}{dx}\Big(y(x)\Big) = 0$$

Now, recall that we have the following notational way of writing the derivative.

$$\frac{d}{dx}\Big(y\left(x\right)\Big) = \frac{dy}{dx} = y'$$

Using this we get the following,

$$y + xy' = 0$$

Note that we dropped the (x) on the y as it was only there to remind us that the y was a function of x and now that we've taken the derivative it's no longer really needed. We just wanted it in the equation to recognize the product rule when we took the derivative.

So, let's now recall just what were we after. We were after the derivative, y', and notice that there is now a y' in the equation. So, to get the derivative all that we need to do is solve the equation for y'.

$$y' = -\frac{y}{x}$$

There it is. Using the second solution technique this is our answer. This is not what we got from the first solution however. Or at least it doesn't look like the same derivative that we got from the first solution. Recall however, that we really do know what y is in terms of x and if we plug that in we will get,

$$y' = -\frac{1/x}{x} = -\frac{1}{x^2}$$

which is what we got from the first solution. Regardless of the solution technique used we should get the same derivative.

The process that we used in the second solution to the previous example is called **implicit differentiation** and that is the subject of this section. In the previous example we were able to just solve for y and avoid implicit differentiation. However, in the remainder of the examples in this section we either won't be able to solve for y or, as we'll see in one of the examples below, the answer will not be in a form that we can deal with.

In the second solution above we replaced the y with $y\left(x\right)$ and then did the derivative. Recall that we did this to remind us that y is in fact a function of x. We'll be doing this quite a bit in these problems, although we rarely actually write $y\left(x\right)$. So, before we actually work anymore implicit differentiation problems let's do a quick set of "simple" derivatives that will hopefully help us with doing derivatives of functions that also contain a $y\left(x\right)$.

Example 2

Differentiate each of the following.

(a) $\left(5x^3 - 7x + 1\right)^5$, $\left[f(x)\right]^5$, $\left[y(x)\right]^5$

(b) $\sin(3 - 6x)$, $\sin\left(y(x)\right)$

(c) $\mathbf{e}^{x^2 - 9x}$, $\mathbf{e}^{y(x)}$

Solution

These are written a little differently from what we're used to seeing here. This is because we want to match up these problems with what we'll be doing in this section. Also, each of these parts has several functions to differentiate starting with a specific function followed by a general function. This again, is to help us with some specific parts of the implicit differentiation process that we'll be doing.

(a) $\left(5x^3 - 7x + 1\right)^5$, $\left[f(x)\right]^5$, $\left[y(x)\right]^5$

With the first function here we're being asked to do the following,

$$\frac{d}{dx}\left[\left(5x^3 - 7x + 1\right)^5\right] = 5\left(5x^3 - 7x + 1\right)^4\left(15x^2 - 7\right)$$

and this is just the chain rule. We differentiated the outside function (the exponent of 5) and then multiplied that by the derivative of the inside function (the stuff inside the parenthesis).

For the second function we're going to do basically the same thing. We're going to need to use the chain rule. The outside function is still the exponent of 5 while the inside function this time is simply $f(x)$. We don't have a specific function here, but that doesn't mean that we can't at least write down the chain rule for this function. Here is the derivative for this function,

$$\frac{d}{dx}\left[f(x)\right]^5 = 5\left[f(x)\right]^4 f'(x)$$

We don't actually know what $f(x)$ is so when we do the derivative of the inside function all we can do is write down notation for the derivative, *i.e.* $f'(x)$.

With the final function here we simply replaced the f in the second function with a y since most of our work in this section will involve y's instead of f's. Outside of that this function is identical to the second. So, the derivative is,

$$\frac{d}{dx}\left[y(x)\right]^5 = 5\left[y(x)\right]^4 y'(x)$$

(b) sin $(3-6x)$, sin $(y(x))$

The first function to differentiate here is just a quick chain rule problem again so here is it's derivative,

$$\frac{d}{dx}\Big[\sin(3-6x)\Big] = -6\cos(3-6x)$$

For the second function we didn't bother this time with using $f(x)$ and just jumped straight to $y(x)$ for the general version. This is still just a general version of what we did for the first function. The outside function is still the sine and the inside is given by $y(x)$ and while we don't have a formula for $y(x)$ and so we can't actually take its derivative we do have a notation for its derivative. Here is the derivative for this function,

$$\frac{d}{dx}\Big[\sin(y(x))\Big] = y'(x)\cos(y(x))$$

(c) $\mathbf{e}^{x^2-9x}$, $\mathbf{e}^{y(x)}$

In this part we'll just give the answers for each and leave out the explanation that we had in the first two parts.

$$\frac{d}{dx}\left(\mathbf{e}^{x^2-9x}\right) = (2x-9)\,\mathbf{e}^{x^2-9x} \qquad \frac{d}{dx}\left(\mathbf{e}^{y(x)}\right) = y'(x)\,\mathbf{e}^{y(x)}$$

So, in this set of examples we were just doing some chain rule problems where the inside function was $y(x)$ instead of a specific function. This kind of derivative shows up all the time in doing implicit differentiation so we need to make sure that we can do them. Also note that we only did this for three kinds of functions but there are many more kinds of functions that we could have used here.

So, it's now time to do our first problem where implicit differentiation is required, unlike the first example where we could actually avoid implicit differentiation by solving for y.

Example 3

Find y' for the following function.

$$x^2 + y^2 = 9$$

Solution

Now, this is just a circle and we can solve for y which would give,

$$y = \pm\sqrt{9-x^2}$$

Prior to starting this problem, we stated that we had to do implicit differentiation here because we couldn't just solve for y and yet that's what we just did. So, why can't we use "normal" differentiation here? The problem is the "$\pm$". With this in the "solution" for y we see that y is in fact two different functions. Which should we use? Should we use both? We only want a single function for the derivative and at best we have two functions here.

So, in this example we really are going to need to do implicit differentiation so we can avoid this. In this example we'll do the same thing we did in the first example and remind ourselves that y is really a function of x and write y as $y(x)$. Once we've done this all we need to do is differentiate each term with respect to x.

$$\frac{d}{dx}\left(x^2 + \left[y(x)\right]^2\right) = \frac{d}{dx}(9)$$

As with the first example the right side is easy. The left side is also pretty easy since all we need to do is take the derivative of each term and note that the second term will be similar the part (a) of the second example. All we need to do for the second term is use the chain rule.

After taking the derivative we have,

$$2x + 2\left[y(x)\right]^1 y'(x) = 0$$

At this point we can drop the (x) part as it was only in the problem to help with the differentiation process. The final step is to simply solve the resulting equation for y'.

$$\begin{aligned} 2x + 2yy' &= 0 \\ y' &= -\frac{x}{y} \end{aligned}$$

Unlike the first example we can't just plug in for y since we wouldn't know which of the two functions to use. Most answers from implicit differentiation will involve both x and y so don't get excited about that when it happens.

As always, we can't forget our interpretations of derivatives.

Example 4

Find the equation of the tangent line to

$$x^2 + y^2 = 9$$

at the point $(2,\ \sqrt{5})$.

Solution

First note that unlike all the other tangent line problems we've done in previous sections we need to be given both the x and the y values of the point. Notice as well that this point does lie on the graph of the circle (you can check by plugging the points into the equation) and so it's okay to talk about the tangent line at this point.

Recall that to write down the tangent line all we need is the slope of the tangent line and this is nothing more than the derivative evaluated at the given point. We've got the derivative from the previous example so all we need to do is plug in the given point.

$$m = y'\big|_{x=2,\,y=\sqrt{5}} = -\frac{2}{\sqrt{5}}$$

The tangent line is then.

$$y = \sqrt{5} - \frac{2}{\sqrt{5}}\,(x-2)$$

Now, let's work some more examples. In the remaining examples we will no longer write $y\,(x)$ for y. This is just something that we were doing to remind ourselves that y is really a function of x to help with the derivatives. Seeing the $y\,(x)$ reminded us that we needed to do the chain rule on that portion of the problem. From this point on we'll leave the y's written as y's and in our head we'll need to remember that they really are $y\,(x)$ and that we'll need to do the chain rule.

There is an easy way to remember how to do the chain rule in these problems. The chain rule really tells us to differentiate the function as we usually would, except we need to add on a derivative of the inside function. In implicit differentiation this means that every time we are differentiating a term with y in it the inside function is the y and we will need to add a y' onto the term since that will be the derivative of the inside function.

Let's see a couple of examples.

Example 5

Find y' for each of the following.

(a) $x^3y^5 + 3x = 8y^3 + 1$

(b) $x^2 \tan(y) + y^{10} \sec(x) = 2x$

(c) $\mathbf{e}^{2x+3y} = x^2 - \ln(xy^3)$

Solution

(a) $x^3y^5 + 3x = 8y^3 + 1$

First differentiate both sides with respect to x and remember that each y is really $y(x)$ we just aren't going to write it that way anymore. This means that the first term on the left will be a product rule.

We differentiated these kinds of functions involving y's to a power with the chain rule in the Example 2 above. Also, recall the discussion prior to the start of this problem. When doing this kind of chain rule problem all that we need to do is differentiate the y's as normal and then add on a y', which is nothing more than the derivative of the "inside function".

Here is the differentiation of each side for this function.

$$3x^2y^5 + 5x^3y^4y' + 3 = 24y^2y'$$

Now all that we need to do is solve for the derivative, y'. This is just basic solving algebra that you are capable of doing. The main problem is that it's liable to be messier than what you're used to doing. All we need to do is get all the terms with y' in them on one side and all the terms without y' in them on the other. Then factor y' out of all the terms containing it and divide both sides by the "coefficient" of the y'. Here is the solving work for this one,

$$\begin{aligned} 3x^2y^5 + 3 &= 24y^2y' - 5x^3y^4y' \\ 3x^2y^5 + 3 &= \left(24y^2 - 5x^3y^4\right) y' \\ y' &= \frac{3x^2y^5 + 3}{24y^2 - 5x^3y^4} \end{aligned}$$

The algebra in these problems can be quite messy so be careful with that.

(b) $x^2 \tan(y) + y^{10} \sec(x) = 2x$

We've got two product rules to deal with this time. Here is the derivative of this function.

$$2x \tan(y) + x^2 \sec^2(y)\, y' + 10y^9 y' \sec(x) + y^{10} \sec(x) \tan(x) = 2$$

Notice the derivative tacked onto the secant! Again, this is just a chain rule problem similar to the second part of Example 2 above.

Now, solve for the derivative.

$$\begin{aligned}\left(x^2 \sec^2(y) + 10y^9 \sec(x)\right) y' &= 2 - y^{10} \sec(x) \tan(x) - 2x \tan(y)\\ y' &= \frac{2 - y^{10} \sec(x) \tan(x) - 2x \tan(y)}{x^2 \sec^2(y) + 10y^9 \sec(x)}\end{aligned}$$

(c) $\mathbf{e}^{2x+3y} = x^2 - \ln\left(xy^3\right)$

We're going to need to be careful with this problem. We've got a couple chain rules that we're going to need to deal with here that are a little different from those that we've dealt with prior to this problem.

In both the exponential and the logarithm we've got a "standard" chain rule in that there is something other than just an x or y inside the exponential and logarithm. So, this means we'll do the chain rule as usual here and then when we do the derivative of the inside function for each term we'll have to deal with differentiating y's.

Here is the derivative of this equation.

$$\mathbf{e}^{2x+3y}\left(2 + 3y'\right) = 2x - \frac{y^3 + 3xy^2 y'}{xy^3}$$

In both of the chain rules note that the y' didn't get tacked on until we actually differentiated the y's in that term.

Now we need to solve for the derivative and this is liable to be somewhat messy. In order to get the y' on one side we'll need to multiply the exponential through the parenthesis and break up the quotient.

$$\begin{aligned}2\mathbf{e}^{2x+3y} + 3y'\mathbf{e}^{2x+3y} &= 2x - \frac{y^3}{xy^3} - \frac{3xy^2 y'}{xy^3}\\ 2\mathbf{e}^{2x+3y} + 3y'\mathbf{e}^{2x+3y} &= 2x - \frac{1}{x} - \frac{3y'}{y}\\ \left(3\mathbf{e}^{2x+3y} + 3y^{-1}\right) y' &= 2x - x^{-1} - 2\mathbf{e}^{2x+3y}\\ y' &= \frac{2x - x^{-1} - 2\mathbf{e}^{2x+3y}}{3\mathbf{e}^{2x+3y} + 3y^{-1}}\end{aligned}$$

Note that to make the derivative at least look a little nicer we converted all the fractions to negative exponents.

Okay, we've seen one application of implicit differentiation in the tangent line example above. However, there is another application that we will be seeing in every problem in the next section.

In some cases we will have two (or more) functions all of which are functions of a third variable. So, we might have $x(t)$ and $y(t)$, for example and in these cases, we will be differentiating with respect to t. This is just implicit differentiation like we did in the previous examples, but there is a difference however.

In the previous examples we have functions involving x's and y's and thinking of y as $y(x)$. In these problems we differentiated with respect to x and so when faced with x's in the function we differentiated as normal and when faced with y's we differentiated as normal except we then added a y' onto that term because we were really doing a chain rule.

In the new example we want to look at we're assuming that $x = x(t)$ and that $y = y(t)$ and differentiating with respect to t. This means that every time we are faced with an x or a y we'll be doing the chain rule. This in turn means that when we differentiate an x we will need to add on an x' and whenever we differentiate a y we will add on a y'.

These new types of problems are really the same kind of problem we've been doing in this section. They are just expanded out a little to include more than one function that will require a chain rule.

Let's take a look at an example of this kind of problem.

Example 6

Assume that $x = x(t)$ and $y = y(t)$ and differentiate the following equation with respect to t.

$$x^3y^6 + \mathbf{e}^{1-x} - \cos(5y) = y^2$$

Solution

So, just differentiate as normal and add on an appropriate derivative at each step. Note as well that the first term will be a product rule since both x and y are functions of t.

$$3x^2x'y^6 + 6x^3y^5y' - x'\mathbf{e}^{1-x} + 5y'\sin(5y) = 2yy'$$

There really isn't all that much to this problem. Since there are two derivatives in the problem we won't be bothering to solve for one of them. When we do this kind of problem in the next section the problem will imply which one we need to solve for.

At this point there doesn't seem be any real reason for doing this kind of problem, but as we'll see in the next section every problem that we'll be doing there will involve this kind of implicit differentiation.

3.11 Related Rates

In this section we are going to look at an application of implicit differentiation. Most of the applications of derivatives are in the next chapter however there are a couple of reasons for placing it in this chapter as opposed to putting it into the next chapter with the other applications. The first reason is that it's an application of implicit differentiation and so putting it right after that section means that we won't have forgotten how to do implicit differentiation. The other reason is simply that after doing all these derivatives we need to be reminded that there really are actual applications to derivatives. Sometimes it is easy to forget there really is a reason that we're spending all this time on derivatives.

For these related rates problems, it's usually best to just jump right into some problems and see how they work.

Example 1

Air is being pumped into a spherical balloon at a rate of 5 cm^3/min. Determine the rate at which the radius of the balloon is increasing when the diameter of the balloon is 20 cm.

Solution

The first thing that we'll need to do here is to identify what information that we've been given and what we want to find. Before we do that let's notice that both the volume of the balloon and the radius of the balloon will vary with time and so are really functions of time, *i.e.* $V(t)$ and $r(t)$.

We know that air is being pumped into the balloon at a rate of 5 cm^3/min. This is the rate at which the volume is increasing. Recall that rates of change are nothing more than derivatives and so we know that,

$$V'(t) = 5$$

We want to determine the rate at which the radius is changing. Again, rates are derivatives and so it looks like we want to determine,

$$r'(t) = ? \qquad \text{when} \qquad r(t) = \frac{d}{2} = 10\,\text{cm}$$

Note that we needed to convert the diameter to a radius.

Now that we've identified what we have been given and what we want to find we need to relate these two quantities to each other. In this case we can relate the volume and the radius with the formula for the volume of a sphere.

$$V(t) = \frac{4}{3}\pi[r(t)]^3$$

As in the previous section when we looked at implicit differentiation, we will typically not use the (t) part of things in the formulas, but since this is the first time through one of these we will do that to remind ourselves that they are really functions of t.

Now we don't really want a relationship between the volume and the radius. What we really want is a relationship between their derivatives. We can do this by differentiating both sides with respect to t. In other words, we will need to do implicit differentiation on the above formula. Doing this gives,

$$V' = 4\pi r^2 r'$$

Note that at this point we went ahead and dropped the (t) from each of the terms. Now all that we need to do is plug in what we know and solve for what we want to find.

$$5 = 4\pi\left(10^2\right)r' \qquad \Rightarrow \qquad r' = \frac{1}{80\pi}\ \text{cm/min}$$

We can get the units of the derivative by recalling that,

$$r' = \frac{dr}{dt}$$

The units of the derivative will be the units of the numerator (cm in the previous example) divided by the units of the denominator (min in the previous example).

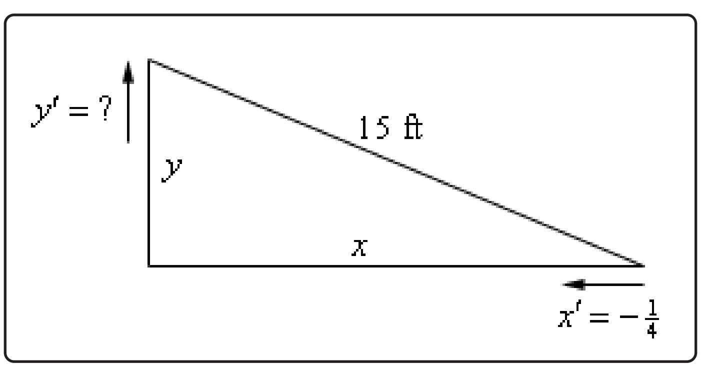
Let's work some more examples.

Example 2

A 15 foot ladder is resting against the wall. The bottom is initially 10 feet away from the wall and is being pushed towards the wall at a rate of $\frac{1}{4}$ ft/sec. How fast is the top of the ladder moving up the wall 12 seconds after we start pushing?

Solution

The first thing to do in this case is to sketch picture that shows us what is going on.



We've defined the distance of the bottom of the ladder from the wall to be x and the distance of the top of the ladder from the floor to be y. Note as well that these are changing with time and so we really should write $x(t)$ and $y(t)$. However, as is often the case with related rates/implicit differentiation problems we don't write the (t) part just try to remember this in our heads as we proceed with the problem.

Next, we need to identify what we know and what we want to find. We know that the rate at which the bottom of the ladder is moving towards the wall. This is,

$$x' = -\frac{1}{4}$$

Note as well that the rate is negative since the distance from the wall, x, is decreasing. We always need to be careful with signs with these problems.

We want to find the rate at which the top of the ladder is moving away from the floor. This is y'. Note as well that this quantity should be positive since y will be increasing.

As with the first example we first need a relationship between x and y. We can get this using Pythagorean theorem.

$$x^2 + y^2 = (15)^2 = 225$$

All that we need to do at this point is to differentiate both sides with respect to t, remembering that x and y are really functions of t and so we'll need to do implicit differentiation. Doing this gives an equation that shows the relationship between the derivatives.

$$2xx' + 2yy' = 0 \tag{3.3}$$

Next, let's see which of the various parts of this equation that we know and what we need to find. We know x' and are being asked to determine y' so it's okay that we don't know that. However, we still need to determine x and y.

Determining x and y is actually fairly simple. We know that initially $x = 10$ and the end is being pushed in towards the wall at a rate of $\frac{1}{4}$ ft/sec and that we are interested in what has happened after 12 seconds. We know that,

$$\begin{aligned} \text{distance} &= \text{rate} \times \text{time} \\ &= \left(\frac{1}{4}\right)(12) = 3 \end{aligned}$$

So, the end of the ladder has been pushed in 3 feet and so after 12 seconds we must have $x = 7$. Note that we could have computed this in one step as follows,

$$x = 10 - \frac{1}{4}(12) = 7$$

To find y (after 12 seconds) all that we need to do is reuse the Pythagorean Theorem with the values of x that we just found above.

$$y = \sqrt{225 - x^2} = \sqrt{225 - 49} = \sqrt{176}$$

Now all that we need to do is plug into Equation 3.3 and solve for y'.

$$2\left(7\right)\left(-\frac{1}{4}\right) + 2\left(\sqrt{176}\right)y' = 0 \Rightarrow y' = \frac{7/4}{\sqrt{176}} = \frac{7}{4\sqrt{176}} = 0.1319 \text{ ft/sec}$$

Notice that we got the correct sign for y'. If we'd gotten a negative value we'd have known that we had made a mistake and we could go back and look for it.

Before working another example, we need to make a comment about the set up of the previous problem. When we labeled our sketch, we acknowledged that the hypotenuse is constant and so just called it 15 ft. A common mistake that students will sometimes make here is to also label the hypotenuse as a letter, say z, in this case.

Well, it's not really a mistake to label with a letter, but it will often lead to problem down the road. Had we labeled the hypotenuse z then the Pythagorean theorem and its derivative would have been,

$$x^2 + y^2 = z^2 \qquad \rightarrow \qquad 2xx' + 2yy' = 2zz'$$

Again, there is nothing wrong with doing this but it does require that we acknowledge the values of two more quantities, z and z'. Because z is just the hypotenuse that is clearly $z = 15$. The problem that some students then sometimes run into is determining the value of z'. In this case, we have to remember that because the ladder, and hence the hypotenuse has a fixed length, its length can't be changing and so $z' = 0$.

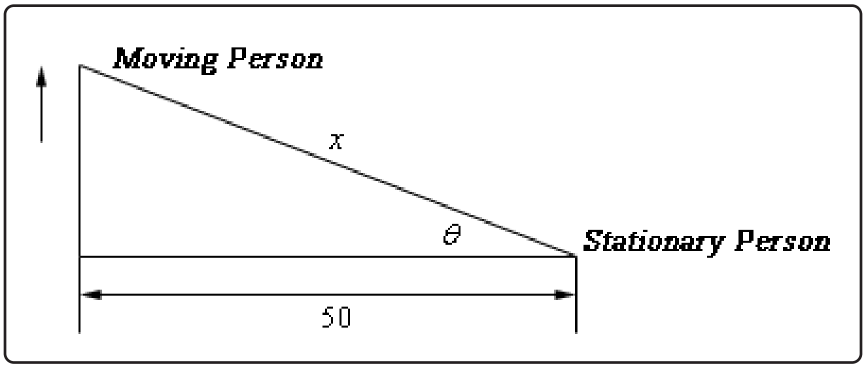
Plugging both of these values into the derivative give us same equation that we got in the example but required a little more effort to get to. It would have been easier to just label the hypotenuse 15 to start off with and not have to worry about remembering that $z' = 0$.

When labeling a fixed quantity (the length of the ladder in this example) with a letter it is sometimes easy to forget that it is a fixed quantity and so it's derivative must be zero. If you don't remember this, the problem becomes impossible to finish as you will have two unknown quantities that you have to deal with. In any problem were a quantity is fixed and will never over the course of the problem change it is always best to just acknowledge that and label it with its value rather than with a letter.

Of course, if we'd had a sliding ladder that was allowed to change length then we would have to label it with a letter. However, for that kind of problem we would also need some more information in the problem statement in order to actually do the problem. The practice problems in this section have several problems in which all three sides of a right triangle are changing. You should check them out and see if you can work them.

Example 3

Two people are 50 feet apart. One of them starts walking north at a rate so that the angle shown in the diagram below is changing at a constant rate of 0.01 rad/min. At what rate is distance between the two people changing when $\theta = 0.5$ radians?



Solution

This example is not as tricky as it might at first appear. Let's call the distance between them at any point in time x as noted above. We can then relate all the known quantities by one of two trig formulas.

$$\cos\theta = \frac{50}{x} \qquad \sec\theta = \frac{x}{50}$$

We want to find x' and we could find x if we wanted to at the point in question using cosine since we also know the angle at that point in time. However, if we use the second formula we won't need to know x as you'll see. So, let's differentiate that formula.

$$\sec\theta\tan\theta\ \theta' = \frac{x'}{50}$$

As noted, there are no x's in this formula. We want to determine x' and we know that $\theta = 0.5$ and $\theta' = 0.01$ (do you agree with it being positive?). So, just plug in and solve.

$$(50)(0.01)\sec(0.5)\tan(0.5) = x' \qquad \Rightarrow \qquad x' = 0.311254 \text{ ft/min}$$

So far we we've seen three related rates problems. While each one was worked in a very different manner the process was essentially the same in each. In each problem we identified what we were given and what we wanted to find. We next wrote down a relationship between all the various quantities and used implicit differentiation to arrive at a relationship between the various derivatives in the problem. Finally, we plugged the known quantities into the equation to find the value we were after.

So, in a general sense each problem was worked in pretty much the same manner. The only real difference between them was coming up with the relationship between the known and unknown quantities. This is often the hardest part of the problem. In many problems the best way to come up with the relationship is to sketch a diagram that shows the situation. This often seems like a silly step but can make all the difference in whether we can find the relationship or not.

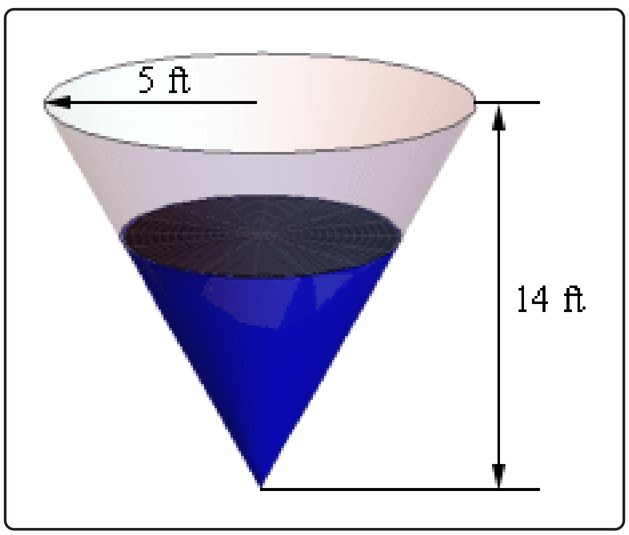
Let's work another problem that uses some different ideas and shows some of the different kinds of things that can show up in related rates problems.

Example 4

A tank of water in the shape of a cone is leaking water at a constant rate of $2\,\text{ft}^3/\text{hour}$. The base radius of the tank is 5 ft and the height of the tank is 14 ft.

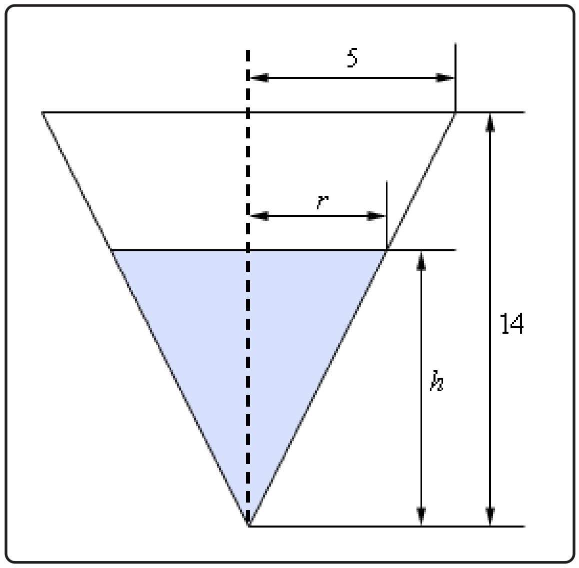
(a) At what rate is the depth of the water in the tank changing when the depth of the water is 6 ft?

(b) At what rate is the radius of the top of the water in the tank changing when the depth of the water is 6 ft?



Solution

Okay, we should probably start off with a quick sketch (probably not to scale) of what is going on here. We'll also be doing the sketch as if we were looking at the tank from directly in front of it (and so the 3D of the tank will not be visible) as this will help a little with seeing what is going on. Showing the 3D nature of the tank is liable to just get in the way. So here is the sketch of the tank with some water in it.



As we can see, the water in the tank actually forms a smaller cone/triangle (depending on which image we are looking at) with the same central angle as the tank itself. The radius of the "water" cone at any time is given by r and the height of the "water" cone at any time is given by h. The volume of water in the tank at any time t is given by,

$$V = \frac{1}{3}\pi r^2 h$$

and we've been given that $V' = -2$.

(a) At what rate is the depth of the water in the tank changing when the depth of the water is 6 ft?

For this part we need to determine h' when $h = 6$ and now we have a problem. The only formula that we've got that will relate the volume to the height also includes the radius and so if we were to differentiate this with respect to t we would get,

$$V' = \frac{2}{3}\pi r r' h + \frac{1}{3}\pi r^2 h'$$

So, in this equation we know V' and h and want to find h', but we don't know r and r'. As we'll see finding r isn't too bad, but we just don't have enough information, at this point, that will allow us to find r' and h' simultaneously.

To fix this we'll need to eliminate the r from the volume formula in some way. This is actually easier than it might at first look. If we go back to our sketch above and look at just the right half of the tank we see that we have two similar triangles and when we say similar we mean similar in the geometric sense. Recall that two triangles are called similar if their angles are identical, which is the case here. When we have two similar triangles then ratios of any two sides will be equal. For our set this means that we have,

$$\frac{r}{h} = \frac{5}{14} \qquad \Rightarrow \qquad r = \frac{5}{14}h$$

If we take this and plug it into our volume formula we have,

$$V = \frac{1}{3}\pi r^2 h = \frac{1}{3}\pi\left(\frac{5}{14}h\right)^2 h = \frac{25}{588}\pi h^3$$

This gives us a volume formula that only involved the volume and the height of the water. Note however that this volume formula is only valid for our cone, so don't be tempted to use it for other cones! If we now differentiate this we have,

$$V' = \frac{25}{196}\pi h^2 h'$$

At this point all we need to do is plug in what we know and solve for h'.

$$-2 = \frac{25}{196}\pi\left(6^2\right)h' \qquad \Rightarrow \qquad h' = \frac{-98}{225\pi} = -0.1386$$

So, it looks like the height is decreasing at a rate of 0.1386 ft/hr.

(b) At what rate is the radius of the top of the water in the tank changing when the depth of the water is 6 ft?

In this case we are asking for r' and there is an easy way to do this part and a difficult (well, more difficult than the easy way anyway....) way to do it. The "difficult" way is to redo the work in part (a) above only this time use,

$$\frac{h}{r} = \frac{14}{5} \qquad \Rightarrow \qquad h = \frac{14}{5}r$$

to get the volume in terms of V and r and then proceed as before.

That's not terribly difficult, but it is more work that we need to so. Recall from the first part that we have,

$$r = \frac{5}{14}h \qquad \Rightarrow \qquad r' = \frac{5}{14}h'$$

So, as we can see if we take the relationship that relates r and h that we used in the first part and differentiate it we get a relationship between r' and h'. At this point all we need to do here is use the result from the first part to get,

$$r' = \frac{5}{14}\left(\frac{-98}{225\pi}\right) = -\frac{7}{45\pi} = -0.04951$$

Much easier that redoing all of the first part. Note however, that we were only able to do this the "easier" way because it was asking for r' at exactly the same time that we asked for h' in the first part. If we hadn't been using the same time then we would have had no choice but to do this the "difficult" way.

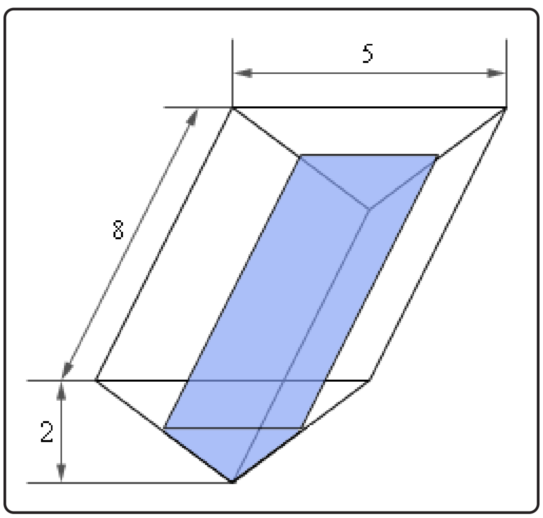
In the second part of the previous problem we saw an important idea in dealing with related rates. In order to find the asked for rate all we need is an equation that relates the rate we're looking for to a rate that we already know. Sometimes there are multiple equations that we can use and sometimes one will be easier than another.

Also, this problem showed us that we will often have an equation that contains more variables that we have information about and so, in these cases, we will need to eliminate one (or more) of the variables. In this problem we eliminated the extra variable using the idea of similar triangles. This will not always be how we do this, but many of these problems do use similar triangles so make sure you can use that idea.

Let's work some more problems.

Example 5

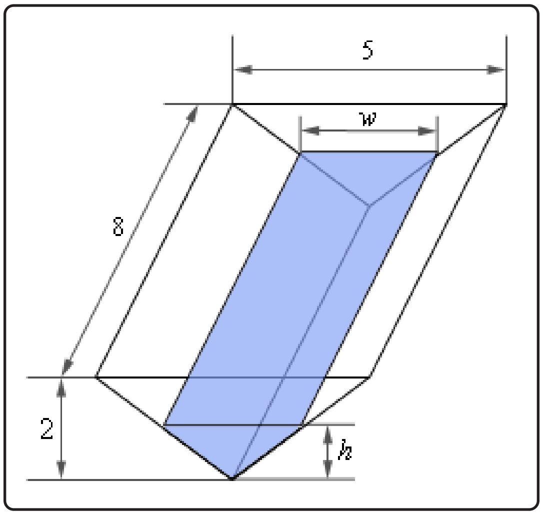
A trough of water is 8 meters in length and its ends are in the shape of isosceles triangles whose width is 5 meters and height is 2 meters. If water is being pumped in at a constant rate of 6 m^3/sec. At what rate is the height of the water changing when the water has a height of 120 cm? At what rate is the width of the water changing when the water has a height of 120cm?



Solution

Note that an isosceles triangle is just a triangle in which two of the sides are the same length. In our case sides of the tank have the same length.

Let's add in some dimensions for the water to the sketch from above.



Now, in this problem we know that $V' = 6\,\text{m}^3$/sec and we want to determine h' when $h = 1.2\,\text{m}$. Note that because V' is in terms of meters we need to convert h into meters as well. So, we need an equation that will relate these two quantities and the volume of the tank will do it.

The volume of this kind of tank is simple to compute. The volume is the area of the end times the depth. For our case the volume of the water in the tank is,

$$\begin{aligned} V &= (\text{Area of End})\,(\text{depth}) \\ &= \left(\frac{1}{2}\text{base} \times \text{height}\right)(\text{depth}) \\ &= \frac{1}{2}hw\,(8) \\ &= 4hw \end{aligned}$$

As with the previous example we've got an extra quantity here, w, that is also changing with time and so we need to eliminate it from the problem. To do this we'll again make use of the idea of similar triangles. If we look at the end of the tank we'll see that we again have two similar triangles. One for the tank itself and one formed by the water in the tank. Again, remember that with similar triangles ratios of sides must be equal. In our case we'll use,

$$\frac{w}{5} = \frac{h}{2} \qquad \Rightarrow \qquad w = \frac{5}{2}h$$

Plugging this into the volume gives a formula for the volume (and only for this tank) that only involved the height of the water.

$$V = 4hw = 4h\left(\frac{5}{2}h\right) = 10h^2$$

We can now differentiate this to get,

$$V' = 20hh'$$

Finally, all we need to do is plug in and solve for h'.

$$6 = 20\,(1.2)\,h' \qquad \Rightarrow \qquad h' = 0.25 \text{ m/sec}$$

So, the height of the water is rising at a rate of 0.25 m/sec.

In order to answer the second part of this question is not all that difficult.

We will need w' to answer this part and we have the following equation from the similar triangle that relate the width to the height and we can quickly differentiate it to get a relationship between w' and h'.

$$w = \frac{5}{2}h \qquad \Rightarrow \qquad w' = \frac{5}{2}h'$$

From the first part we know the value of h' and so all wee need to do is plug that into this equation and we'll have the answer.

$$w' = \frac{5}{2}\,(0.25) = 0.625 \text{ m/sec}$$

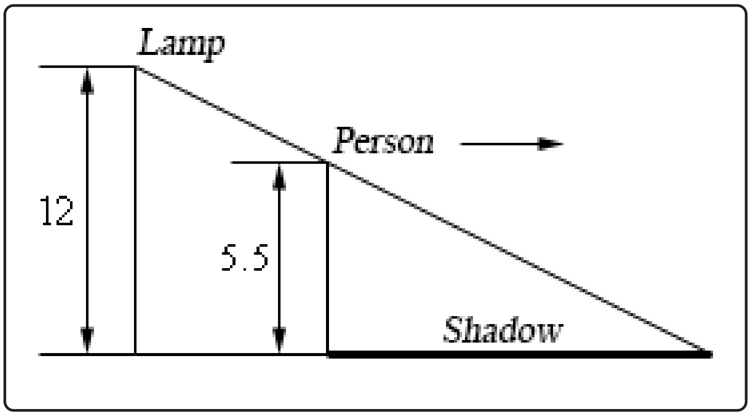
Therefore the width is increasing at a rate of 0.625 m/sec.

Example 6

A light is on the top of a 12 ft tall pole and a 5 ft 6 in tall person is walking away from the pole at a rate of 2 ft/sec.

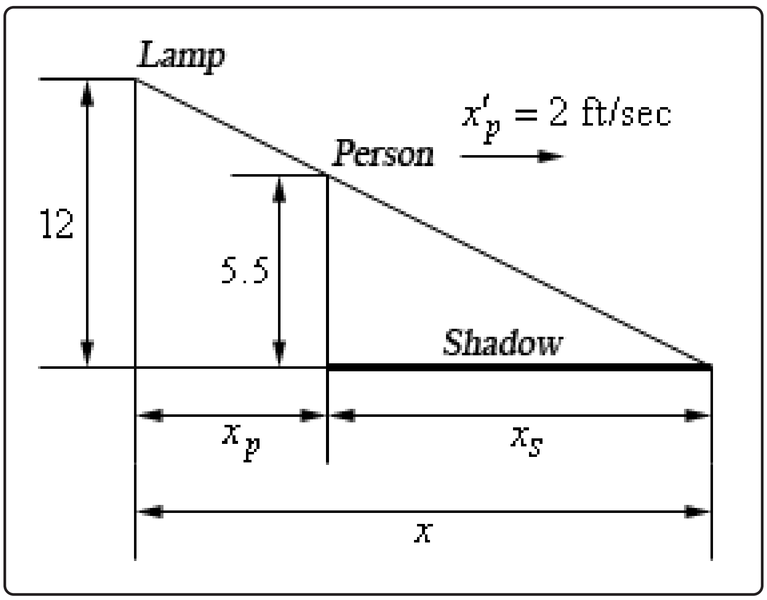
(a) At what rate is the tip of the shadow moving away from the pole when the person is 25 ft from the pole?

(b) At what rate is the tip of the shadow moving away from the person when the person is 25 ft from the pole?



Solution

Let's start off with putting all the relevant quantities into the sketch from above.



Here x is the distance of the tip of the shadow from the pole, x_p is the distance of the person from the pole and x_s is the length of the shadow. Also note that we converted the persons height over to 5.5 feet since all the other measurements are in feet.

The tip of the shadow is defined by the rays of light just getting past the person and so we can see they form a set of similar triangles. This will be useful down the road.

(a) At what rate is the tip of the shadow moving away from the pole when the person is 25 ft from the pole?

In this case we want to determine x' when $x_p = 25$ given that $x_p' = 2$.

The equation we'll need here is,

$$x = x_p + x_s$$

but we'll need to eliminate x_s from the equation in order to get an answer. To do this we can again make use of the fact that the two triangles are similar to get,

$$\frac{5.5}{12} = \frac{x_s}{x} \qquad \text{Note : } \frac{5.5}{12} = \frac{\frac{11}{2}}{12} = \frac{11}{24}$$

From this we can quickly see that,

$$x_s = \frac{11}{24}x$$

We can then plug this into the equation above and solve for x as follows.

$$x = x_p + x_s = x_p + \frac{11}{24}x \qquad \Rightarrow \qquad x = \frac{24}{13}x_p$$

Now all that we need to do is differentiate this, plug in and solve for x'.

$$x' = \frac{24}{13}x_p' \qquad \Rightarrow \qquad x' = \frac{24}{13}(2) = 3.6923\text{ft/sec}$$

The tip of the shadow is then moving away from the pole at a rate of 3.6923 ft/sec. Notice as well that we never actually had to use the fact that $x_p = 25$ for this problem. That will happen on rare occasions.

(b) At what rate is the tip of the shadow moving away from the person when the person is 25 ft from the pole?

This part is actually quite simple if we have the answer from (a) in hand, which we do of course. In this case we know that x_s represents the length of the shadow, or the distance of the tip of the shadow from the person so it looks like we want to determine x_s' when $x_p = 25$.

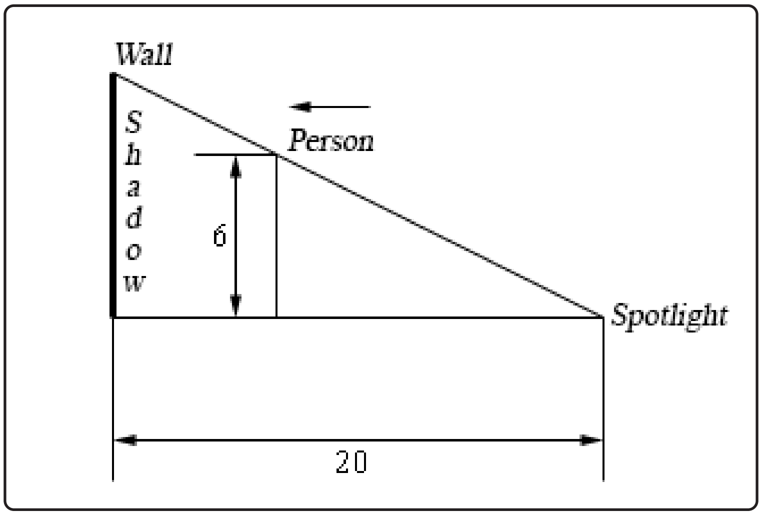
Again, we can use $x = x_p + x_s$, however unlike the first part we now know that $x_p' = 2$ and $x' = 3.6923\text{ft/sec}$ so in this case all we need to do is differentiate the equation and plug in for all the known quantities.

$$\begin{aligned} x' &= x'_p + x'_s \\ 3.6923 &= 2 + x'_s \qquad\qquad x'_s = 1.6923\text{ft/sec} \end{aligned}$$

The tip of the shadow is then moving away from the person at a rate of 1.6923 ft/sec.

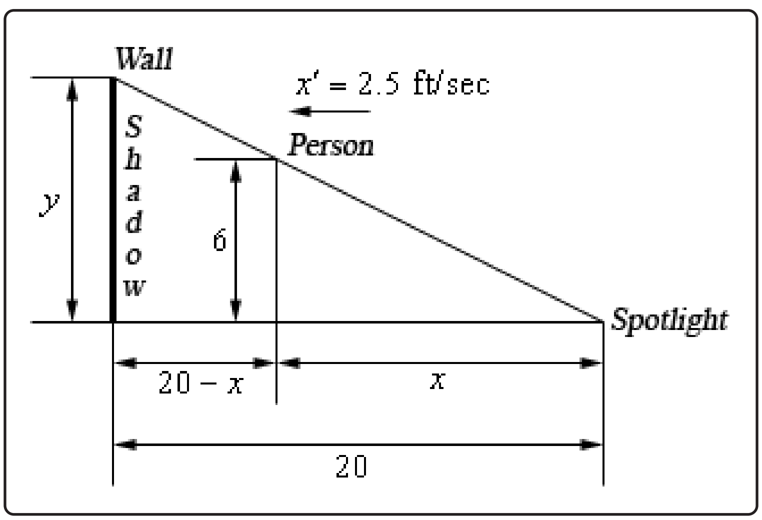
Example 7

A spot light is on the ground 20 ft away from a wall and a 6 ft tall person is walking towards the wall at a rate of 2.5 ft/sec. How fast is the height of the shadow changing when the person is 8 feet from the wall? Is the shadow increasing or decreasing in height at this time?



Solution

Below is a copy of the sketch in the problem statement with all the relevant quantities added in. The top of the shadow will be defined by the light rays going over the head of the person and so we again get yet another set of similar triangles.



In this case we want to determine y' when the person is 8 ft from wall or $x = 12$ ft. Also, if the person is moving towards the wall at 2.5 ft/sec then the person must be moving away from the spotlight at 2.5 ft/sec and so we also know that $x' = 2.5$.

In all the previous problems that used similar triangles we used the similar triangles to eliminate one of the variables from the equation we were working with. In this case however, we can get the equation that relates x and y directly from the two similar triangles. In this case the equation we're going to work with is,

$$\frac{y}{6} = \frac{20}{x} \qquad \Rightarrow \qquad y = \frac{120}{x}$$

Now all that we need to do is differentiate and plug values into solve to get y'.

$$y' = -\frac{120}{x^2}x' \qquad \Rightarrow \qquad y' = -\frac{120}{12^2}(2.5) = -2.0833\text{ft/sec}$$

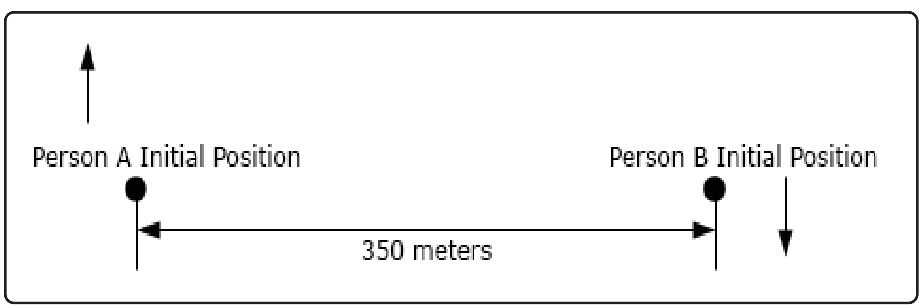
The height of the shadow is then decreasing at a rate of 2.0833 ft/sec.

Okay, we've worked quite a few problems now that involved similar triangles in one form or another so make sure you can do these kinds of problems.

It's now time to do a problem that while similar to some of the problems we've done to this point is also sufficiently different that it can cause problems until you've seen how to do it.

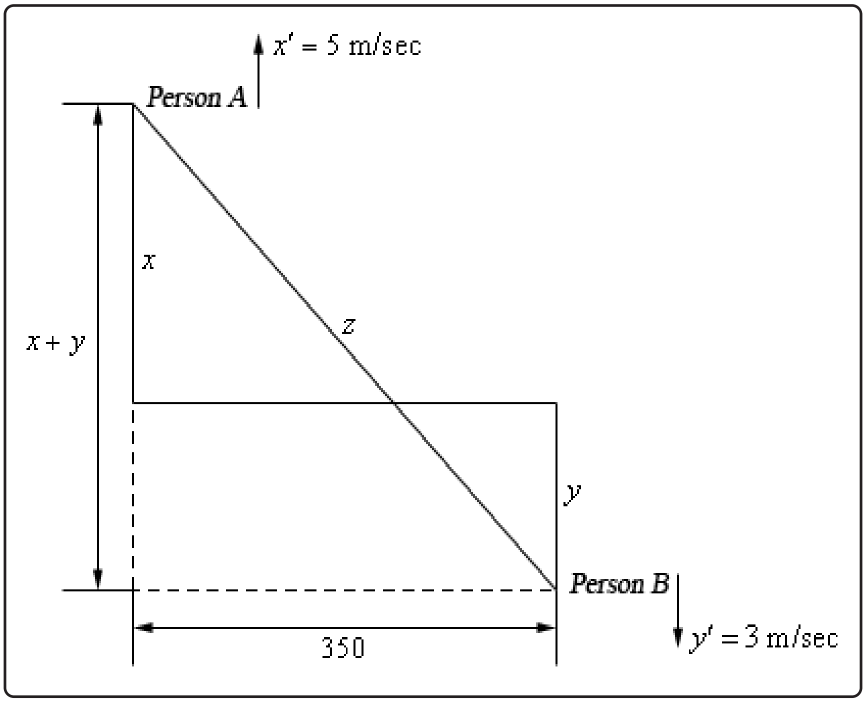
Example 8

Two people on bikes are separated by 350 meters. Person A starts riding north at a rate of 5 m/sec and 7 minutes later Person B starts riding south at 3 m/sec. At what rate is the distance separating the two people changing 25 minutes after Person A starts riding?



Solution

There is a lot to digest here with this problem. Let's start off with a sketch of the situation that shows each person's location sometime after both people start riding.



Now we are after z' and we know that $x' = 5$ and $y' = 3$. We want to know z' after Person A had been riding for 25 minutes and Person B has been riding for $25 - 7 = 18$ minutes. After converting these times to seconds (because our rates are all in m/sec) this means that at the time we're interested in each of the bike riders has rode,

$$x = 5\,(25 \times 60) = 7500\text{m} \qquad\qquad y = 3\,(18 \times 60) = 3240\text{m}$$

Next, the Pythagorean theorem tells us that,

$$z^2 = (x+y)^2 + 350^2 \tag{3.4}$$

Therefore, 25 minutes after Person A starts riding the two bike riders are

$$z = \sqrt{(x+y)^2 + 350^2} = \sqrt{(7500+3240)^2 + 350^2} = 10745.7015 \text{ m}$$

apart.

To determine the rate at which the two riders are moving apart all we need to do then is differentiate Equation 3.4 and plug in all the quantities that we know to find z'.

$$\begin{aligned} 2zz' &= 2\,(x+y)\,(x'+y') \\ 2\,(10745.7015)\,z' &= 2\,(7500+3240)\,(5+3) \\ z' &= 7.9958\text{m/sec} \end{aligned}$$

So, the two riders are moving apart at a rate of 7.9958 m/sec.

Every problem that we've worked to this point has come down to needing a geometric formula and we should probably work a quick problem that is not geometric in nature.

Example 9

Suppose that we have two resistors connected in parallel with resistances R_1 and R_2 measured in ohms (Ω). The total resistance, R, is then given by,

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$

Suppose that R_1 is increasing at a rate of 0.4 Ω/min and R_2 is decreasing at a rate of 0.7 Ω/min. At what rate is R changing when $R_1 = 80\,\Omega$ and $R_2 = 105\,\Omega$?

Solution

Okay, unlike the previous problems there really isn't a whole lot to do here. First, let's note that we're looking for R' and that we know $R'_1 = 0.4$ and $R'_2 = -0.7$. Be careful with the signs here.

Also, since we'll eventually need it let's determine R at the time we're interested in.

$$\frac{1}{R} = \frac{1}{80} + \frac{1}{105} = \frac{37}{1680} \qquad \Rightarrow \qquad R = \frac{1680}{37} = 45.4054\,\Omega$$

Next, we need to differentiate the equation given in the problem statement.

$$\begin{aligned} -\frac{1}{R^2}R' &= -\frac{1}{(R_1)^2}R'_1 - \frac{1}{(R_2)^2}R'_2 \\ R' &= R^2\left(\frac{1}{(R_1)^2}R'_1 + \frac{1}{(R_2)^2}R'_2\right) \end{aligned}$$

Finally, all we need to do is plug into this and do some quick computations.

$$R' = (45.4054)^2\left(\frac{1}{80^2}(0.4) + \frac{1}{105^2}(-0.7)\right) = -0.002045$$

So, it looks like R is decreasing at a rate of 0.002045 Ω/min.

We've seen quite a few related rates problems in this section that cover a wide variety of possible problems. There are still many more different kinds of related rates problems out there in the world, but the ones that we've worked here should give you a pretty good idea on how to at least start most of the problems that you're liable to run into.

3.12 Higher Order Derivatives

Let's start this section with the following function.

$$f(x) = 5x^3 - 3x^2 + 10x - 5$$

By this point we should be able to differentiate this function without any problems. Doing this we get,

$$f'(x) = 15x^2 - 6x + 10$$

Now, this is a function and so it can be differentiated. Here is the notation that we'll use for that, as well as the derivative.

$$f''(x) = \left(f'(x)\right)' = 30x - 6$$

This is called the **second derivative** and $f'(x)$ is now called the **first derivative**.

Again, this is a function, so we can differentiate it again. This will be called the **third derivative**. Here is that derivative as well as the notation for the third derivative.

$$f'''(x) = \left(f''(x)\right)' = 30$$

Continuing, we can differentiate again. This is called, oddly enough, the **fourth derivative**. We're also going to be changing notation at this point. We can keep adding on primes, but that will get cumbersome after a while.

$$f^{(4)}(x) = \left(f'''(x)\right)' = 0$$

This process can continue but notice that we will get zero for all derivatives after this point. This set of derivatives leads us to the following fact about the differentiation of polynomials.

Fact

If $p(x)$ is a polynomial of degree n (*i.e.* the largest exponent in the polynomial) then,

$$p^{(k)}(x) = 0 \qquad \text{for } k \geq n + 1$$

We will need to be careful with the "non-prime" notation for derivatives. Consider each of the following.

$$f^{(2)}(x) = f''(x)$$
$$f^2(x) = \left[f(x)\right]^2$$

The presence of parenthesis in the exponent denotes differentiation while the absence of parenthesis denotes exponentiation.

Collectively the second, third, fourth, *etc.* derivatives are called **higher order derivatives**.

Let's take a look at some examples of higher order derivatives.

Example 1

Find the first four derivatives for each of the following.

(a) $R(t) = 3t^2 + 8t^{\frac{1}{2}} + \mathbf{e}^t$

(b) $y = \cos(x)$

(c) $f(y) = \sin(3y) + \mathbf{e}^{-2y} + \ln(7y)$

Solution

(a) $R(t) = 3t^2 + 8t^{\frac{1}{2}} + \mathbf{e}^t$

There really isn't a lot to do here other than do the derivatives.

$$\begin{aligned} R'(t) &= 6t + 4t^{-\frac{1}{2}} + \mathbf{e}^t \\ R''(t) &= 6 - 2t^{-\frac{3}{2}} + \mathbf{e}^t \\ R'''(t) &= 3t^{-\frac{5}{2}} + \mathbf{e}^t \\ R^{(4)}(t) &= -\frac{15}{2}t^{-\frac{7}{2}} + \mathbf{e}^t \end{aligned}$$

Notice that differentiating an exponential function is very simple. It doesn't change with each differentiation.

(b) $y = \cos(x)$

Again, let's just do some derivatives.

$$\begin{aligned} y &= \cos(x) \\ y' &= -\sin(x) \\ y'' &= -\cos(x) \\ y''' &= \sin(x) \\ y^{(4)} &= \cos(x) \end{aligned}$$

Note that cosine (and sine) will repeat every four derivatives. The other four trig functions will not exhibit this behavior. You might want to take a few derivatives to convince yourself of this.

(c) $f(y) = \sin(3y) + \mathbf{e}^{-2y} + \ln(7y)$

In the previous two examples we saw some patterns in the differentiation of exponential functions, cosines and sines. We need to be careful however since they only work if there is just a t or an x in the argument. This is the point of this example. In this example we will need to use the chain rule on each derivative.

$$\begin{aligned} f'(y) &= 3\cos(3y) - 2\mathbf{e}^{-2y} + \frac{1}{y} = 3\cos(3y) - 2\mathbf{e}^{-2y} + y^{-1} \\ f''(y) &= -9\sin(3y) + 4\mathbf{e}^{-2y} - y^{-2} \\ f'''(y) &= -27\cos(3y) - 8\mathbf{e}^{-2y} + 2y^{-3} \\ f^{(4)}(y) &= 81\sin(3y) + 16\mathbf{e}^{-2y} - 6y^{-4} \end{aligned}$$

So, we can see with slightly more complicated arguments the patterns that we saw for exponential functions, sines and cosines no longer completely hold.

Let's do a couple more examples to make a couple of points.

Example 2

Find the second derivative for each of the following functions.

(a) $Q(t) = \sec(5t)$

(b) $g(w) = \mathbf{e}^{1-2w^3}$

(c) $f(t) = \ln(1 + t^2)$

Solution

(a) $Q(t) = \sec(5t)$

Here's the first derivative.

$$Q'(t) = 5\sec(5t)\tan(5t)$$

Notice that the second derivative will now require the product rule.

$$\begin{aligned} Q''(t) &= 25\sec(5t)\tan(5t)\tan(5t) + 25\sec(5t)\sec^2(5t) \\ &= 25\sec(5t)\tan^2(5t) + 25\sec^3(5t) \end{aligned}$$

Notice that each successive derivative will require a product and/or chain rule and that as noted above this will not end up returning back to just a secant after four (or another other number for that matter) derivatives as sine and cosine will.

(b) $g(w) = \mathbf{e}^{1-2w^3}$

Again, let's start with the first derivative.

$$g'(w) = -6w^2\mathbf{e}^{1-2w^3}$$

As with the first example we will need the product rule for the second derivative.

$$\begin{aligned} g''(w) &= -12w\mathbf{e}^{1-2w^3} - 6w^2\left(-6w^2\right)\mathbf{e}^{1-2w^3} \\ &= -12w\mathbf{e}^{1-2w^3} + 36w^4\mathbf{e}^{1-2w^3} \end{aligned}$$

(c) $f(t) = \ln\left(1+t^2\right)$

Same thing here.

$$f'(t) = \frac{2t}{1+t^2}$$

The second derivative this time will require the quotient rule.

$$\begin{aligned} f''(t) &= \frac{2\left(1+t^2\right) - (2t)(2t)}{\left(1+t^2\right)^2} \\ &= \frac{2-2t^2}{\left(1+t^2\right)^2} \end{aligned}$$

As we saw in this last set of examples we will often need to use the product or quotient rule for the higher order derivatives, even when the first derivative didn't require these rules.

Let's work one more example that will illustrate how to use implicit differentiation to find higher order derivatives.

Example 3

Find y'' for

$$x^2 + y^4 = 10$$

Solution

Okay, we know that in order to get the second derivative we need the first derivative and in order to get that we'll need to do implicit differentiation. Here is the work for that.

$$\begin{aligned} 2x + 4y^3y' &= 0 \\ y' &= -\frac{x}{2y^3} \end{aligned}$$

Now, this is the first derivative. We get the second derivative by differentiating this, which will require implicit differentiation again.

$$\begin{aligned} y'' &= \left(-\frac{x}{2y^3}\right)' \\ &= -\frac{2y^3 - x\left(6y^2y'\right)}{\left(2y^3\right)^2} \\ &= -\frac{2y^3 - 6xy^2y'}{4y^6} = -\frac{y - 3xy'}{2y^4} \end{aligned}$$

This is fine as far as it goes. However, we would like there to be no derivatives in the answer. We don't, generally, mind having x's and/or y's in the answer when doing implicit differentiation, but we really don't like derivatives in the answer. We can get rid of the derivative however by acknowledging that we know what the first derivative is and substituting this into the second derivative equation. Doing this gives,

$$\begin{aligned} y'' &= -\frac{y - 3xy'}{2y^4} \\ &= -\frac{y - 3x\left(-\frac{x}{2y^3}\right)}{2y^4} = -\frac{y + \frac{3}{2}x^2y^{-3}}{2y^4} \end{aligned}$$

Now that we've found some higher order derivatives we should probably talk about an interpretation of the second derivative.

If the position of an object is given by $s(t)$ we know that the velocity is the first derivative of the position.

$$v(t) = s'(t)$$

The acceleration of the object is the first derivative of the velocity, but since this is the first derivative of the position function we can also think of the acceleration as the second derivative of the position function.

$$a(t) = v'(t) = s''(t)$$

Alternate Notation

There is some alternate notation for higher order derivatives as well. Recall that there was a fractional notation for the first derivative.

$$f'(x) = \frac{df}{dx}$$

We can extend this to higher order derivatives.

$$f''(x) = \frac{d^2 f}{dx^2} \qquad f'''(x) = \frac{d^3 f}{dx^3} \qquad etc.$$

3.13 Logarithmic Differentiation

There is one last topic to discuss in this section. Taking the derivatives of some complicated functions can be simplified by using logarithms. This is called **logarithmic differentiation**.

It's easiest to see how this works in an example.

Example 1

Differentiate the function.

$$y = \frac{x^5}{(1-10x)\sqrt{x^2+2}}$$

Solution

Differentiating this function could be done with a product rule and a quotient rule. However, that would be a fairly messy process. We can simplify things somewhat by taking logarithms of both sides.

$$\ln(y) = \ln\left(\frac{x^5}{(1-10x)\sqrt{x^2+2}}\right)$$

Of course, this isn't really simpler. What we need to do is use the properties of logarithms to expand the right side as follows.

$$\begin{aligned}\ln(y) &= \ln\left(x^5\right) - \ln\left((1-10x)\sqrt{x^2+2}\right)\\ \ln(y) &= \ln\left(x^5\right) - \ln(1-10x) - \ln\left(\sqrt{x^2+2}\right)\end{aligned}$$

This doesn't look all that simple. However, the differentiation process will be simpler. What we need to do at this point is differentiate both sides with respect to x. Note that this is really implicit differentiation.

$$\begin{aligned}\frac{y'}{y} &= \frac{5x^4}{x^5} - \frac{-10}{1-10x} - \frac{\frac{1}{2}\left(x^2+2\right)^{-\frac{1}{2}}(2x)}{\left(x^2+2\right)^{\frac{1}{2}}}\\ \frac{y'}{y} &= \frac{5}{x} + \frac{10}{1-10x} - \frac{x}{x^2+2}\end{aligned}$$

To finish the problem all that we need to do is multiply both sides by y and the plug in for y since we do know what that is.

$$\begin{aligned}y' &= y\left(\frac{5}{x} + \frac{10}{1-10x} - \frac{x}{x^2+2}\right)\\ &= \frac{x^5}{(1-10x)\sqrt{x^2+2}}\left(\frac{5}{x} + \frac{10}{1-10x} - \frac{x}{x^2+2}\right)\end{aligned}$$

Depending upon the person, doing this would probably be slightly easier than doing both the product and quotient rule. The answer is almost definitely simpler than what we would have gotten using the product and quotient rule.

So, as the first example has shown we can use logarithmic differentiation to avoid using the product rule and/or quotient rule.

We can also use logarithmic differentiation to differentiate functions in the form.

$$y = \left(f\left(x\right)\right)^{g(x)}$$

Let's take a quick look at a simple example of this.

Example 2

Differentiate $y = x^x$.

Solution

We've seen two functions similar to this at this point.

$$\frac{d}{dx}\left(x^n\right) = nx^{n-1} \qquad \frac{d}{dx}\left(a^x\right) = a^x \ln(a)$$

Neither of these two will work here since both require either the base or the exponent to be a constant. In this case both the base and the exponent are variables and so we have no way to differentiate this function using only known rules from previous sections.

With logarithmic differentiation we can do this however. First take the logarithm of both sides as we did in the first example and use the logarithm properties to simplify things a little.

$$\begin{aligned} \ln(y) &= \ln x^x \\ \ln(y) &= x \ln(x) \end{aligned}$$

Differentiate both sides using implicit differentiation.

$$\frac{y'}{y} = \ln(x) + x\left(\frac{1}{x}\right) = \ln(x) + 1$$

As with the first example multiply by y and substitute back in for y.

$$\begin{aligned} y' &= y\big(1 + \ln(x)\big) \\ &= x^x\big(1 + \ln(x)\big) \end{aligned}$$

Let's take a look at a more complicated example of this.

Example 3

Differentiate $y = (1 - 3x)^{\cos(x)}$.

Solution

Now, this looks much more complicated than the previous example, but is in fact only slightly more complicated. The process is pretty much identical, so we first take the log of both sides and then simplify the right side.

$$\ln(y) = \ln\left[(1 - 3x)^{\cos(x)}\right] = \cos(x)\ln(1 - 3x)$$

Next, do some implicit differentiation.

$$\frac{y'}{y} = -\sin(x)\ln(1 - 3x) + \cos(x)\frac{-3}{1 - 3x} = -\sin(x)\ln(1 - 3x) - \cos(x)\frac{3}{1 - 3x}$$

Finally, solve for y' and substitute back in for y.

$$\begin{aligned} y' &= -y\left(\sin(x)\ln(1 - 3x) + \cos(x)\frac{3}{1 - 3x}\right) \\ &= -(1 - 3x)^{\cos(x)}\left(\sin(x)\ln(1 - 3x) + \cos(x)\frac{3}{1 - 3x}\right) \end{aligned}$$

A messy answer but there it is.

We'll close this section out with a quick recap of all the various ways we've seen of differentiating functions with exponents. It is important to not get all of these confused.

$\frac{d}{dx}\left(a^b\right) = 0$	This is a constant
$\frac{d}{dx}\left(x^n\right) = nx^{n-1}$	Power Rule
$\frac{d}{dx}\left(a^x\right) = a^x\ln(a)$	Derivative of an exponential function
$\frac{d}{dx}\left(x^x\right) = x^x\left(1 + \ln(x)\right)$	Logarithmic Differentiation

It is sometimes easy to get these various functions confused and use the wrong rule for differentiation. Always remember that each rule has very specific rules for where the variable and constants must be. For example, the Power Rule requires that the base be a variable and the exponent be a constant, while the exponential function requires exactly the opposite.

If you can keep straight all the rules you can't go wrong with these.

4 Derivative Applications

In the previous chapter we focused almost exclusively, with the exception of Related Rates, on the computation and interpretation of derivatives. In this chapter will focus on applications of derivatives and it is important to always remember that we didn't spend a whole chapter talking about how to compute derivatives just to be talking about them. Each of the applications here will require us to compute at least one derivative and that computation will often be the very first step in the problem. So, if you are rusty with your differentiation skills you will need to go back to the previous chapter and scrap some of that rust off, so to speak, or you will find yourself struggling a lot in this chapter.

There are quite a few important applications to derivatives but almost all of them require the use of something called a critical point. So, we will first need to define just what a critical point is and make sure we are comfortable with finding critical points.

Once we have a good understanding of critical points we will turn our focus to the first application that we'll be spending quite bit of time on in this chapter. Namely, how we can use derivatives to find some important information about a function. We already know how to determine if a function is increasing or decreasing as we discussed that and worked a few problems on that in the last chapter. We will work some more problems involving increasing and decreasing functions to make sure we are clear on how that works. As we know from the last chapter we use the first derivative to determine where a function is increasing and decreasing. So we will then move on to see what the second derivative can tell us about a function. As we will see the second derivative can be used to determine the concavity of a function. The concavity of a function gives, in some way, the "curvature" of a function.

Once we have discussed all the information that derivatives can tell us about a function we'll use that information to get a sketch of the graph of a function without any kind of computational aid outside of occasionally needing a calculator to compute the value of the function at a few points. As we'll see we will often get a fairly good sketch of the graph from just this information.

The other topic that we will focus on in this chapter will be optimizing functions. By optimizing a function we mean finding the minimum and maximum value that a function can take. In addition, we will, on occasion, include a constraint on the function we are trying to optimize. The constraint will be an additional equation that the variable(s) in the function we are optimizing must also satisfy.

We will also take a quick look at a couple of other applications. These will include linear approximations (*i.e.* find a linear function that can approximate the function for at least a range of variables), Newton's Method (*i.e.* approximating the solution to an equation) as well as a couple of applications of derivatives to some business applications.

We will also briefly revisit limits to discuss L'Hospital's Rule. This is a method of computing some limits of functions that are of "indeterminate form" (defined later) that we cannot, at this point, compute. So, why did we not discuss L'Hospital's Rule back in the Limits chapter? That is a valid question and it has a simple answer. We couldn't discuss L'Hospitals's Rule until this point because it involved taking some derivatives which we (clearly) did not yet know when we first looked at limits.

4.1 Rates of Change

The purpose of this section is to remind us of one of the more important applications of derivatives. That is the fact that $f'(x)$ represents the rate of change of $f(x)$. This is an application that we repeatedly saw in the previous chapter. Almost every section in the previous chapter contained at least one problem dealing with this application of derivatives. While this application will arise occasionally in this chapter we are going to focus more on other applications in this chapter.

So, to make sure that we don't forget about this application here is a brief set of examples concentrating on the rate of change application of derivatives. Note that the point of these examples is to remind you of material covered in the previous chapter and not to teach you how to do these kinds of problems. If you don't recall how to do these kinds of examples you'll need to go back and review the previous chapter.

Example 1

Determine all the points where the following function is not changing.

$$g(x) = 5 - 6x - 10\cos(2x)$$

Solution

First, we'll need to take the derivative of the function.

$$g'(x) = -6 + 20\sin(2x)$$

Now, the function will not be changing if the rate of change is zero and so to answer this question we need to determine where the derivative is zero. So, let's set this equal to zero and solve.

$$-6 + 20\sin(2x) = 0 \qquad \Rightarrow \qquad \sin(2x) = \frac{6}{20} = 0.3$$

The solution to this is then,

$$\begin{aligned} 2x &= 0.3047 + 2\pi n \qquad \text{OR} \qquad & 2x &= 2.8369 + 2\pi n \qquad n = 0, \pm 1, \pm 2, \ldots \\ x &= 0.1524 + \pi n \qquad \text{OR} \qquad & x &= 1.4185 + \pi n \qquad n = 0, \pm 1, \pm 2, \ldots \end{aligned}$$

If you don't recall how to solve trig equations check out the Solving Trig Equations sections in the Review Chapter.

Example 2

Determine where the following function is increasing and decreasing.

$$A(t) = 27t^5 - 45t^4 - 130t^3 + 150$$

Solution

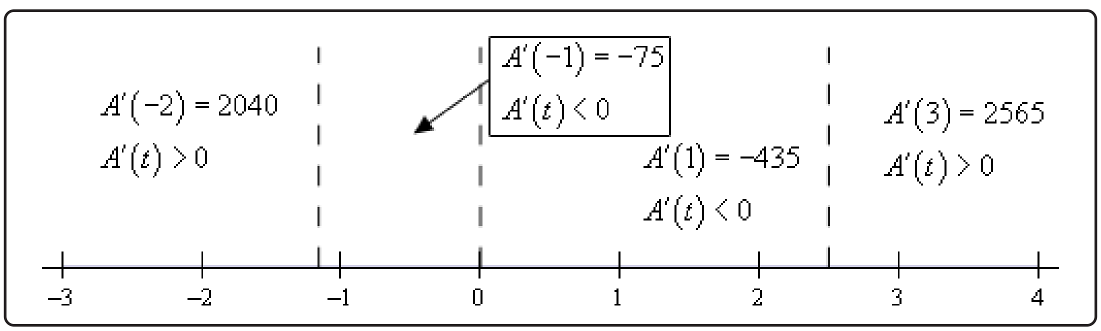
As with the first problem we first need to take the derivative of the function.

$$A'(t) = 135t^4 - 180t^3 - 390t^2 = 15t^2\left(9t^2 - 12t - 26\right)$$

Next, we need to determine where the function isn't changing. This is at,

$$\begin{aligned} t &= 0 \\ t &= \frac{12 \pm \sqrt{144 - 4(9)(-26)}}{18} = \frac{12 \pm \sqrt{1080}}{18} \\ &= \frac{12 \pm 6\sqrt{30}}{18} = \frac{2 \pm \sqrt{30}}{3} = -1.159, \quad 2.492 \end{aligned}$$

So, the function is not changing at three values of t. Finally, to determine where the function is increasing or decreasing we need to determine where the derivative is positive or negative. Recall that if the derivative is positive then the function must be increasing and if the derivative is negative then the function must be decreasing. The following number line gives this information.



So, from this number line we can see that we have the following increasing and decreasing information.

$$\text{Increasing} : -\infty < t < -1.159, \quad 2.492 < t < \infty$$
$$\text{Decreasing} : -1.159 < t < 0, \quad 0 < t < 2.492$$

If you don't remember how to solve polynomial and rational inequalities then you should check out the appropriate sections in the Review Chapter.

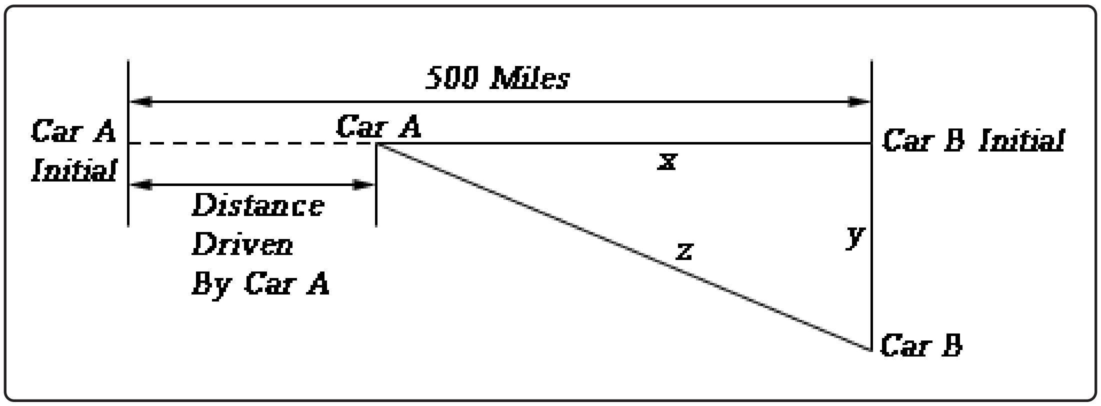
Finally, we can't forget about Related Rates problems.

Example 3

Two cars start out 500 miles apart. Car A is to the west of Car B and starts driving to the east (*i.e.* towards Car B) at 35 mph and at the same time Car B starts driving south at 50 mph. After 3 hours of driving at what rate is the distance between the two cars changing? Is it increasing or decreasing?

Solution

The first thing to do here is to get sketch a figure showing the situation.



In this figure y represents the distance driven by Car B and x represents the distance separating Car A from Car B's initial position and z represents the distance separating the two cars. After 3 hours driving time with have the following values of x and y.

$$x = 500 - 35(3) = 395 \qquad\qquad y = 50(3) = 150$$

We can use the Pythagorean theorem to find z at this time as follows,

$$z^2 = 395^2 + 150^2 = 178525 \qquad \Rightarrow \qquad z = \sqrt{178525} = 422.5222$$

Now, to answer this question we will need to determine z' given that $x' = -35$ and $y' = 50$. Do you agree with the signs on the two given rates? Remember that a rate is negative if the quantity is decreasing and positive if the quantity is increasing.

We can again use the Pythagorean theorem here. First, write it down and the remember that x, y, and z are all changing with time and so differentiate the equation using Implicit Differentiation.

$$z^2 = x^2 + y^2 \qquad \Rightarrow \qquad 2zz' = 2xx' + 2yy'$$

Finally, all we need to do is cancel a 2 from everything, plug in for the known quantities and solve for z'.

$$z'\left(422.5222\right) = \left(395\right)\left(-35\right) + \left(150\right)\left(50\right) \qquad \Rightarrow \qquad z' = \frac{-6325}{422.5222} = -14.9696$$

So, after three hours the distance between them is decreasing at a rate of 14.9696 mph.

So, in this section we covered three "standard" problems using the idea that the derivative of a function gives the rate of change of the function. As mentioned earlier, this chapter will be focusing more on other applications than the idea of rate of change, however, we can't forget this application as it is a very important one.

4.2 Critical Points

Critical points will show up throughout a majority of this chapter so we first need to define them and work a few examples before getting into the sections that actually use them.

Definition

We say that $x = c$ is a critical point of the function $f(x)$ if $f(c)$ exists and if either of the following are true.

$$f'(c) = 0 \qquad \text{OR} \qquad f'(c) \ \text{ doesn't exist}$$

Note that we require that $f(c)$ exists in order for $x = c$ to actually be a critical point. This is an important, and often overlooked, point. What this is really saying is that all critical points must be in the domain of the function. If a point is not in the domain of the function then it is not a critical point.

Note as well that, at this point, we only work with real numbers and so any complex numbers that might arise in finding critical points (and they will arise on occasion) will be ignored. There are portions of calculus that work a little differently when working with complex numbers and so in a first calculus class such as this we ignore complex numbers and only work with real numbers. Calculus with complex numbers is beyond the scope of this course and is usually taught in higher level mathematics courses.

The main point of this section is to work some examples finding critical points. So, let's work some examples.

Example 1

Determine all the critical points for the function.

$$f(x) = 6x^5 + 33x^4 - 30x^3 + 100$$

Solution

We first need the derivative of the function in order to find the critical points and so let's get that and notice that we'll factor it as much as possible to make our life easier when we go to find the critical points.

$$\begin{aligned} f'(x) &= 30x^4 + 132x^3 - 90x^2 \\ &= 6x^2\left(5x^2 + 22x - 15\right) \\ &= 6x^2(5x - 3)(x + 5) \end{aligned}$$

Now, our derivative is a polynomial and so will exist everywhere. Therefore, the only critical points will be those values of x which make the derivative zero. So, we must solve.

$$6x^2\left(5x-3\right)\left(x+5\right)=0$$

Because this is the factored form of the derivative it's pretty easy to identify the three critical points. They are,

$$x=-5,\quad x=0,\quad x=\frac{3}{5}$$

Polynomials are usually fairly simple functions to find critical points for provided the degree doesn't get so large that we have trouble finding the roots of the derivative.

Most of the more "interesting" functions for finding critical points aren't polynomials however. So let's take a look at some functions that require a little more effort on our part.

Example 2

Determine all the critical points for the function.

$$g\left(t\right)=\sqrt[3]{t^2}\left(2t-1\right)$$

Solution

To find the derivative it's probably easiest to do a little simplification before we actually differentiate. Let's multiply the root through the parenthesis and simplify as much as possible. This will allow us to avoid using the product rule when taking the derivative.

$$g\left(t\right)=t^{\frac{2}{3}}\left(2t-1\right)=2t^{\frac{5}{3}}-t^{\frac{2}{3}}$$

Now differentiate.

$$g'\left(t\right)=\frac{10}{3}t^{\frac{2}{3}}-\frac{2}{3}t^{-\frac{1}{3}}=\frac{10t^{\frac{2}{3}}}{3}-\frac{2}{3t^{\frac{1}{3}}}$$

We will need to be careful with this problem. When faced with a negative exponent it is often best to eliminate the minus sign in the exponent as we did above. This isn't really required but it can make our life easier on occasion if we do that.

Notice as well that eliminating the negative exponent in the second term allows us to correctly identify why $t=0$ is a critical point for this function. Once we move the second term to the denominator we can clearly see that the derivative doesn't exist at $t=0$ and so this will be a critical point. If you don't get rid of the negative exponent in the second term many people will incorrectly state that $t=0$ is a critical point because the derivative is zero at $t=0$. While this may seem like a silly point, after all in each case $t = 0$ is identified as a critical point, it is sometimes important to know why a point is a critical point. In fact, in a couple of sections we'll see a fact that only works for critical points in which the derivative is zero.

So, we've found one critical point (where the derivative doesn't exist), but we now need to determine where the derivative is zero (provided it is of course...). To help with this it's usually best to combine the two terms into a single rational expression. So, getting a common denominator and combining gives us,

$$g'(t) = \frac{10t - 2}{3t^{\frac{1}{3}}}$$

Notice that we still have $t = 0$ as a critical point. Doing this kind of combining should never lose critical points, it's only being done to help us find them. As we can see it's now become much easier to quickly determine where the derivative will be zero. Recall that a rational expression will only be zero if its numerator is zero (and provided the denominator isn't also zero at that point of course).

So, in this case we can see that the numerator will be zero if $t = \frac{1}{5}$ and so there are two critical points for this function.

$$t = 0 \qquad \text{and} \qquad t = \frac{1}{5}$$

Example 3

Determine all the critical points for the function.

$$R(w) = \frac{w^2 + 1}{w^2 - w - 6}$$

Solution

We'll leave it to you to verify that using the quotient rule, along with some simplification, we get that the derivative is,

$$R'(w) = \frac{-w^2 - 14w + 1}{(w^2 - w - 6)^2} = -\frac{w^2 + 14w - 1}{(w^2 - w - 6)^2}$$

Notice that we factored a "-1" out of the numerator to help a little with finding the critical points. This negative out in front will not affect the derivative whether or not the derivative is zero or not exist but will make our work a little easier.

Now, we have two issues to deal with. First the derivative will not exist if there is division by zero in the denominator. So we need to solve,

$$w^2 - w - 6 = (w-3)(w+2) = 0$$

We didn't bother squaring this since if this is zero, then zero squared is still zero and if it isn't zero then squaring it won't make it zero.

So, we can see from this that the derivative will not exist at $w = 3$ and $w = -2$. However, these are NOT critical points since the function will also not exist at these points. Recall that in order for a point to be a critical point the function must actually exist at that point.

At this point we need to be careful. The numerator doesn't factor, but that doesn't mean that there aren't any critical points where the derivative is zero. We can use the quadratic formula on the numerator to determine if the fraction as a whole is ever zero.

$$w = \frac{-14 \pm \sqrt{(14)^2 - 4(1)(-1)}}{2(1)} = \frac{-14 \pm \sqrt{200}}{2} = \frac{-14 \pm 10\sqrt{2}}{2} = -7 \pm 5\sqrt{2}$$

So, we get two critical points. Also, these are not "nice" integers or fractions. This will happen on occasion. Don't get too locked into answers always being "nice". Often they aren't.

Note as well that we only use real numbers for critical points. So, if upon solving the quadratic in the numerator, we had gotten complex number these would not have been considered critical points.

Summarizing, we have two critical points. They are,

$$w = -7 + 5\sqrt{2}, \quad w = -7 - 5\sqrt{2}$$

Again, remember that while the derivative doesn't exist at $w = 3$ and $w = -2$ neither does the function and so these two points are not critical points for this function.

In the previous example we had to use the quadratic formula to determine some potential critical points. We know that sometimes we will get complex numbers out of the quadratic formula. Just remember that, as mentioned at the start of this section, when that happens we will ignore the complex numbers that arise.

So far all the examples have not had any trig functions, exponential functions, *etc.* in them. We shouldn't expect that to always be the case. So, let's take a look at some examples that don't just involve powers of x.

Example 4

Determine all the critical points for the function.

$$y = 6x - 4\cos(3x)$$

Solution

First get the derivative and don't forget to use the chain rule on the second term.

$$y' = 6 + 12\sin(3x)$$

Now, this will exist everywhere and so there won't be any critical points for which the derivative doesn't exist. The only critical points will come from points that make the derivative zero. We will need to solve,

$$\begin{aligned} 6 + 12\sin(3x) &= 0 \\ \sin(3x) &= -\frac{1}{2} \end{aligned}$$

Solving this equation gives the following.

$$\begin{aligned} 3x &= 3.6652 + 2\pi n, \qquad n = 0, \pm 1, \pm 2, \ldots \\ 3x &= 5.7596 + 2\pi n, \qquad n = 0, \pm 1, \pm 2, \ldots \end{aligned}$$

Don't forget the $2\pi n$ on these! There will be problems down the road in which we will miss solutions without this! Also make sure that it gets put on at this stage! Now divide by 3 to get all the critical points for this function.

$$\begin{aligned} x &= 1.2217 + \frac{2\pi n}{3}, \qquad n = 0, \pm 1, \pm 2, \ldots \\ x &= 1.9199 + \frac{2\pi n}{3}, \qquad n = 0, \pm 1, \pm 2, \ldots \end{aligned}$$

Notice that in the previous example we got an infinite number of critical points. That will happen on occasion so don't worry about it when it happens.

Example 5

Determine all the critical points for the function.

$$h(t) = 10t\mathbf{e}^{3-t^2}$$

Solution

Here's the derivative for this function.

$$h'(t) = 10\mathbf{e}^{3-t^2} + 10t\mathbf{e}^{3-t^2}(-2t) = 10\mathbf{e}^{3-t^2} - 20t^2\mathbf{e}^{3-t^2}$$

Now, this looks unpleasant, however with a little factoring we can clean things up a little as follows,

$$h'(t) = 10\mathbf{e}^{3-t^2}\left(1 - 2t^2\right)$$

This function will exist everywhere, so no critical points will come from the derivative not existing. Determining where this is zero is easier than it looks. We know that exponentials are never zero and so the only way the derivative will be zero is if,

$$\begin{aligned} 1 - 2t^2 &= 0 \\ 1 &= 2t^2 \\ \frac{1}{2} &= t^2 \end{aligned}$$

We will have two critical points for this function.

$$t = \pm\frac{1}{\sqrt{2}}$$

Example 6

Determine all the critical points for the function.

$$f(x) = x^2 \ln(3x) + 6$$

Solution

Before getting the derivative let's notice that since we can't take the log of a negative number or zero we will only be able to look at $x > 0$.

The derivative is then,

$$\begin{aligned} f'(x) &= 2x\ln(3x) + x^2\left(\frac{3}{3x}\right) \\ &= 2x\ln(3x) + x \\ &= x(2\ln(3x) + 1) \end{aligned}$$

Now, this derivative will not exist if x is a negative number or if $x = 0$, but then again neither will the function and so these are not critical points. Remember that the function will only exist if $x > 0$ and nicely enough the derivative will also only exist if $x > 0$ and so the only thing we need to worry about is where the derivative is zero.

First note that, despite appearances, the derivative will not be zero for $x = 0$. As noted above the derivative doesn't exist at $x = 0$ because of the natural logarithm and so the derivative can't be zero there!

So, the derivative will only be zero if,

$$\begin{aligned} 2\ln(3x) + 1 &= 0 \\ \ln(3x) &= -\frac{1}{2} \end{aligned}$$

Recall that we can solve this by exponentiating both sides.

$$\begin{aligned} \mathbf{e}^{\ln(3x)} &= \mathbf{e}^{-\frac{1}{2}} \\ 3x &= \mathbf{e}^{-\frac{1}{2}} \\ x &= \frac{1}{3}\mathbf{e}^{-\frac{1}{2}} = \frac{1}{3\sqrt{\mathbf{e}}} \end{aligned}$$

There is a single critical point for this function.

Let's work one more problem to make a point.

Example 7

Determine all the critical points for the function.

$$f(x) = x\mathbf{e}^{x^2}$$

Solution

Note that this function is not much different from the function used in Example 5. In this case the derivative is,

$$f'(x) = \mathbf{e}^{x^2} + x\mathbf{e}^{x^2}(2x) = \mathbf{e}^{x^2}\left(1 + 2x^2\right)$$

This function will never be zero for any real value of x. The exponential is never zero of course and the polynomial will only be zero if x is complex and recall that we only want real values of x for critical points.

Therefore, this function will not have any critical points.

It is important to note that not all functions will have critical points! In this course most of the functions that we will be looking at do have critical points. That is only because those problems make for more interesting examples. Do not let this fact lead you to always expect that a function will have critical points. Sometimes they don't as this final example has shown.

4.3 Minimum and Maximum Values

Many of our applications in this chapter will revolve around minimum and maximum values of a function. While we can all visualize the minimum and maximum values of a function we want to be a little more specific in our work here. In particular, we want to differentiate between two types of minimum or maximum values. The following definition gives the types of minimums and/or maximums values that we'll be looking at.

Minimums & Maximums

Definition

1. We say that $f(x)$ has an **absolute (or global) maximum** at $x = c$ if $f(x) \leq f(c)$ for every x in the domain we are working on.
2. We say that $f(x)$ has a **relative (or local) maximum** at $x = c$ if $f(x) \leq f(c)$ for every x in some open interval around $x = c$.
3. We say that $f(x)$ has an **absolute (or global) minimum** at $x = c$ if $f(x) \geq f(c)$ for every x in the domain we are working on.
4. We say that $f(x)$ has a **relative (or local) minimum** at $x = c$ if $f(x) \geq f(c)$ for every x in some open interval around $x = c$.

Note that when we say an "open interval around $x = c$" we mean that we can find some interval (a, b), not including the endpoints, such that $a < c < b$. Or, in other words, c will be contained somewhere inside the interval and will not be either of the endpoints.

Also, we will collectively call the minimum and maximum points of a function the **extrema** of the function. So, relative extrema will refer to the relative minimums and maximums while absolute extrema refer to the absolute minimums and maximums.

Now, let's talk a little bit about the subtle difference between the absolute and relative in the definition above.

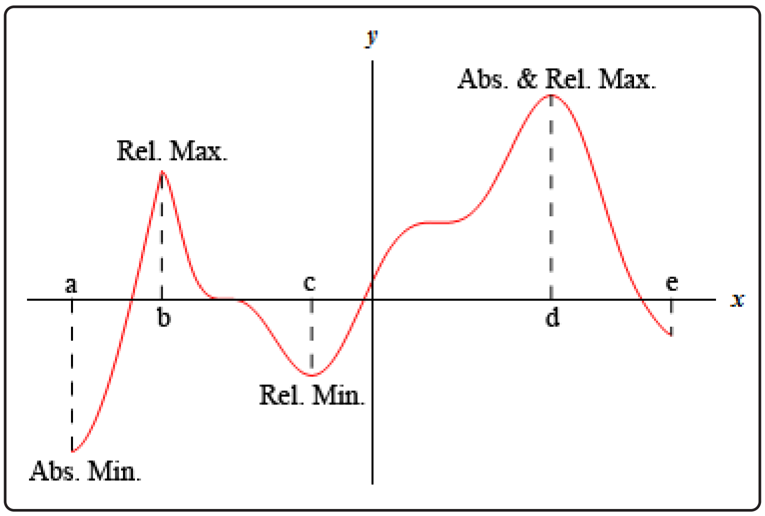
We will have an absolute maximum (or minimum) at $x = c$ provided $f(c)$ is the largest (or smallest) value that the function will ever take on the domain that we are working on. Also, when we say the "domain we are working on" this simply means the range of x's that we have chosen to work with for a given problem. There may be other values of x that we can actually plug into the function but have excluded them for some reason.

A relative maximum or minimum is slightly different. All that's required for a point to be a relative maximum or minimum is for that point to be a maximum or minimum in some interval of x's around $x = c$. There may be larger or smaller values of the function at some other place, but relative to $x = c$, or local to $x = c$, $f(c)$ is larger or smaller than all the other function values that are near it.

Note as well that in order for a point to be a relative extrema we must be able to look at function values on both sides of $x = c$ to see if it really is a maximum or minimum at that point. This means that relative extrema do not occur at the end points of a domain. They can only occur interior to the domain.

There is actually some debate on the preceding point. Some folks do feel that relative extrema can occur on the end points of a domain. However, in this class we will be using the definition that says that they can't occur at the end points of a domain. This will be discussed in a little more detail at the end of the section once we have a relevant fact taken care of.

It's usually easier to get a feel for the definitions by taking a quick look at a graph.



For the function shown in this graph we have relative maximums at $x = b$ and $x = d$. Both of these points are relative maximums since they are interior to the domain shown and are the largest point on the graph in some interval around the point. We also have a relative minimum at $x = c$ since this point is interior to the domain and is the lowest point on the graph in an interval around it. The far-right end point, $x = e$, will not be a relative minimum since it is an end point.

The function will have an absolute maximum at $x = d$ and an absolute minimum at $x = a$. These two points are the largest and smallest that the function will ever be. We can also notice that the absolute extrema for a function will occur at either the endpoints of the domain or at relative extrema. We will use this idea in later sections so it's more important than it might seem at the present time.

Let's take a quick look at some examples to make sure that we have the definitions of absolute extrema and relative extrema straight.

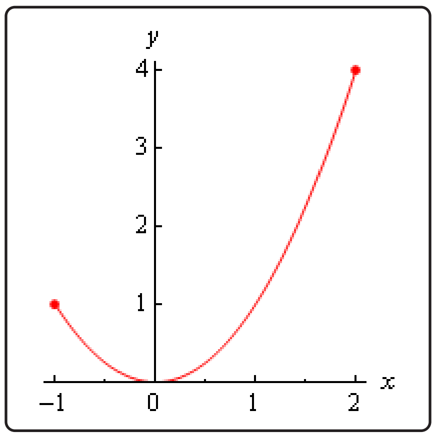
Example 1

Identify the absolute extrema and relative extrema for the following function.

$$f(x) = x^2 \qquad \text{on} \qquad [-1, 2]$$

Solution

Since this function is easy enough to graph let's do that. However, we only want the graph on the interval $[-1, 2]$. Here is the graph,



Note that we used dots at the end of the graph to remind us that the graph ends at these points.

We can now identify the extrema from the graph. It looks like we've got a relative and absolute minimum of zero at $x = 0$ and an absolute maximum of four at $x = 2$. Note that $x = -1$ is not a relative maximum since it is at the end point of the interval.

This function doesn't have any relative maximums.

As we saw in the previous example functions do not have to have relative extrema. It is completely possible for a function to not have a relative maximum and/or a relative minimum.

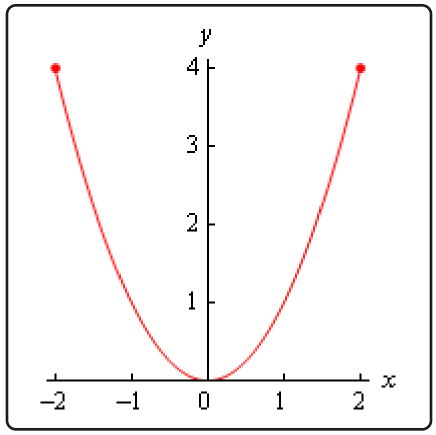
Example 2

Identify the absolute extrema and relative extrema for the following function.

$$f(x) = x^2 \qquad \text{on} \qquad [-2, 2]$$

Solution

Here is the graph for this function.



In this case we still have a relative and absolute minimum of zero at $x = 0$. We also still have an absolute maximum of four. However, unlike the first example this will occur at two points, $x = -2$ and $x = 2$.

Again, the function doesn't have any relative maximums.

As this example has shown there can only be a single absolute maximum or absolute minimum value, but they can occur at more than one place in the domain.

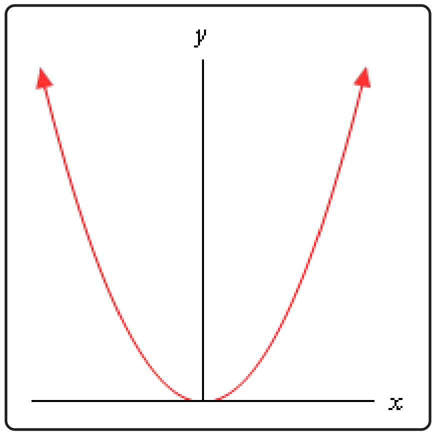
Example 3

Identify the absolute extrema and relative extrema for the following function.

$$f(x) = x^2$$

Solution

In this case we've given no domain and so the assumption is that we will take the largest possible domain. For this function that means all the real numbers. Here is the graph.



In this case the graph doesn't stop increasing at either end and so there are no maximums of any kind for this function. No matter which point we pick on the graph there will be points both larger and smaller than it on either side so we can't have any maximums (of any kind, relative or absolute) in a graph.

We still have a relative and absolute minimum value of zero at $x = 0$.

So, some graphs can have minimums but not maximums. Likewise, a graph could have maximums but not minimums.

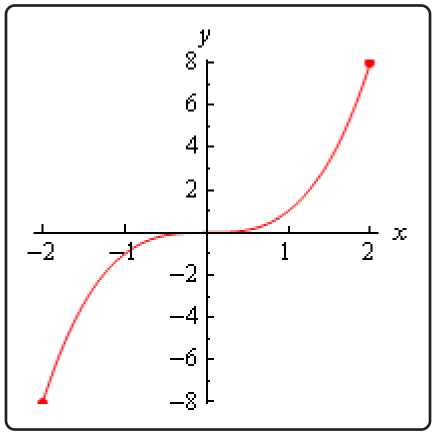
Example 4

Identify the absolute extrema and relative extrema for the following function.

$$f(x) = x^3 \qquad \text{on} \qquad [-2, 2]$$

Solution

Here is the graph for this function.



This function has an absolute maximum of eight at $x = 2$ and an absolute minimum of negative eight at $x = -2$. This function has no relative extrema.

So, a function doesn't have to have relative extrema as this example has shown.

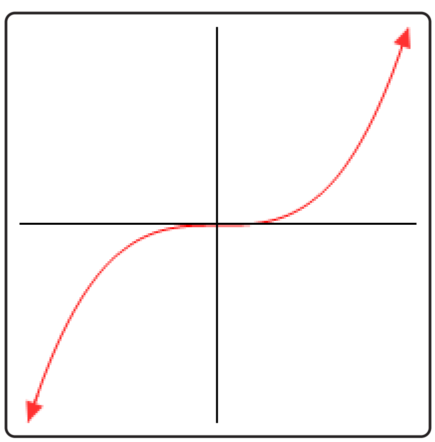
Example 5

Identify the absolute extrema and relative extrema for the following function.

$$f(x) = x^3$$

Solution

Again, we aren't restricting the domain this time so here's the graph.



In this case the function has no relative extrema and no absolute extrema.

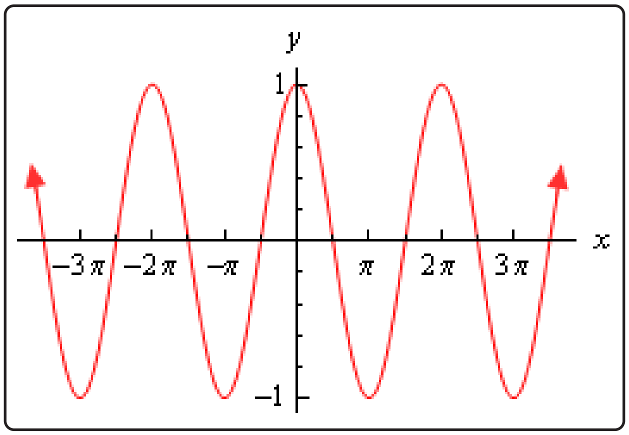
As we've seen in the previous example functions don't have to have any kind of extrema, relative or absolute.

Example 6

$$f(x) = \cos(x)$$

Solution

We've not restricted the domain for this function. Here is the graph.



Cosine has extrema (relative and absolute) that occur at many points. Cosine has both relative and absolute maximums of 1 at

$$x = \ldots - 4\pi,\ -2\pi,\ 0,\ 2\pi,\ 4\pi, \ldots$$

Cosine also has both relative and absolute minimums of -1 at

$$x = \ldots - 3\pi,\ -\pi,\ \pi,\ 3\pi, \ldots$$

As this example has shown a graph can in fact have extrema occurring at a large number (infinite in this case) of points.

We've now worked quite a few examples and we can use these examples to see a nice fact about absolute extrema. First let's notice that all the functions above were continuous functions. Next notice that every time we restricted the domain to a closed interval (*i.e.* the interval contains its end points) we got absolute maximums and absolute minimums. Finally, in only one of the three examples in which we did not restrict the domain did we get both an absolute maximum and an absolute minimum.

These observations lead us the following theorem.

Extreme Value Theorem

Suppose that $f(x)$ is continuous on the interval $[a, b]$ then there are two numbers $a \le c, d \le b$ so that $f(c)$ is an absolute maximum for the function and $f(d)$ is an absolute minimum for the function.

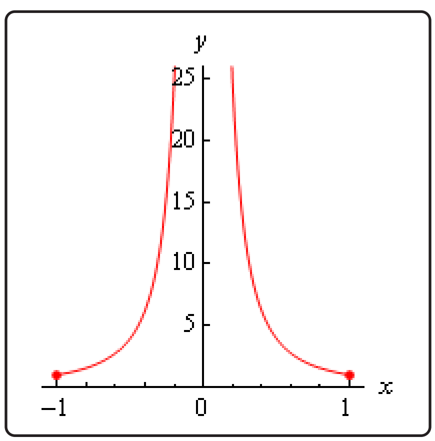
So, if we have a continuous function on an interval $[a, b]$ then we are guaranteed to have both an absolute maximum and an absolute minimum for the function somewhere in the interval. The theorem doesn't tell us where they will occur or if they will occur more than once, but at least it tells us that they do exist somewhere. Sometimes, all that we need to know is that they do exist.

This theorem doesn't say anything about absolute extrema if we aren't working on an interval. We saw examples of functions above that had both absolute extrema, one absolute extrema, and no absolute extrema when we didn't restrict ourselves down to an interval.

The requirement that a function be continuous is also required in order for us to use the theorem. Consider the case of

$$f(x) = \frac{1}{x^2} \qquad \text{on} \qquad [-1, 1]$$

Here's the graph.



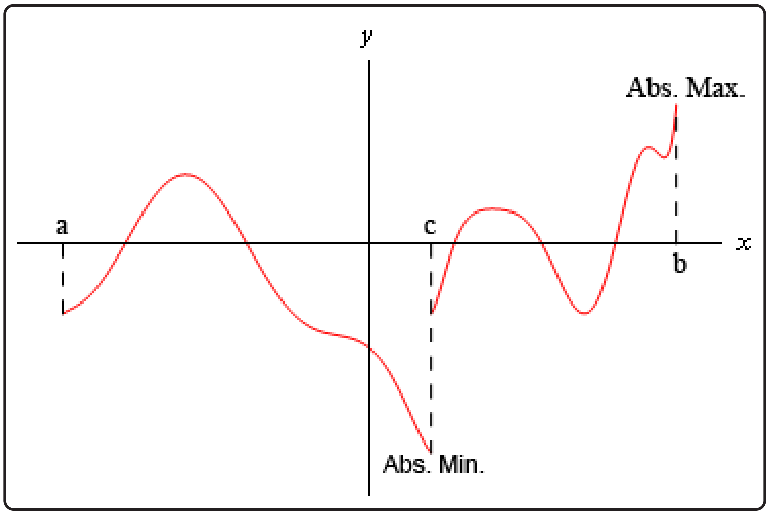
This function is not continuous at $x = 0$ as we move in towards zero the function is approaching infinity. So, the function does not have an absolute maximum. Note that it does have an absolute minimum however. In fact the absolute minimum occurs twice at both $x = -1$ and $x = 1$.

If we changed the interval a little to say,

$$f(x) = \frac{1}{x^2} \qquad \text{on} \qquad \left[\frac{1}{2}, 1\right]$$

the function would now have both absolute extrema. We may only run into problems if the interval contains the point of discontinuity. If it doesn't then the theorem will hold.

We should also point out that just because a function is not continuous at a point that doesn't mean that it won't have both absolute extrema in an interval that contains that point. Below is the graph of a function that is not continuous at a point in the given interval and yet has both absolute extrema.



This graph is not continuous at $x = c$, yet it does have both an absolute maximum ($x = b$) and an absolute minimum ($x = c$). Also note that, in this case one of the absolute extrema occurred at the point of discontinuity, but it doesn't need to. The absolute minimum could just have easily been at the other end point or at some other point interior to the region. The point here is that this graph is not continuous and yet does have both absolute extrema

The point of all this is that we need to be careful to only use the Extreme Value Theorem when the conditions of the theorem are met and not misinterpret the results if the conditions aren't met.

In order to use the Extreme Value Theorem we must have an interval that includes its endpoints, often called a closed interval, and the function must be continuous on that interval. If we don't have a closed interval and/or the function isn't continuous on the interval then the function may or may not have absolute extrema.

We need to discuss one final topic in this section before moving on to the first major application of the derivative that we're going to be looking at in this chapter.

Fermat's Theorem

If $f(x)$ has a relative extrema at $x = c$ and $f'(c)$ exists then $x = c$ is a critical point of $f(x)$. In fact, it will be a critical point such that $f'(c) = 0$.

To see the proof of this theorem see the Proofs From Derivative Applications section of the Extras appendix.

Also note that we can say that $f'(c) = 0$ because we are also assuming that $f'(c)$ exists.

This theorem tells us that there is a nice relationship between relative extrema and critical points. In fact, it will allow us to get a list of all possible relative extrema. Since a relative extrema must be a critical point the list of all critical points will give us a list of all possible relative extrema.

Consider the case of $f(x) = x^2$. We saw that this function had a relative minimum at $x = 0$ in several earlier examples. So according to Fermat's theorem $x = 0$ should be a critical point. The derivative of the function is,

$$f'(x) = 2x$$

Sure enough $x = 0$ is a critical point.

Be careful not to misuse this theorem. It doesn't say that a critical point will be a relative extrema. To see this, consider the following case.

$$f(x) = x^3 \qquad\qquad f'(x) = 3x^2$$

Clearly $x = 0$ is a critical point. However, we saw in an earlier example this function has no relative extrema of any kind. So, critical points do not have to be relative extrema.

Also note that this theorem says nothing about absolute extrema. An absolute extrema may or may not be a critical point.

Before we leave this section we need to discuss a couple of issues.

First, Fermat's Theorem only works for critical points in which $f'(c) = 0$. This does not, however, mean that relative extrema won't occur at critical points where the derivative does not exist. To see this consider $f(x) = |x|$. This function clearly has a relative minimum at $x = 0$ and yet in a previous section we showed in an example that $f'(0)$ does not exist.

What this all means is that if we want to locate relative extrema all we really need to do is look at the critical points as those are the places where relative extrema may exist.

Finally, recall that at that start of the section we stated that relative extrema will not exist at endpoints of the interval we are looking at. The reason for this is that if we allowed relative extrema to occur there it may well (and in fact most of the time) violate Fermat's Theorem. There is no reason to expect end points of intervals to be critical points of any kind. Therefore, we do not allow relative extrema to exist at the endpoints of intervals.

4.4 Finding Absolute Extrema

It's now time to see our first major application of derivatives in this chapter. Given a continuous function, $f(x)$, on an interval $[a, b]$ we want to determine the absolute extrema of the function. To do this we will need many of the ideas that we looked at in the previous section.

First, since we have a closed interval (*i.e.* and interval that includes the endpoints) and we are assuming that the function is continuous the Extreme Value Theorem tells us that we can in fact do this. This is a good thing of course. We don't want to be trying to find something that may not exist.

Next, we saw in the previous section that absolute extrema can occur at endpoints or at relative extrema. Also, from the previous section that we know that the list of critical points is also a list of all possible relative extrema. So, the endpoints along with the list of all critical points will in fact be a list of all possible absolute extrema.

Now we just need to recall that the absolute extrema are nothing more than the largest and smallest values that a function will take so all that we really need to do is get a list of possible absolute extrema, plug these points into our function and then identify the largest and smallest values.

Here is the procedure for finding absolute extrema.

Finding Absolute Extrema of $f(x)$ on $[a, b]$

0. Verify that the function is continuous on the interval $[a, b]$.
1. Find all critical points of $f(x)$ that are in the interval $[a, b]$. This makes sense if you think about it. Since we are only interested in what the function is doing in this interval we don't care about critical points that fall outside the interval.
2. Evaluate the function at the critical points found in step 1 and the end points.
3. Identify the absolute extrema.

There really isn't a whole lot to this procedure. We called the first step in the process step 0, mostly because all of the functions that we're going to look at here are going to be continuous, but it is something that we do need to be careful with. This process will only work if we have a function that is continuous on the given interval. The most labor intensive step of this process is the second step (step 1) where we find the critical points. It is also important to note that all we want are the critical points that are in the interval.

Let's do some examples.

Example 1

Determine the absolute extrema for the following function and interval.

$$g(t) = 2t^3 + 3t^2 - 12t + 4 \qquad \text{on} \qquad [-4, 2]$$

Solution

All we really need to do here is follow the procedure given above. So, first notice that this is a polynomial and so is continuous everywhere and therefore is continuous on the given interval.

Now, we need to get the derivative so that we can find the critical points of the function.

$$g'(t) = 6t^2 + 6t - 12 = 6(t+2)(t-1)$$

It looks like we'll have two critical points, $t = -2$ and $t = 1$. Note that we actually want something more than just the critical points. We only want the critical points of the function that lie in the interval in question. Both of these do fall in the interval as so we will use both of them. That may seem like a silly thing to mention at this point, but it is often forgotten, usually when it becomes important, and so we will mention it at every opportunity to make sure it's not forgotten.

Now we evaluate the function at the critical points and the end points of the interval.

$$g(-2) = 24 \qquad\qquad g(1) = -3$$
$$g(-4) = -28 \qquad\qquad g(2) = 8$$

Absolute extrema are the largest and smallest the function will ever be and these four points represent the only places in the interval where the absolute extrema can occur. So, from this list we see that the absolute maximum of $g(t)$ is 24 and it occurs at $t = -2$ (a critical point) and the absolute minimum of $g(t)$ is -28 which occurs at $t = -4$ (an endpoint).

In this example we saw that absolute extrema can and will occur at both endpoints and critical points. One of the biggest mistakes that students make with these problems is to forget to check the endpoints of the interval.

Example 2

Determine the absolute extrema for the following function and interval.

$$g(t) = 2t^3 + 3t^2 - 12t + 4 \qquad \text{on} \qquad [0, 2]$$

Solution

Note that this problem is almost identical to the first problem. The only difference is the interval that we're working on. This small change will completely change our answer however. With this change we have excluded both of the answers from the first example.

The first step is to again find the critical points. From the first example we know these are $t = -2$ and $t = 1$.. At this point it's important to recall that we only want the critical points that actually fall in the interval in question. This means that we only want $t = 1$ since $t = -2$ falls outside the interval.

Now evaluate the function at the single critical point in the interval and the two endpoints.

$$g(1) = -3 \qquad\qquad g(0) = 4 \qquad\qquad g(2) = 8$$

From this list of values we see that the absolute maximum is 8 and will occur at $t = 2$ and the absolute minimum is -3 which occurs at $t = 1$.

As we saw in this example a simple change in the interval can completely change the answer. It also has shown us that we do need to be careful to exclude critical points that aren't in the interval. Had we forgotten this and included $t = -2$ we would have gotten the wrong absolute maximum!

This is the other big mistakes that students make in these problems. All too often they forget to exclude critical points that aren't in the interval. If your instructor is anything like me this will mean that you will get the wrong answer. It's not too hard to make sure that a critical point outside of the interval is larger or smaller than any of the points in the interval.

Example 3

Suppose that the population (in thousands) of a certain kind of insect after t months is given by the following formula.

$$P(t) = 3t + \sin(4t) + 100$$

Determine the minimum and maximum population in the first 4 months.

Solution

The question that we're really asking is to find the absolute extrema of $P(t)$ on the interval $[0, 4]$. Since this function is continuous everywhere we know we can do this.

Let's start with the derivative.

$$P'(t) = 3 + 4\cos(4t)$$

We need the critical points of the function. The derivative exists everywhere so there are no critical points from that. So, all we need to do is determine where the derivative is zero.

$$\begin{aligned} 3 + 4\cos(4t) &= 0 \\ \cos(4t) &= -\frac{3}{4} \end{aligned}$$

The solutions to this are,

$$\begin{aligned} 4t &= 2.4189 + 2\pi n, \quad n = 0, \pm 1, \pm 2, \ldots \\ 4t &= 3.8643 + 2\pi n, \quad n = 0, \pm 1, \pm 2, \ldots \end{aligned} \quad \Rightarrow \quad \begin{aligned} t &= 0.6047 + \frac{\pi n}{2}, \quad n = 0, \pm 1, \pm 2, \ldots \\ t &= 0.9661 + \frac{\pi n}{2}, \quad n = 0, \pm 1, \pm 2, \ldots \end{aligned}$$

So, these are all the critical points. We need to determine the ones that fall in the interval $[0, 4]$. There's nothing to do except plug some n's into the formulas until we get all of them.

$n = 0$:

$$t = 0.6047 \qquad\qquad t = 0.9661$$

We'll need both of these critical points.

$n = 1$:

$$t = 0.6047 + \frac{\pi}{2} = 2.1755 \qquad\qquad t = 0.9661 + \frac{\pi}{2} = 2.5369$$

We'll need these.

$n = 2$:

$$t = 0.6047 + \pi = 3.7463 \qquad\qquad t = 0.9661 + \pi = 4.1077$$

In this case we only need the first one since the second is out of the interval.

There are five critical points that are in the interval. They are,

$$0.6047,\;\; 0.9661,\;\; 2.1755,\;\; 2.5369,\;\; 3.7463$$

Finally, to determine the absolute minimum and maximum population we only need to plug these values into the function as well as the two end points. Here are the function evaluations.

$$\begin{aligned} P(0) &= 100.0 & P(4) &= 111.7121 \\ P(0.6047) &= 102.4756 & P(0.9661) &= 102.2368 \\ P(2.1755) &= 107.1880 & P(2.5369) &= 106.9492 \\ P(3.7463) &= 111.9004 & & \end{aligned}$$

From these evaluations it appears that the minimum population is 100,000 (remember that P is in thousands...) which occurs at $t = 0$ and the maximum population is 111,900 which occurs at $t = 3.7463$.

Make sure that you can correctly solve trig equations. If we had forgotten the $2\pi n$ we would have missed the last three critical points in the interval and hence gotten the wrong answer since the maximum population was at the final critical point.

Also, note that we do really need to be very careful with rounding answers here. If we'd rounded to the nearest integer, for instance, it would appear that the maximum population would have occurred at two different locations instead of only one.

Example 4

Suppose that the amount of money in a bank account after t years is given by,

$$A(t) = 2000 - 10t\mathbf{e}^{5-\frac{t^2}{8}}$$

Determine the minimum and maximum amount of money in the account during the first 10 years that it is open.

Solution

Here we are really asking for the absolute extrema of $A(t)$ on the interval $[0, 10]$. As with the previous examples this function is continuous everywhere and so we know that this can be done.

We'll first need the derivative so we can find the critical points.

$$\begin{aligned} A'(t) &= -10\mathbf{e}^{5-\frac{t^2}{8}} - 10t\mathbf{e}^{5-\frac{t^2}{8}}\left(-\frac{t}{4}\right) \\ &= 10\mathbf{e}^{5-\frac{t^2}{8}}\left(-1+\frac{t^2}{4}\right) \end{aligned}$$

The derivative exists everywhere and the exponential is never zero. Therefore, the derivative will only be zero where,

$$-1+\frac{t^2}{4}=0 \qquad \Rightarrow \qquad t^2=4 \qquad \Rightarrow \qquad t=\pm 2$$

We've got two critical points, however only $t = 2$ is actually in the interval so that is only critical point that we'll use.

Let's now evaluate the function at the lone critical point and the end points of the interval. Here are those function evaluations.

$$A(0) = 2000 \qquad A(2) = 199.66 \qquad A(10) = 1999.94$$

So, the maximum amount in the account will be \$2000 which occurs at $t = 0$ and the minimum amount in the account will be \$199.66 which occurs at the 2 year mark.

In this example there are two important things to note. First, if we had included the second critical point we would have gotten an incorrect answer for the maximum amount so it's important to be careful with which critical points to include and which to exclude.

All of the problems that we've worked to this point had derivatives that existed everywhere and so the only critical points that we looked at were those for which the derivative is zero. Do not get too locked into this always happening. Most of the problems that we run into will be like this, but they won't all be like this.

Let's work another example to make this point.

Example 5

Determine the absolute extrema for the following function and interval.

$$Q(y) = 3y(y+4)^{\frac{2}{3}} \qquad \text{on} \qquad [-5, -1]$$

Solution

Again, as with all the other examples here, this function is continuous on the given interval and so we know that this can be done.

First, we'll need the derivative and make sure you can do the simplification that we did here to make the work for finding the critical points easier.

$$\begin{aligned} Q'(y) &= 3(y+4)^{\frac{2}{3}} + 3y\left(\frac{2}{3}\right)(y+4)^{-\frac{1}{3}} \\ &= 3(y+4)^{\frac{2}{3}} + \frac{2y}{(y+4)^{\frac{1}{3}}} \\ &= \frac{3(y+4)+2y}{(y+4)^{\frac{1}{3}}} \\ &= \frac{5y+12}{(y+4)^{\frac{1}{3}}} \end{aligned}$$

So, it looks like we've got two critical points.

$y = -4$ Because the derivative doesn't exist here.

$y = -\frac{12}{5}$ Because the derivative is zero here.

Both of these are in the interval so let's evaluate the function at these points and the end points of the interval.

$$Q(-4) = 0 \qquad Q\left(-\frac{12}{5}\right) = -9.849$$

$$Q(-5) = -15 \qquad Q(-1) = -6.241$$

The function has an absolute maximum of zero at $y = -4$ and the function will have an absolute minimum of -15 at $y = -5$.

So, if we had ignored or forgotten about the critical point where the derivative doesn't exist ($y = -4$) we would not have gotten the correct answer.

In this section we've seen how we can use a derivative to identify the absolute extrema of a function. This is an important application of derivatives that will arise from time to time so don't forget about it.

4.5 The Shape of a Graph, Part I

In the previous section we saw how to use the derivative to determine the absolute minimum and maximum values of a function. However, there is a lot more information about a graph that can be determined from the first derivative of a function. We will start looking at that information in this section. The main idea we'll be looking at in this section will be identifying all the relative extrema of a function.

Let's start this section off by revisiting a familiar topic from the previous chapter. Let's suppose that we have a function, $f(x)$. We know from our work in the previous chapter that the first derivative, $f'(x)$, is the rate of change of the function. We used this idea to identify where a function was increasing, decreasing or not changing.

Before reviewing this idea let's first write down the mathematical definition of increasing and decreasing. We all know what the graph of an increasing/decreasing function looks like but sometimes it is nice to have a mathematical definition as well. Here it is.

Definition

1. Given any x_1 and x_2 from an interval I with $x_1 < x_2$ if $f(x_1) < f(x_2)$ then $f(x)$ is ***increasing*** on I.
2. Given any x_1 and x_2 from an interval I with $x_1 < x_2$ if $f(x_1) > f(x_2)$ then $f(x)$ is ***decreasing*** on I.

This definition will actually be used in the proof of the next fact in this section.

Now, recall that in the previous chapter we constantly used the idea that if the derivative of a function was positive at a point then the function was increasing at that point and if the derivative was negative at a point then the function was decreasing at that point. We also used the fact that if the derivative of a function was zero at a point then the function was not changing at that point. We used these ideas to identify the intervals in which a function is increasing and decreasing.

The following fact summarizes up what we were doing in the previous chapter.

Fact

1. If $f'(x) > 0$ for every x on some interval I, then $f(x)$ is increasing on the interval.
2. If $f'(x) < 0$ for every x on some interval I, then $f(x)$ is decreasing on the interval.
3. If $f'(x) = 0$ for every x on some interval I, then $f(x)$ is constant on the interval.

The proof of this fact is in the Proofs From Derivative Applications section of the Extras appendix.

Let's take a look at an example. This example has two purposes. First, it will remind us of the increasing/decreasing type of problems that we were doing in the previous chapter. Secondly, and maybe more importantly, it will now incorporate critical points into the solution. We didn't know about critical points in the previous chapter, but if you go back and look at those examples, the first step in almost every increasing/decreasing problem is to find the critical points of the function and so the process we'll be using in the following example should be familiar.

Example 1

Determine all intervals where the following function is increasing or decreasing.

$$f(x) = -x^5 + \frac{5}{2}x^4 + \frac{40}{3}x^3 + 5$$

Solution

To determine if the function is increasing or decreasing we will need the derivative.

$$\begin{aligned} f'(x) &= -5x^4 + 10x^3 + 40x^2 \\ &= -5x^2\left(x^2 - 2x - 8\right) \\ &= -5x^2(x-4)(x+2) \end{aligned}$$

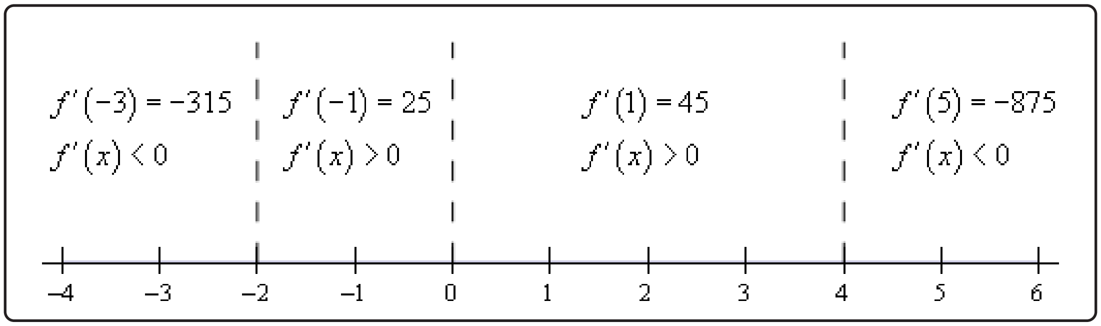
Note that when we factored the derivative we first factored a "-1" out to make the rest of the factoring a little easier.

From the factored form of the derivative we see that we have three critical points : $x = -2$, $x = 0$, and $x = 4$. We'll need these in a bit.

We now need to determine where the derivative is positive and where it's negative. We've done this several times now in both the Review chapter and the previous chapter. Since the derivative is a polynomial it is continuous and so we know that the only way for it to change signs is to first go through zero.

In other words, the only place that the derivative *may* change signs is at the critical points of the function. We've now got another use for critical points. So, we'll build a number line, graph the critical points and pick test points from each region to see if the derivative is positive or negative in each region.

Here is the number line and the test points for the derivative.



Make sure that you test your points in the derivative. One of the more common mistakes here is to test the points in the function instead! Recall that we know that the derivative will be the same sign in each region. The only place that the derivative can change signs is at the critical points and we've marked the only critical points on the number line.

So, it looks like we've got the following intervals of increase and decrease.

$$\text{Increase : } -2 < x < 0 \text{ and } 0 < x < 4$$
$$\text{Decrease : } -\infty < x < -2 \text{ and } 4 < x < \infty$$

In this example we used the fact that the only place that a derivative can change sign is at the critical points. Also, the critical points for this function were those for which the derivative was zero. However, the same thing can be said for critical points where the derivative doesn't exist. This is nice to know. A function, in this section a derivative, can change signs where it is zero or doesn't exist. In the previous chapter all our examples of this type had only critical points where the derivative was zero. Now, that we know more about critical points we'll also see an example or two later on with critical points where the derivative doesn't exist.

If you aren't sure you believe that functions (they don't have to be derivatives of course) can change sign where they don't exist consider $f(x) = \frac{1}{x}$. This function clearly does not exist at $x = 0$ and is negative if $x < 0$ and positive if $x > 0$ and so does change sign at the point where it does not exist. Be careful to not assume this will always be true however. Take $f(x) = \frac{1}{x^2}$ for example. Again, this clearly does not exist at $x = 0$ and yet is positive on both sides of $x = 0$.

So, just to reiterate one more time. Functions, regardless of whether they are derivatives or not, may (but are not guaranteed to) change sign where they are either zero or do not exist.

Now that we have the previous "reminder" example out of the way let's move into some new material. Once we have the intervals of increasing and decreasing for a function we can use this information to get a sketch of the graph. Note that the sketch, at this point, may not be super accurate when it comes to the curvature of the graph, but it will at least have the basic shape correct. To get the curvature of the graph correct we'll need the information from the next section.

Let's attempt to get a sketch of the graph of the function we used in the previous example.

Example 2

Sketch the graph of the following function.

$$f(x) = -x^5 + \frac{5}{2}x^4 + \frac{40}{3}x^3 + 5$$

Solution

There really isn't a whole lot to this example. Whenever we sketch a graph it's nice to have a few points on the graph to give us a starting place. So we'll start by the function at the critical points. These will give us some starting points when we go to sketch the graph. These points are,

$$f(-2) = -\frac{89}{3} = -29.67 \qquad f(0) = 5 \qquad f(4) = \frac{1423}{3} = 474.33$$

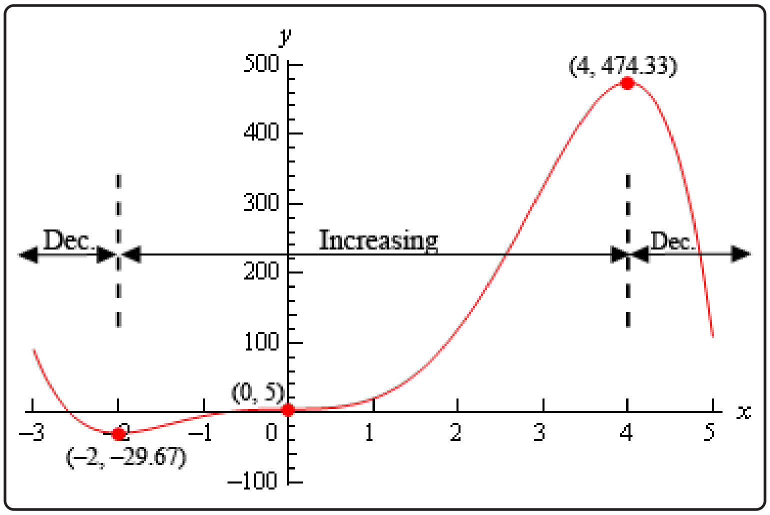
Once these points are graphed we go to the increasing and decreasing information and start sketching. For reference purposes here is the increasing/decreasing information.

$$\text{Increase} : -2 < x < 0 \text{ and } 0 < x < 4$$
$$\text{Decrease} : -\infty < x < -2 \text{ and } 4 < x < \infty$$

Note that we are only after a sketch of the graph. As noted before we started this example we won't be able to accurately predict the curvature of the graph at this point. However, even without this information we will still be able to get a basic idea of what the graph should look like.

To get this sketch we start at the very left of the graph and knowing that the graph must be decreasing and will continue to decrease until we get to $x = -2$. At this point the function will continue to increase until it gets to $x = 4$. However, note that during the increasing phase it does need to go through the point at $x = 0$ and at this point we also know that the derivative is zero here and so the graph goes through $x = 0$ horizontally. Finally, once we hit $x = 4$ the graph starts, and continues, to decrease. Also, note that just like at $x = 0$ the graph will need to be horizontal when it goes through the other two critical points as well.

Here is the graph of the function. We, of course, used a graphical program to generate this graph, however, outside of some potential curvature issues if you followed the increasing/decreasing information and had all the critical points plotted first you should have something similar to this.



Let's use the sketch from this example to give us a very nice test for classifying critical points as relative maximums, relative minimums or neither minimums or maximums.

Recall from the Minimum and Maximum Values section that all relative extrema of a function come from the list of critical points. The graph in the previous example has two relative extrema and both occur at critical points as the we predicted in that section. Note as well that we've got a critical point that isn't a relative extrema ($x = 0$). This is okay since there is no reason to think that all critical points will be relative extrema. We only know that relative extrema will come from the list of critical points.

In the sketch of the graph from the previous example we can see that to the left of $x = -2$ the graph is decreasing and to the right of $x = -2$ the graph is increasing and $x = -2$ is a relative minimum. In other words, the graph is behaving around the minimum exactly as it would have to be in order for $x = -2$ to be a minimum. The same thing can be said for the relative maximum at $x = 4$. The graph is increasing on the left and decreasing on the right exactly as it must be in order for $x = 4$ to be a maximum. Finally, the graph is increasing on both sides of $x = 0$ and so this critical point can't be a minimum or a maximum.

These ideas can be generalized to arrive at a nice way to test if a critical point is a relative minimum, relative maximum or neither. If $x = c$ is a critical point and the function is decreasing to the left of $x = c$ and is increasing to the right then $x = c$ must be a relative minimum of the function. Likewise, if the function is increasing to the left of $x = c$ and decreasing to the right then $x = c$ must be a relative maximum of the function. Finally, if the function is increasing on both sides of $x = c$ or decreasing on both sides of $x = c$ then $x = c$ can be neither a relative minimum nor a relative maximum.

These ideas can be summarized up in the following test.

First Derivative Test

Suppose that $x = c$ is a critical point of $f(x)$ then,

1. If $f'(x) > 0$ to the left of $x = c$ and $f'(x) < 0$ to the right of $x = c$ then $x = c$ is a relative maximum.
2. If $f'(x) < 0$ to the left of $x = c$ and $f'(x) > 0$ to the right of $x = c$ then $x = c$ is a relative minimum.
3. If $f'(x)$ is the same sign on both sides of $x = c$ then $x = c$ is neither a relative maximum nor a relative minimum.

It is important to note here that the first derivative test will only classify critical points as relative extrema and not as absolute extrema. As we recall from the Finding Absolute Extrema section absolute extrema are largest and smallest function values and may not even exist or be critical points if they do exist.

The first derivative test is exactly that, a test using the first derivative. It doesn't ever use the value of the function and so no conclusions can be drawn from the test about the relative "size" of the function at the critical points (which would be needed to identify absolute extrema) and can't even begin to address the fact that absolute extrema may not occur at critical points.

Let's take at another example.

Example 3

Find and classify all the critical points of the following function. Give the intervals where the function is increasing and decreasing.

$$g(t) = t\sqrt[3]{t^2 - 4}$$

Solution

First, we'll need the derivative so we can get our hands on the critical points. Note as well that we'll do some simplification on the derivative to help us find the critical points.

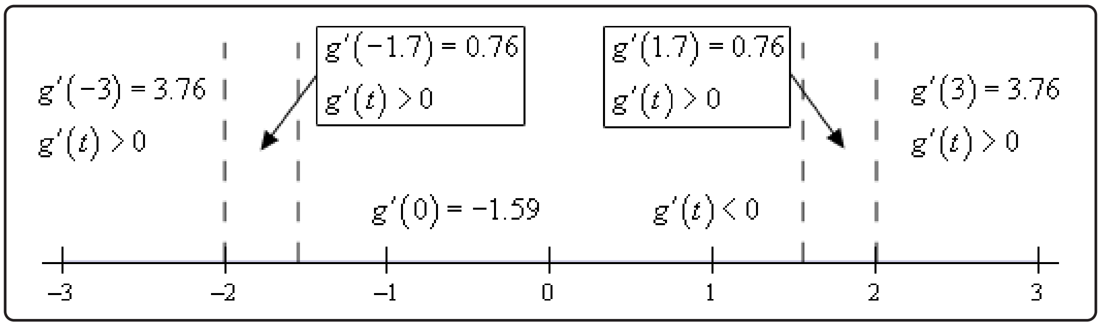
$$\begin{aligned} g'(t) &= \left(t^2 - 4\right)^{\frac{1}{3}} + \frac{2}{3}t^2\left(t^2 - 4\right)^{-\frac{2}{3}} \\ &= \left(t^2 - 4\right)^{\frac{1}{3}} + \frac{2t^2}{3(t^2 - 4)^{\frac{2}{3}}} \\ &= \frac{3\left(t^2 - 4\right) + 2t^2}{3(t^2 - 4)^{\frac{2}{3}}} \\ &= \frac{5t^2 - 12}{3(t^2 - 4)^{\frac{2}{3}}} \end{aligned}$$

So, it looks like we'll have four critical points here. They are,

$$t = \pm 2 \qquad \text{The derivative doesn't exist here.}$$

$$t = \pm\sqrt{\frac{12}{5}} = \pm 1.549 \qquad \text{The derivative is zero here.}$$

Finding the intervals of increasing and decreasing will also give the classification of the critical points so let's get those first. Here is a number line with the critical points graphed and test points.



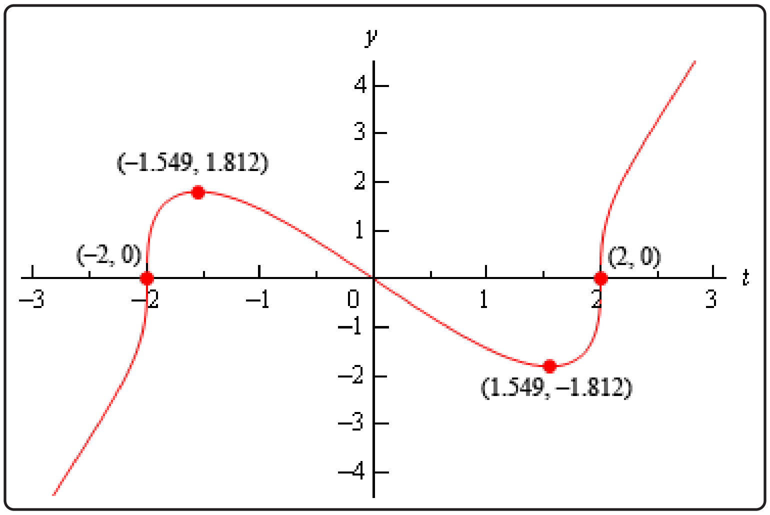
So, it looks like we've got the following intervals of increasing and decreasing.

$$\text{Increase : } -\infty < x < -2, \quad -2 < x < -\sqrt{\frac{12}{5}}, \quad \sqrt{\frac{12}{5}} < x < 2, \quad \& \quad 2 < x < \infty$$

$$\text{Decrease : } \quad -\sqrt{\frac{12}{5}} < x < \sqrt{\frac{12}{5}}$$

From this it looks like $t = -2$ and $t = 2$ are neither relative minimum or relative maximums since the function is increasing on both side of them. On the other hand, $t = -\sqrt{\frac{12}{5}}$ is a relative maximum and $t = \sqrt{\frac{12}{5}}$ is a relative minimum.

For completeness sake here is the graph of the function. Note that this graph is a little trickier to sketch based only on the increasing and decreasing information. It is only presented here for reference so you can see what it looks like.



In the previous example the two critical points where the derivative didn't exist ended up not being relative extrema. Do not read anything into this. They often will be relative extrema. Check out Example 5 in the Absolute Extrema to see an example of one such critical point.

Let's work a couple more examples.

Example 4

Suppose that the elevation above sea level of a road is given by the following function.

$$E(x) = 500 + \cos\left(\frac{x}{4}\right) + \sqrt{3}\sin\left(\frac{x}{4}\right)$$

where x is in miles. Assume that if x is positive we are to the east of the initial point of measurement and if x is negative we are to the west of the initial point of measurement.

If we start 25 miles to the west of the initial point of measurement and drive until we are 25 miles east of the initial point how many miles of our drive were we driving up an incline?

Solution

Okay, this is just a really fancy way of asking what the intervals of increasing and decreasing are for the function on the interval $[-25, 25]$. So, we first need the derivative of the function.

$$E'(x) = -\frac{1}{4}\sin\left(\frac{x}{4}\right) + \frac{\sqrt{3}}{4}\cos\left(\frac{x}{4}\right)$$

Setting this equal to zero gives,

$$-\frac{1}{4}\sin\left(\frac{x}{4}\right)+\frac{\sqrt{3}}{4}\cos\left(\frac{x}{4}\right)=0$$
$$\tan\left(\frac{x}{4}\right)=\sqrt{3}$$

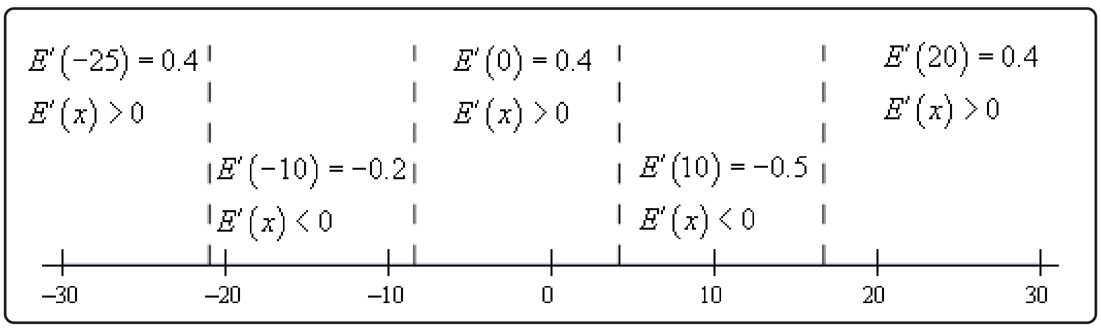
The solutions to this and hence the critical points are,

$$\begin{aligned}\frac{x}{4}&=1.0472+2\pi n,\ n=0,\pm1,\pm2,\ldots\\ \frac{x}{4}&=4.1888+2\pi n,\ n=0,\pm1,\pm2,\ldots\end{aligned}\quad\Rightarrow\quad\begin{aligned}x&=4.1888+8\pi n,\ n=0,\pm1,\pm2,\ldots\\ x&=16.7552+8\pi n,\ n=0,\pm1,\pm2,\ldots\end{aligned}$$

I'll leave it to you to check that the critical points that fall in the interval that we're after are,

$$-20.9439,\ \ -8.3775,\ \ 4.1888,\ \ 16.7552$$

Here is the number line with the critical points and test points.



So, it looks like the intervals of increasing and decreasing are,

Increase : $-25<x<-20.9439,\ \ -8.3775<x<4.1888$ and $16.7552<x<25$

Decrease : $-20.9439<x<-8.3775$ and $4.1888<x<16.7552$

Notice that we had to end our intervals at -25 and 25 since we've done no work outside of these points and so we can't really say anything about the function outside of the interval $[-25,25]$.

From the intervals we can actually answer the question. We were driving on an incline during the intervals of increasing and so the total number of miles is,

$$\begin{aligned}\text{Distance}&=(-20.9439-(-25))+(4.1888-(-8.3775))+(25-16.7552)\\&=24.8652\text{ miles}\end{aligned}$$

Even though the problem didn't ask for it we can also classify the critical points that are in the interval $[-25, 25]$.

$$\text{Relative Maximums : } -20.9439, 4.1888$$
$$\text{Relative Minimums : } -8.3775, 16.7552$$

Example 5

The population of rabbits (in hundreds) after t years in a certain area is given by the following function,

$$P(t) = t^2 \ln(3t) + 6$$

Determine if the population ever decreases in the first two years.

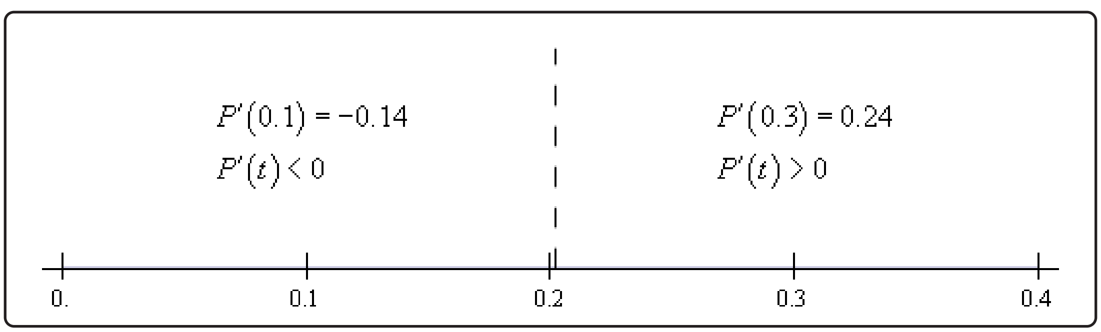
Solution

So, again we are really after the intervals and increasing and decreasing in the interval $[0, 2]$.

We found the only critical point to this function back in the Critical Points section to be,

$$x = \frac{1}{3\sqrt{\mathbf{e}}} = 0.202$$

Here is a number line for the intervals of increasing and decreasing.



So, it looks like the population will decrease for a short period and then continue to increase forever.

Also, while the problem didn't ask for it we can see that the single critical point is a relative minimum.

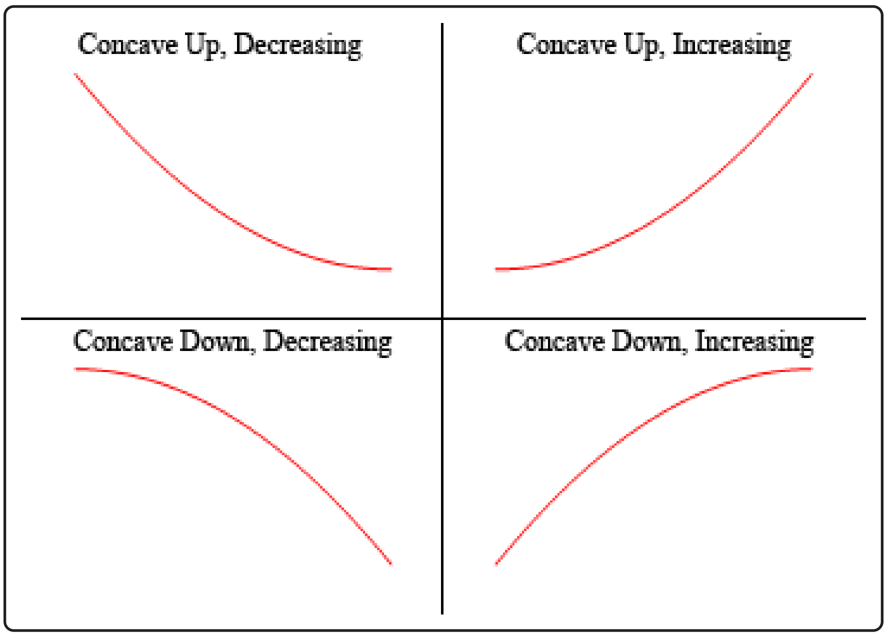
In this section we've seen how we can use the first derivative of a function to give us some information about the shape of a graph and how we can use this information in some applications.

Using the first derivative to give us information about a whether a function is increasing or decreasing is a very important application of derivatives and arises on a fairly regular basis in many areas.

4.6 The Shape of a Graph, Part II

In the previous section we saw how we could use the first derivative of a function to get some information about the graph of a function. In this section we are going to look at the information that the second derivative of a function can give us a about the graph of a function.

Before we do this we will need a couple of definitions out of the way. The main concept that we'll be discussing in this section is concavity. Concavity is easiest to see with a graph (we'll give the mathematical definition in a bit).



So, a function is **concave up** if it "opens" up and the function is **concave down** if it "opens" down. Notice as well that concavity has nothing to do with increasing or decreasing. A function can be concave up and either increasing or decreasing. Similarly, a function can be concave down and either increasing or decreasing.

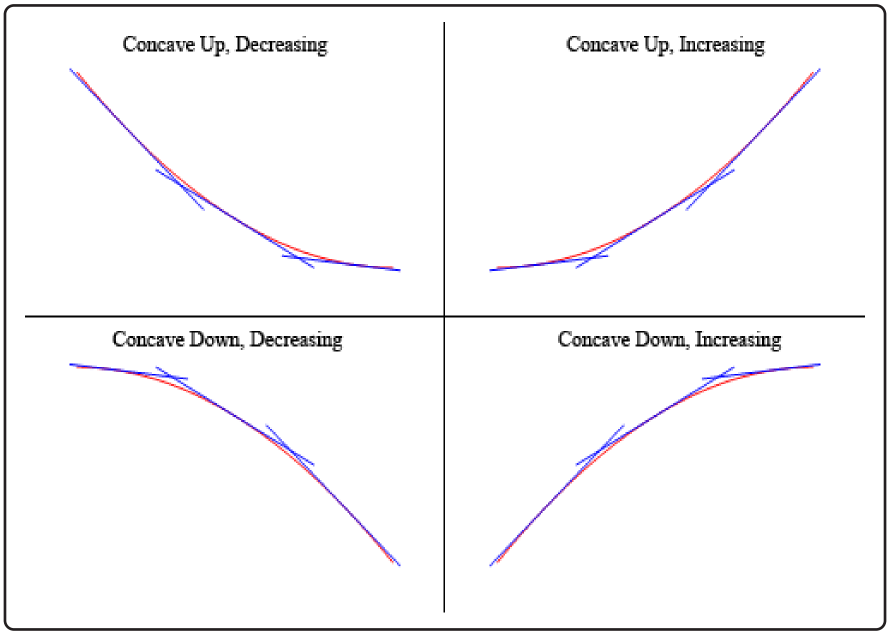
It's probably not the best way to define concavity by saying which way it "opens" since this is a somewhat nebulous definition. Here is the mathematical definition of concavity.

Definition 1

Given the function $f(x)$ then

1. $f(x)$ is **concave up** on an interval I if all of the tangents to the curve on I are below the graph of $f(x)$.
2. $f(x)$ is **concave down** on an interval I if all of the tangents to the curve on I are above the graph of $f(x)$.

To show that the graphs above do in fact have concavity claimed above here is the graph again (blown up a little to make things clearer).



So, as you can see, in the two upper graphs all of the tangent lines sketched in are all below the graph of the function and these are concave up. In the lower two graphs all the tangent lines are above the graph of the function and these are concave down.

Again, notice that concavity and the increasing/decreasing aspect of the function is completely separate and do not have anything to do with each other. This is important to note because students often mix these two up and use information about one to get information about the other.

There's one more definition that we need to get out of the way.

Definition 2

A point $x = c$ is called an **inflection point** if the function is continuous at the point and the concavity of the graph changes at that point.

Now that we have all the concavity definitions out of the way we need to bring the second derivative into the mix. We did after all start off this section saying we were going to be using the second derivative to get information about the graph. The following fact relates the second derivative of a function to its concavity. The proof of this fact is in the Proofs From Derivative Applications section of the Extras appendix.

Fact

Given the function $f(x)$ then,

1. If $f''(x) > 0$ for all x in some interval I then $f(x)$ is concave up on I.
2. If $f''(x) < 0$ for all x in some interval I then $f(x)$ is concave down on I.

So, what this fact tells us is that the inflection points will be all the points where the second derivative changes sign. We saw in the previous chapter that a function may change signs if it is either zero or does not exist. Note that we were working with the first derivative in the previous section but the fact that a function possibly changing signs where it is zero or doesn't exist has nothing to do with the first derivative. It is simply a fact that applies to all functions regardless of whether they are derivatives or not.

This in turn tells us that a list of possible inflection points will be those points where the second derivative is zero or doesn't exist, as these are the only points where the second derivative might change sign.

Be careful however to not make the assumption that just because the second derivative is zero or doesn't exist that the point will be an inflection point. We will only know that it is an inflection point once we determine the concavity on both sides of it. It will only be an inflection point if the concavity is different on both sides of the point.

Now that we know about concavity we can use this information as well as the increasing/decreasing information from the previous section to get a pretty good idea of what a graph should look like. Let's take a look at an example of that.

Example 1

For the following function identify the intervals where the function is increasing and decreasing and the intervals where the function is concave up and concave down. Use this information to sketch the graph.

$$h(x) = 3x^5 - 5x^3 + 3$$

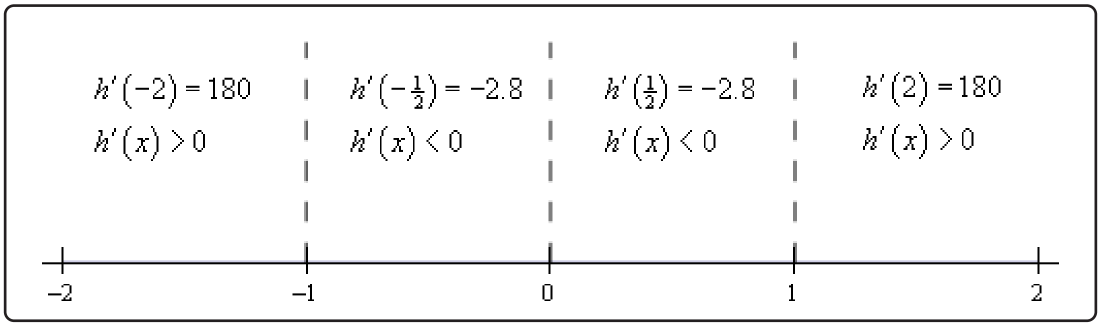
Solution

Okay, we are going to need the first two derivatives so let's get those first.

$$h'(x) = 15x^4 - 15x^2 = 15x^2(x-1)(x+1)$$
$$h''(x) = 60x^3 - 30x = 30x\left(2x^2 - 1\right)$$

Let's start with the increasing/decreasing information since we should be fairly comfortable with that after the last section.

There are three critical points for this function : $x = -1$, $x = 0$, and $x = 1$. Below is the number line for the increasing/decreasing information.



So, it looks like we've got the following intervals of increasing and decreasing.

$$\text{Increasing : } -\infty < x < -1 \text{ and } 1 < x < \infty$$
$$\text{Decreasing : } -1 < x < 0, \;\; 0 < x < 1$$

Note that from the first derivative test we can also say that $x = -1$ is a relative maximum and that $x = 1$ is a relative minimum. Also $x = 0$ is neither a relative minimum or maximum.

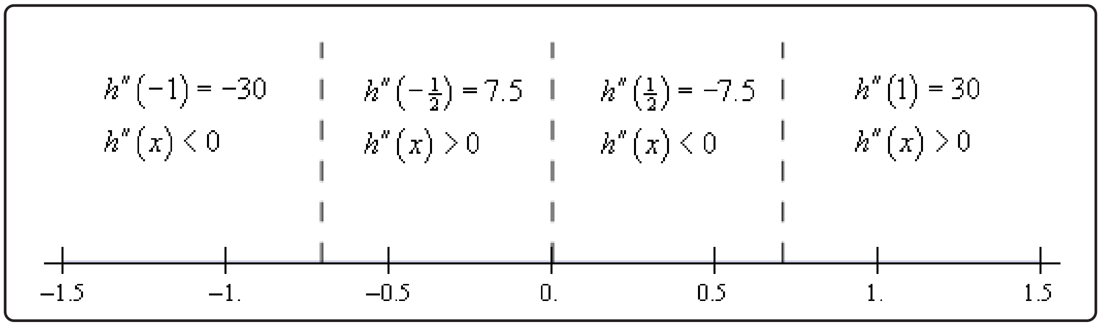
Now let's get the intervals where the function is concave up and concave down. If you think about it this process is almost identical to the process we use to identify the intervals of increasing and decreasing. This only difference is that we will be using the second derivative instead of the first derivative.

The first thing that we need to do is identify the possible inflection points. These will be where the second derivative is zero or doesn't exist. The second derivative in this case is a polynomial and so will exist everywhere. It will be zero at the following points.

$$x = 0, \;\; x = \pm\frac{1}{\sqrt{2}} = \pm 0.7071$$

As with the increasing and decreasing part we can draw a number line up and use these points to divide the number line into regions. In these regions we know that the second derivative will always have the same sign since these three points are the only places where the function *may* change sign. Therefore, all that we need to do is pick a point from each region and plug it into the second derivative. The second derivative will then have that sign in the whole region from which the point came from

Here is the number line for this second derivative.



So, it looks like we've got the following intervals of concavity.

$$\text{Concave Up : } -\frac{1}{\sqrt{2}} < x < 0 \text{ and } \frac{1}{\sqrt{2}} < x < \infty$$
$$\text{Concave Down : } -\infty < x < -\frac{1}{\sqrt{2}} \text{ and } 0 < x < \frac{1}{\sqrt{2}}$$

This also means that

$$x = 0, \; x = \pm\frac{1}{\sqrt{2}} = \pm 0.7071$$

are all inflection points.

All this information can be a little overwhelming when going to sketch the graph. The first thing that we should do is get some starting points. The critical points and inflection points are good starting points. So, first graph these points.

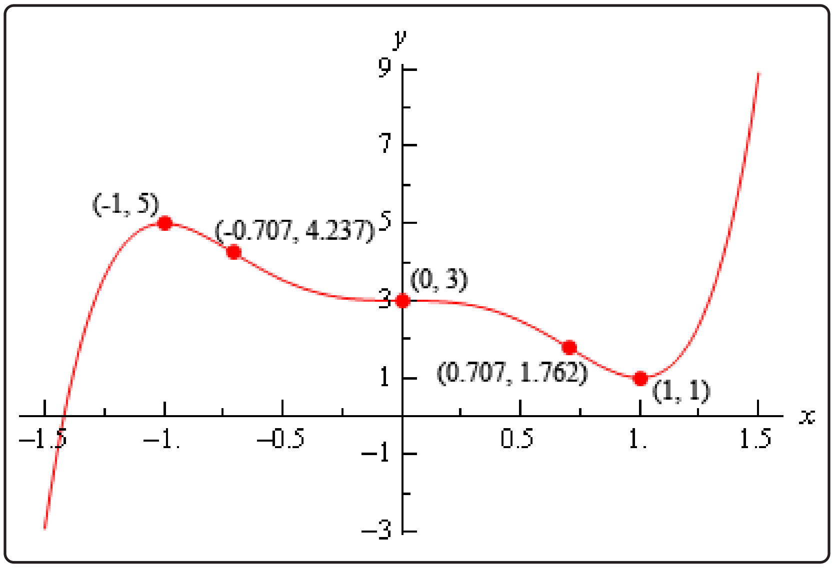
From this point there are several ways to proceed with sketching the graph. The way that we find to be the easiest (although you may not and that is perfectly fine....) is to start with the increasing/decreasing information and start sketching the graph from just that information as we did in the previous section. However, unlike the previous section, this time as we draw an increasing or decreasing portion of the curve we will also pay attention to the concavity of the curve as we are doing this.

So, if we start with $x < -1$ we know that we have an increasing function. At the same time, we know that we also have to be concave down in this range. So, we can start off with sketching an increasing curve that has is also concave down until we reach $x = -1$.

At this point the graph starts to decrease and will continue to decrease until we hit $x = 1$. However, as we decrease the concavity needs to switch to concave up at $x \approx -0.707$ and then switch back to concave down at $x = 0$ with a final switch to concave up at $x \approx 0.707$.

Once we hit $x = 1$ the graph starts to increase and is still concave up and both of these behaviors continue for the rest of the graph.

Putting all this information together will give us the following graph of the function.



We can use the previous example to illustrate another way to classify some of the critical points of a function as relative maximums or relative minimums.

Notice that $x = -1$ is a relative maximum and that the function is concave down at this point. This means that $f''(-1)$ must be negative. Likewise, $x = 1$ is a relative minimum and the function is concave up at this point. This means that $f''(1)$ must be positive.

As we'll see in a bit we will need to be very careful with $x = 0$. In this case the second derivative is zero, but that will not actually mean that $x = 0$ is not a relative minimum or maximum. We'll see some examples of this in a bit, but we need to get some other information taken care of first.

It is also important to note here that all of the critical points in this example were critical points in which the first derivative was zero and this is required for this to work. We will not be able to use this test on critical points where the derivative doesn't exist.

Here is the test that can be used to classify some of the critical points of a function. The proof of this test is in the Proofs of Derivative Applications section of the Extras appendix.

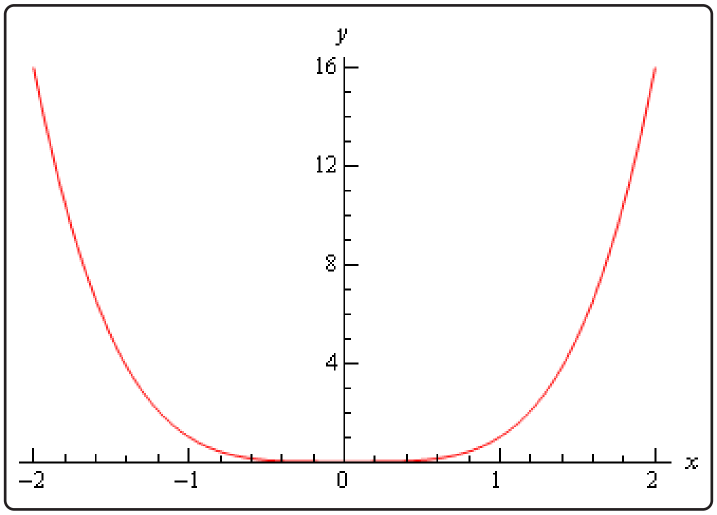
Second Derivative Test

Suppose that $x = c$ is a critical point of $f(x)$ such that $f'(c) = 0$ and that $f''(x)$ is continuous in a region around $x = c$. Then,

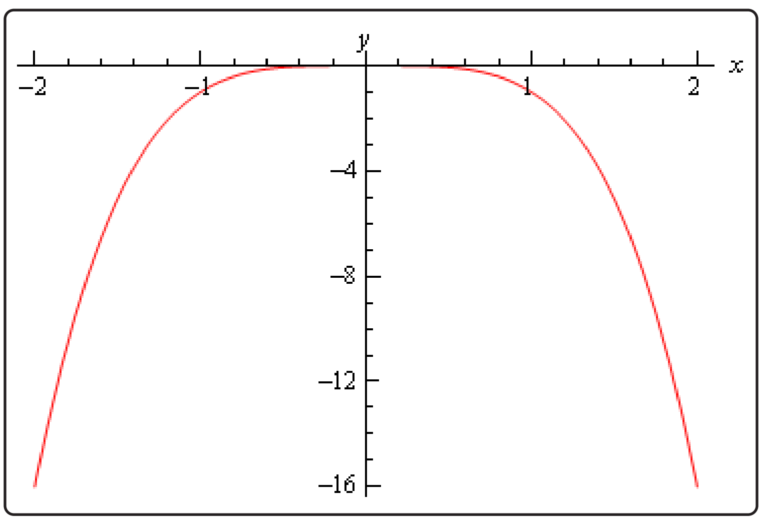
1. If $f''(c) < 0$ then $x = c$ is a relative maximum.
2. If $f''(c) > 0$ then $x = c$ is a relative minimum.
3. If $f''(c) = 0$ then $x = c$ can be a relative maximum, relative minimum or neither.

The third part of the second derivative test is important to notice. If the second derivative is zero then the critical point can be anything. Below are the graphs of three functions all of which have a critical point at $x = 0$, the second derivative of all of the functions is zero at $x = 0$ and yet all three possibilities are exhibited.

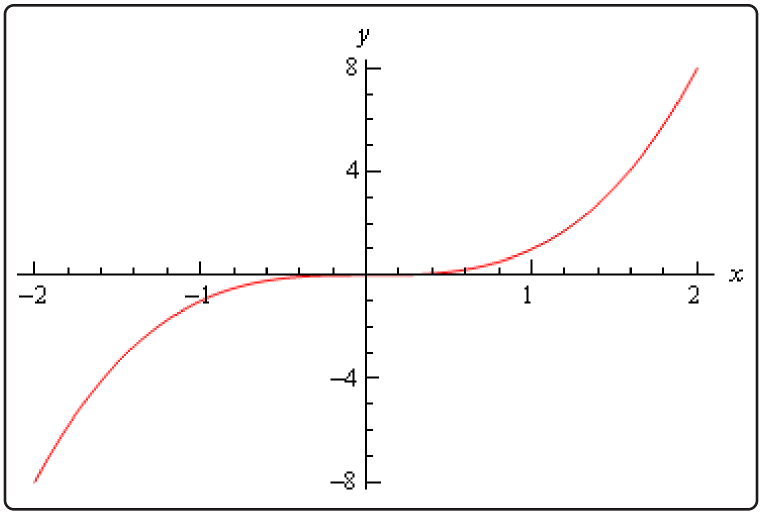
The first is the graph of $f(x) = x^4$. This graph has a relative minimum at $x = 0$.



Next is the graph of $f(x) = -x^4$ which has a relative maximum at $x = 0$.



Finally, there is the graph of $f(x) = x^3$ and this graph had neither a relative minimum or a relative maximum at $x = 0$.



So, we can see that we have to be careful if we fall into the third case. For those times when we do fall into this case we will have to resort to other methods of classifying the critical point. This is usually done with the first derivative test.

Let's go back and take a look at the critical points from the first example and use the Second Derivative Test on them, if possible.

Example 2

Use the second derivative test to classify the critical points of the function,

$$h(x) = 3x^5 - 5x^3 + 3$$

Solution

Note that all we're doing here is verifying the results from the first example. The second derivative is,

$$h''(x) = 60x^3 - 30x$$

The three critical points ($x = -1$, $x = 0$, and $x = 1$) of this function are all critical points where the first derivative is zero so we know that we at least have a chance that the Second Derivative Test will work. The value of the second derivative for each of these are,

$$h''(-1) = -30 \qquad h''(0) = 0 \qquad h''(1) = 30$$

The second derivative at $x = -1$ is negative so by the Second Derivative Test this critical point this is a relative maximum as we saw in the first example. The second derivative at $x = 1$ is positive and so we have a relative minimum here by the Second Derivative Test as we also saw in the first example.

In the case of $x = 0$ the second derivative is zero and so we can't use the Second Derivative Test to classify this critical point. Note however, that we do know from the First Derivative Test we used in the first example that *in this case* the critical point is not a relative extrema.

Let's work one more example.

Example 3

For the following function find the inflection points and use the second derivative test, if possible, to classify the critical points. Also, determine the intervals of increase/decrease and the intervals of concave up/concave down and sketch the graph of the function.

$$f(t) = t(6-t)^{\frac{2}{3}}$$

Solution

We'll need the first and second derivatives to get us started. We'll leave it to you to verify these derivatives but be aware that we did a little simplification after taking each derivative.

$$f'(t) = \frac{18-5t}{3(6-t)^{\frac{1}{3}}} \qquad f''(t) = \frac{10t-72}{9(6-t)^{\frac{4}{3}}}$$

The critical points are,

$$t = \frac{18}{5} = 3.6 \qquad t = 6$$

Notice as well that we won't be able to use the second derivative test on $t = 6$ to classify this critical point since the derivative doesn't exist at this point. To classify this, we'll need the increasing/decreasing information that we'll get to sketch the graph.

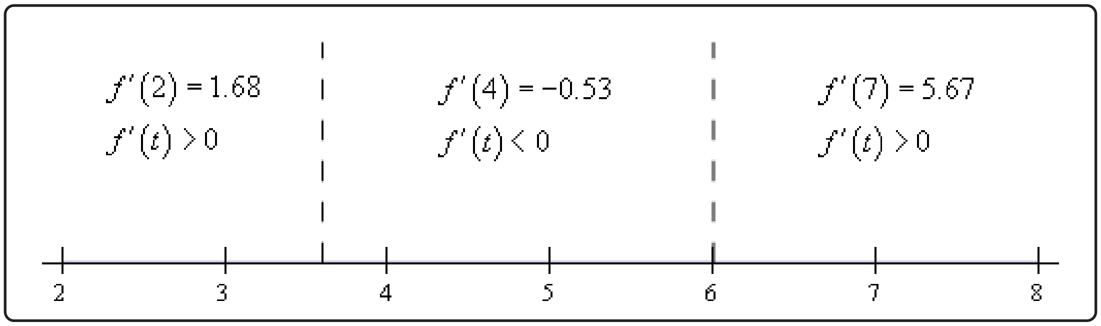
We can however, use the Second Derivative Test to classify the other critical point so let's do that before we proceed with the sketching work. Here is the value of the second derivative at $t = 3.6$.

$$f''(3.6) = -1.245 < 0$$

So, according to the second derivative test $t = 3.6$ is a relative maximum.

Now let's proceed with the work to get the sketch of the graph and notice that once we have the increasing/decreasing information we'll be able to classify $t = 6$.

Here is the number line for the first derivative.



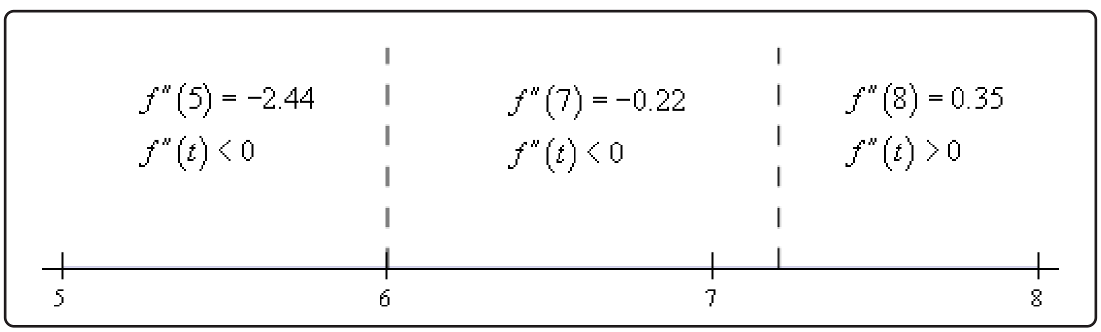
So, according to the first derivative test we can verify that $t = 3.6$ is in fact a relative maximum. We can also see that $t = 6$ is a relative minimum.

Be careful not to assume that a critical point that can't be used in the second derivative test won't be a relative extrema. We've clearly seen now both with this example and in the discussion after we have the test that just because we can't use the Second Derivative Test or the Second Derivative Test doesn't tell us anything about a critical point doesn't mean that the critical point will not be a relative extrema. This is a common mistake that many students make so be careful when using the Second Derivative Test.

Okay, let's finish the problem out. We will need the list of possible inflection points. These are,

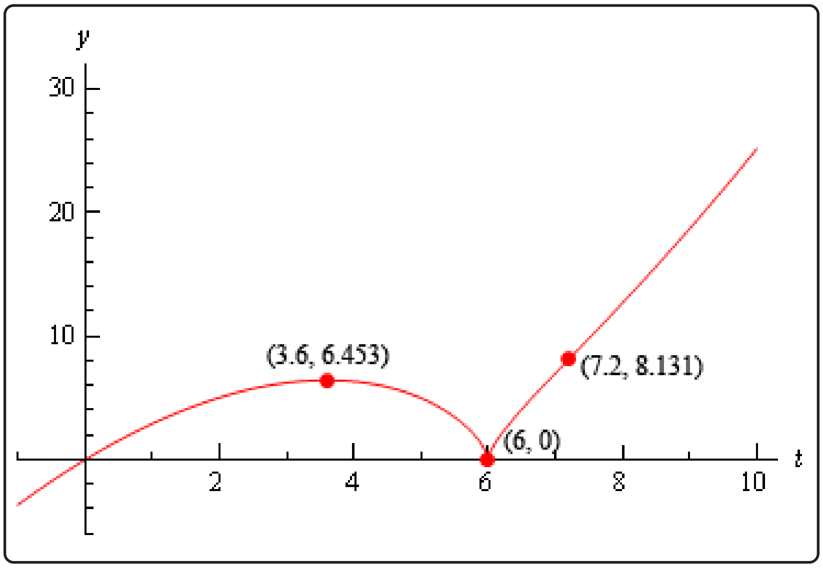
$$t = 6 \qquad\qquad t = \frac{72}{10} = 7.2$$

Here is the number line for the second derivative. Note that we will need this to see if the two points above are in fact inflection points.



So, the concavity only changes at $t = 7.2$ and so this is the only inflection point for this function.

Here is the sketch of the graph.



The change of concavity at $t = 7.2$ is hard to see, but it is there it's just a very subtle change in concavity.

4.7 The Mean Value Theorem

In this section we want to take a look at the Mean Value Theorem. In most traditional textbooks this section comes before the sections containing the First and Second Derivative Tests because many of the proofs in those sections need the Mean Value Theorem. However, we feel that from a logical point of view it's better to put the Shape of a Graph sections right after the absolute extrema section. So, if you've been following the proofs from the previous two sections you've probably already read through this section.

Before we get to the Mean Value Theorem we need to cover the following theorem.

Rolle's Theorem

Suppose $f(x)$ is a function that satisfies all of the following.

1. $f(x)$ is continuous on the closed interval $[a, b]$.
2. $f(x)$ is differentiable on the open interval (a, b).
3. $f(a) = f(b)$

Then there is a number c such that $a < c < b$ and $f'(c) = 0$. Or, in other words $f(x)$ has a critical point in (a, b).

To see the proof of Rolle's Theorem see the Proofs From Derivative Applications section of the Extras appendix.

Let's take a look at a quick example that uses Rolle's Theorem.

Example 1

Show that $f(x) = 4x^5 + x^3 + 7x - 2$ has exactly one real root.

Solution

From basic Algebra principles we know that since $f(x)$ is a 5^{th} degree polynomial it will have five roots. What we're being asked to prove here is that only one of those 5 is a real number and the other 4 must be complex roots.

First, we should show that it does have at least one real root. To do this note that $f(0) = -2$ and that $f(1) = 10$ and so we can see that $f(0) < 0 < f(1)$. Now, because $f(x)$ is a polynomial we know that it is continuous everywhere and so by the Intermediate Value Theorem there is a number c such that $0 < c < 1$ and $f(c) = 0$. In other words $f(x)$ has at least one real root.

We now need to show that this is in fact the only real root. To do this we'll use an argument that is called contradiction proof. What we'll do is assume that $f(x)$ has at least two real roots. This means that we can find real numbers a and b (there might be more, but all we need for this particular argument is two) such that $f(a) = f(b) = 0$. But if we do this then we know from Rolle's Theorem that there must then be another number c such that $f'(c) = 0$.

This is a problem however. The derivative of this function is,

$$f'(x) = 20x^4 + 3x^2 + 7$$

Because the exponents on the first two terms are even we know that the first two terms will always be greater than or equal to zero and we are then going to add a positive number onto that and so we can see that the smallest the derivative will ever be is 7 and this contradicts the statement above that says we MUST have a number c such that $f'(c) = 0$.

We reached these contradictory statements by assuming that $f(x)$ has at least two roots. Since this assumption leads to a contradiction the assumption must be false and so we can only have a single real root.

The reason for covering Rolle's Theorem is that it is needed in the proof of the Mean Value Theorem. To see the proof see the Proofs From Derivative Applications section of the Extras appendix. Here is the theorem.

Mean Value Theorem

Suppose $f(x)$ is a function that satisfies both of the following.

1. $f(x)$ is continuous on the closed interval $[a, b]$.
2. $f(x)$ is differentiable on the open interval (a, b).

Then there is a number c such that $a < c < b$ and

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Or,

$$f(b) - f(a) = f'(c)(b - a)$$

Note that the Mean Value Theorem doesn't tell us what c is. It only tells us that there is at least one number c that will satisfy the conclusion of the theorem.

Also note that if it weren't for the fact that we needed Rolle's Theorem to prove this we could think of Rolle's Theorem as a special case of the Mean Value Theorem. To see that just assume that $f(a) = f(b)$ and then the result of the Mean Value Theorem gives the result of Rolle's Theorem.

Before we take a look at a couple of examples let's think about a geometric interpretation of the Mean Value Theorem. First define $A = (a, f(a))$ and $B = (b, f(b))$ and then we know from the Mean Value theorem that there is a c such that $a < c < b$ and that

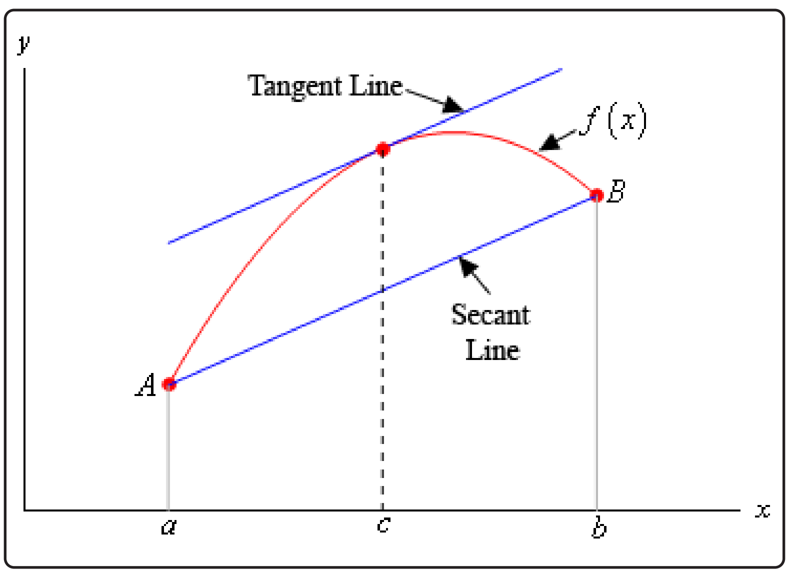
$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Now, if we draw in the secant line connecting A and B then we can know that the slope of the secant line is,

$$\frac{f(b) - f(a)}{b - a}$$

Likewise, if we draw in the tangent line to $f(x)$ at $x = c$ we know that its slope is $f'(c)$.

What the Mean Value Theorem tells us is that these two slopes must be equal or in other words the secant line connecting A and B and the tangent line at $x = c$ must be parallel. We can see this in the following sketch.



Let's now take a look at a couple of examples using the Mean Value Theorem.

Example 2

Determine all the numbers c which satisfy the conclusions of the Mean Value Theorem for the following function.

$$f(x) = x^3 + 2x^2 - x \qquad \text{on} \qquad [-1, 2]$$

Solution

There isn't really a whole lot to this problem other than to notice that since $f(x)$ is a polynomial it is both continuous and differentiable (*i.e.* the derivative exists) on the interval given.

First let's find the derivative.

$$f'(x) = 3x^2 + 4x - 1$$

Now, to find the numbers that satisfy the conclusions of the Mean Value Theorem all we need to do is plug this into the formula given by the Mean Value Theorem.

$$\begin{aligned} f'(c) &= \frac{f(2) - f(-1)}{2 - (-1)} \\ 3c^2 + 4c - 1 &= \frac{14 - 2}{3} = \frac{12}{3} = 4 \end{aligned}$$

Now, this is just a quadratic equation,

$$\begin{aligned} 3c^2 + 4c - 1 &= 4 \\ 3c^2 + 4c - 5 &= 0 \end{aligned}$$

Using the quadratic formula on this we get,

$$c = \frac{-4 \pm \sqrt{16 - 4(3)(-5)}}{6} = \frac{-4 \pm \sqrt{76}}{6}$$

So, solving gives two values of c.

$$c = \frac{-4 + \sqrt{76}}{6} = 0.7863 \qquad\qquad c = \frac{-4 - \sqrt{76}}{6} = -2.1196$$

Notice that only one of these is actually in the interval given in the problem. That means that we will exclude the second one (since it isn't in the interval). The number that we're after in this problem is,

$$c = 0.7863$$

Be careful to not assume that only one of the numbers will work. It is possible for both of them to work.

Example 3

Suppose that we know that $f(x)$ is continuous and differentiable on $[6, 15]$. Let's also suppose that we know that $f(6) = -2$ and that we know that $f'(x) \le 10$. What is the largest possible value for $f(15)$?

Solution

Let's start with the conclusion of the Mean Value Theorem.

$$f(15) - f(6) = f'(c)(15 - 6)$$

Plugging in for the known quantities and rewriting this a little gives,

$$f(15) = f(6) + f'(c)(15 - 6) = -2 + 9f'(c)$$

Now we know that $f'(x) \le 10$ so in particular we know that $f'(c) \le 10$. This gives us the following,

$$\begin{aligned} f(15) &= -2 + 9f'(c) \\ &\le -2 + (9)\,10 \\ &= 88 \end{aligned}$$

All we did was replace $f'(c)$ with its largest possible value.

This means that the largest possible value for $f(15)$ is 88.

Example 4

Suppose that we know that $f(x)$ is continuous and differentiable everywhere. Let's also suppose that we know that $f(x)$ has two roots. Show that $f'(x)$ must have at least one root.

Solution

It is important to note here that all we can say is that $f'(x)$ will have at least one root. We can't say that it will have exactly one root. So don't confuse this problem with the first one we worked.

This is actually a fairly simple thing to prove. Since we know that $f(x)$ has two roots let's suppose that they are a and b. Now, by assumption we know that $f(x)$ is continuous and differentiable everywhere and so in particular it is continuous on $[a, b]$ and differentiable on

(a, b).

Therefore, by the Mean Value Theorem there is a number c that is between a and b (this isn't needed for this problem, but it's true so it should be pointed out) and that,

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

But we now need to recall that a and b are roots of $f(x)$ and so this is,

$$f'(c) = \frac{0 - 0}{b - a} = 0$$

Or, $f'(x)$ has a root at $x = c$.

Again, it is important to note that we don't have a value of c. We have only shown that it exists. We also haven't said anything about c being the only root. It is completely possible for $f'(x)$ to have more than one root.

It is completely possible to generalize the previous example significantly. For instance if we know that $f(x)$ is continuous and differentiable everywhere and has three roots we can then show that not only will $f'(x)$ have at least two roots but that $f''(x)$ will have at least one root. We'll leave it to you to verify this, but the ideas involved are identical to those in the previous example.

We'll close this section out with a couple of nice facts that can be proved using the Mean Value Theorem. Note that in both of these facts we are assuming the functions are continuous and differentiable on the interval $[a, b]$.

Fact 1

If $f'(x) = 0$ for all x in an interval (a, b) then $f(x)$ is constant on (a, b).

This fact is very easy to prove so let's do that here.

First, notice that because we are assuming the derivative exists on (a, b) we know that $f(x)$ is differentiable on (a, b). In addition, we know that if a function is differentiable on an interval then it is also continuous on that interval and so $f(x)$ will also be continuous on (a, b).

Now, take any two x's in the interval (a, b), say x_1 and x_2. Then since $f(x)$ is continuous and differentiable on (a, b) it must also be continuous and differentiable on $[x_1, x_2]$. This means that we can apply the Mean Value Theorem for these two values of x. Doing this gives,

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1)$$

where $x_1 < c < x_2$. But by assumption $f'(x) = 0$ for all x in an interval (a, b) and so in particular we must have,

$$f'(c) = 0$$

Putting this into the equation above gives,

$$f(x_2) - f(x_1) = 0 \qquad \Rightarrow \qquad f(x_2) = f(x_1)$$

Now, since x_1 and x_2 were any two values of x in the interval (a, b) we can see that we must have $f(x_2) = f(x_1)$ for all x_1 and x_2 in the interval and this is exactly what it means for a function to be constant on the interval and so we've proven the fact.

Fact 2

If $f'(x) = g'(x)$ for all x in an interval (a, b) then in this interval we have $f(x) = g(x) + c$ where c is some constant.

This fact is a direct result of the previous fact and is also easy to prove.

If we first define,

$$h(x) = f(x) - g(x)$$

Then since both $f(x)$ and $g(x)$ are continuous and differentiable in the interval (a, b) then so must be $h(x)$. Therefore, the derivative of $h(x)$ is,

$$h'(x) = f'(x) - g'(x)$$

However, by assumption $f'(x) = g'(x)$ for all x in an interval (a, b) and so we must have that $h'(x) = 0$ for all x in an interval (a, b). So, by Fact 1 $h(x)$ must be constant on the interval.

This means that we have,

$$\begin{aligned} h(x) &= c \\ f(x) - g(x) &= c \\ f(x) &= g(x) + c \end{aligned}$$

which is what we were trying to show.

4.8 Optimization

In this section we are going to look at optimization problems. In optimization problems we are looking for the largest value or the smallest value that a function can take. We saw how to solve one kind of optimization problem in the Absolute Extrema section where we found the largest and smallest value that a function would take on an interval.

In this section we are going to look at another type of optimization problem. Here we will be looking for the largest or smallest value of a function subject to some kind of constraint. The constraint will be some condition (that can usually be described by some equation) that must absolutely, positively be true no matter what our solution is. On occasion, the constraint will not be easily described by an equation, but in these problems it will be easy to deal with as we'll see.

This section is generally one of the more difficult for students taking a Calculus course. One of the main reasons for this is that a subtle change of wording can completely change the problem. There is also the problem of identifying the quantity that we'll be optimizing and the quantity that is the constraint and writing down equations for each.

The first step in all of these problems should be to very carefully read the problem. Once you've done that the next step is to identify the quantity to be optimized and the constraint.

In identifying the constraint remember that the constraint is the quantity that must be true regardless of the solution. In almost every one of the problems we'll be looking at here one quantity will be clearly indicated as having a fixed value and so must be the constraint. Once you've got that identified the quantity to be optimized should be fairly simple to get. It is however easy to confuse the two if you just skim the problem so make sure you carefully read the problem first!

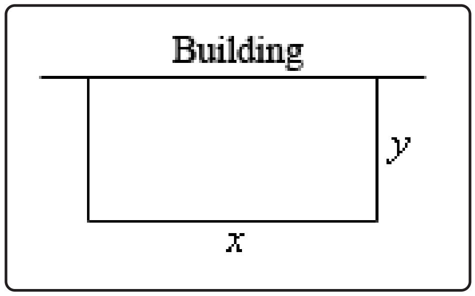
Let's start the section off with a simple problem to illustrate the kinds of issues we will be dealing with here.

Example 1

We need to enclose a rectangular field with a fence. We have 500 feet of fencing material and a building is on one side of the field and so won't need any fencing. Determine the dimensions of the field that will enclose the largest area.

Solution

In all of these problems we will have two functions. The first is the function that we are actually trying to optimize and the second will be the constraint. Sketching the situation will often help us to arrive at these equations so let's do that.



In this problem we want to maximize the area of a field and we know that will use 500 ft of fencing material. So, the area will be the function we are trying to optimize and the amount of fencing is the constraint. The two equations for these are,

$$\text{Maximize : } A = xy$$
$$\text{Constraint : } 500 = x + 2y$$

Okay, we know how to find the largest or smallest value of a function provided it's only got a single variable. The area function (as well as the constraint) has two variables in it and so what we know about finding absolute extrema won't work. However, if we solve the constraint for one of the two variables we can substitute this into the area and we will then have a function of a single variable.

So, let's solve the constraint for x. Note that we could have just as easily solved for y but that would have led to fractions and so, in this case, solving for x will probably be best.

$$x = 500 - 2y$$

Substituting this into the area function gives a function of y.

$$A(y) = (500 - 2y)\,y = 500y - 2y^2$$

Now we want to find the largest value this will have on the interval $[0, 250]$. The limits in this interval corresponds to taking $y = 0$ (*i.e.* no sides to the fence) and $y = 250$ (*i.e.* only two sides and no width, also if there are two sides each must be 250 ft to use the whole 500 ft).

Note that the endpoints of the interval won't make any sense from a physical standpoint if we actually want to enclose some area because they would both give zero area. They do, however, give us a set of limits on y and so the Extreme Value Theorem tells us that we will have a maximum value of the area somewhere between the two endpoints. Having these limits will also mean that we can use the process we discussed in the Finding Absolute Extrema section earlier in the chapter to find the maximum value of the area.

So, recall that the maximum value of a continuous function (which we've got here) on a closed interval (which we also have here) will occur at critical points and/or end points. As we've already pointed out the end points in this case will give zero area and so don't make any sense. That means our only option will be the critical points.

So, let's get the derivative and find the critical points.

$$A'(y) = 500 - 4y$$

Setting this equal to zero and solving gives a lone critical point of $y = 125$. Plugging this into the area gives an area of $A(125) = 31250\,\text{ft}^2$. So according to the method from Absolute Extrema section this must be the largest possible area, since the area at either endpoint is zero.

Finally, let's not forget to get the value of x and then we'll have the dimensions since this is what the problem statement asked for. We can get the x by plugging in our y into the constraint.

$$x = 500 - 2(125) = 250$$

The dimensions of the field that will give the largest area, subject to the fact that we used exactly 500 ft of fencing material, are 250 x 125.

Don't forget to actually read the problem and give the answer that was asked for. These types of problems can take a fair amount of time/effort to solve and it's not hard to sometimes forget what the problem was actually asking for.

In the previous problem we used the method from the Finding Absolute Extrema section to find the maximum value of the function we wanted to optimize. However, as we'll see in later examples it will not always be easy to find endpoints. Also, even if we can find the endpoints we will see that sometimes dealing with the endpoints may not be easy either. Not only that, but this method requires that the function we're optimizing be continuous on the interval we're looking at, including the endpoints, and that may not always be the case.

So, before proceeding with any more examples let's spend a little time discussing some methods for determining if our solution is in fact the absolute minimum/maximum value that we're looking for. In some examples all of these will work while in others one or more won't be all that useful. However, we will always need to use some method for making sure that our answer is in fact that optimal value that we're after.

Method 1 : : Use the method used in Finding Absolute Extrema.

This is the method used in the first example above. Recall that in order to use this method the interval of possible values of the independent variable in the function we are optimizing, let's call it I, must have finite endpoints. Also, the function we're optimizing (once it's down to a single variable) must be continuous on I, including the endpoints. If these conditions are met then we know that the optimal value, either the maximum or minimum depending on the problem, will occur at either the endpoints of the range or at a critical point that is inside the range of possible solutions.

There are two main issues that will often prevent this method from being used however. First, not every problem will actually have a range of possible solutions that have finite endpoints at both ends. We'll see at least one example of this as we work through the remaining examples. Also, many of the functions we'll be optimizing will not be continuous once we reduce them down to a single variable and this will prevent us from using this method.

Method 2 : Use a variant of the First Derivative Test.

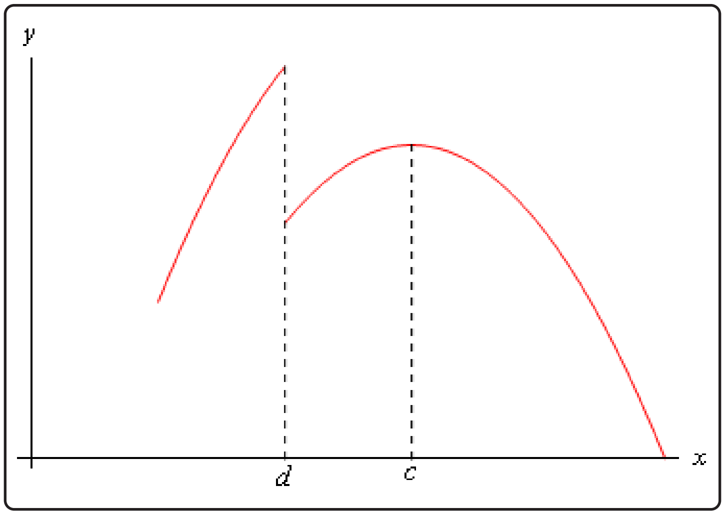
In this method we also will need an interval of possible values of the independent variable in the function we are optimizing, I. However, in this case, unlike the previous method the endpoints do not need to be finite. Also, we will need to require that the function be continuous on the interior of the interval I and we will only need the function to be continuous at the end points if the endpoint is finite and the function actually exists at the endpoint. We'll see several problems where the function we're optimizing doesn't actually exist at one of the endpoints. This will not prevent this method from being used.

Let's suppose that $x = c$ is a critical point of the function we're trying to optimize, $f(x)$. We already know from the First Derivative Test that if $f'(x) > 0$ immediately to the left of $x = c$ (*i.e.* the function is increasing immediately to the left) and if $f'(x) < 0$ immediately to the right of $x = c$ (*i.e.* the function is decreasing immediately to the right) then $x = c$ will be a relative maximum for $f(x)$.

Now, this does not mean that the absolute maximum of $f(x)$ will occur at $x = c$. However, suppose that we knew a little bit more information. Suppose that in fact we knew that $f'(x) > 0$ for all x in I such that $x < c$. Likewise, suppose that we knew that $f'(x) < 0$ for all x in I such that $x > c$. In this case we know that to the left of $x = c$, provided we stay in I of course, the function is always increasing and to the right of $x = c$, again staying in I, we are always decreasing. In this case we can say that the absolute maximum of $f(x)$ in I will occur at $x = c$.

Similarly, if we know that to the left of $x = c$ the function is always decreasing and to the right of $x = c$ the function is always increasing then the absolute minimum of $f(x)$ in I will occur at $x = c$.

Before we give a summary of this method let's discuss the continuity requirement a little. Nowhere in the above discussion did the continuity requirement apparently come into play. We require that the function we're optimizing to be continuous in I to prevent the following situation.



In this case, a relative maximum of the function clearly occurs at $x = c$. Also, the function is always decreasing to the right and is always increasing to the left. However, because of the discontinuity at $x = d$, we can clearly see that $f(d) > f(c)$ and so the absolute maximum of the function does not occur at $x = c$. Had the discontinuity at $x = d$ not been there this would not have happened and the absolute maximum would have occurred at $x = c$.

Here is a summary of this method.

First Derivative Test for Absolute Extrema

Let I be the interval of all possible values of x in $f(x)$, the function we want to optimize, and further suppose that $f(x)$ is continuous on I , except possibly at the endpoints. Finally suppose that $x = c$ is a critical point of $f(x)$ and that c is in the interval I. If we restrict x to values from I (*i.e.* we only consider possible optimal values of the function) then,

1. If $f'(x) > 0$ for all $x < c$ and if $f'(x) < 0$ for all $x > c$ then $f(c)$ will be the absolute maximum value of $f(x)$ on the interval I.
2. If $f'(x) < 0$ for all $x < c$ and if $f'(x) > 0$ for all $x > c$ then $f(c)$ will be the absolute minimum value of $f(x)$ on the interval I.

Method 3 : Use the second derivative.

There are actually two ways to use the second derivative to help us identify the optimal value of a function and both use the Second Derivative Test to one extent or another.

The first way to use the second derivative doesn't actually help us to identify the optimal value. What it does do is allow us to potentially exclude values and knowing this can simplify our work somewhat and so is not a bad thing to do.

Suppose that we are looking for the absolute maximum of a function and after finding the critical points we find that we have multiple critical points. Let's also suppose that we run all of them through the second derivative test and determine that some of them are in fact relative minimums of the function. Since we are after the absolute maximum we know that a maximum (of any kind) can't occur at relative minimums and so we immediately know that we can exclude these points from further consideration. We could do a similar check if we were looking for the absolute minimum. Doing this may not seem like all that great of a thing to do, but it can, on occasion, lead to a nice reduction in the amount of work that we need to do in later steps.

The second way of using the second derivative to identify the optimal value of a function is in fact very similar to the second method above. In fact, we will have the same requirements for this method as we did in that method. We need an interval of possible values of the independent variable in function we are optimizing, call it I as before, and the endpoint(s) may or may not be finite. We'll also need to require that the function, $f(x)$ be continuous everywhere in I except possibly at the endpoints as above.

Now, suppose that $x = c$ is a critical point and that $f''(c) > 0$. The second derivative test tells us that $x = c$ must be a relative minimum of the function. Suppose however that we also knew that $f''(x) > 0$ for all x in I. In this case we would know that the function was concave up in all of I and that would in turn mean that the absolute minimum of $f(x)$ in I would in fact have to be at $x = c$.

Likewise, if $x = c$ is a critical point and $f''(x) < 0$ for all x in I then we would know that the function was concave down in I and that the absolute maximum of $f(x)$ in I would have to be at $x = c$.

Here is a summary of this method.

Second Derivative Test for Absolute Extrema

Let I be the interval of all possible values of x in $f(x)$, the function we want to optimize, and suppose that $f(x)$ is continuous on I , except possibly at the endpoints. Finally suppose that $x = c$ is a critical point of $f(x)$ and that c is in the interval I. Then,

1. If $f''(x) > 0$ for all x in I then $f(c)$ will be the absolute minimum value of $f(x)$ on the interval I.
2. If $f''(x) < 0$ for all x in I then $f(c)$ will be the absolute maximum value of $f(x)$ on the interval I.

As we work examples over the next two sections we will use each of these methods as needed in the examples. In some cases, the method we use will be the only method we could use, in others it will be the easiest method to use and in others it will simply be the method we chose to use for that example. It is important to realize that we won't be able to use each of the methods for every example. With some examples one method will be easiest to use or may be the only method that can be used, however, each of the methods described above will be used at least a couple of times through out all of the examples.

It is also important to be aware that some problems don't allow any of the methods discussed above to be used exactly as outlined above. We may need to modify one of them or use a combination of them to fully work the problem. There is an example in the next section where none of the methods above work easily, although we do also present an alternative solution method in which we can use at least one of the methods discussed above.

Next, the vast majority of the examples worked over the course of the next section will only have a single critical point. Problems with more than one critical point are often difficult to know which critical point(s) give the optimal value. There are a couple of examples in the next two sections with more than one critical point including one in the next section mentioned above in which none of the methods discussed above easily work. In that example you can see some of the ideas you might need to do in order to find the optimal value.

Finally, in all of the methods above we referenced an interval I. This was done to make the discussion a little easier. However, in all of the examples over the next two sections we will never explicitly say "this is the interval I". Just remember that the interval I is just the largest interval of possible values of the independent variable in the function we are optimizing.

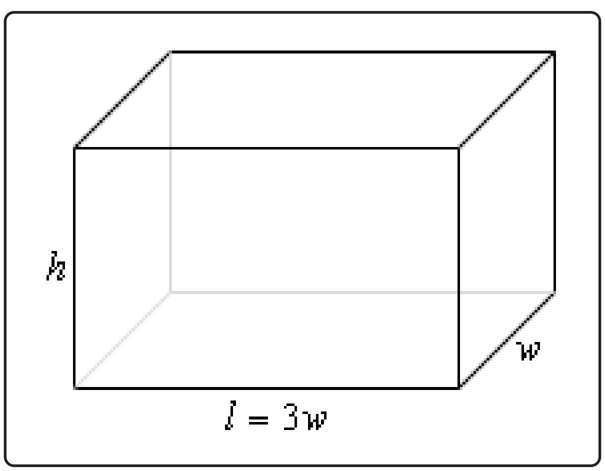
Okay, let's work some more examples.

Example 2

We want to construct a box whose base length is 3 times the base width. The material used to build the top and bottom cost \$10/ft^2 and the material used to build the sides cost \$6/ft^2. If the box must have a volume of 50ft^3 determine the dimensions that will minimize the cost to build the box.

Solution

First, a quick figure (probably not to scale...).



We want to minimize the cost of the materials subject to the constraint that the volume must be 50ft^3. Note as well that the cost for each side is just the area of that side times the appropriate cost.

The two functions we'll be working with here this time are,

$$\text{Minimize : } C = 10\,(2lw) + 6\,(2wh + 2lh)\; = 60w^2 + 48wh$$
$$\text{Constraint : } 50 = lwh = 3w^2h$$

As with the first example, we will solve the constraint for one of the variables and plug this into the cost. It will definitely be easier to solve the constraint for h so let's do that.

$$h = \frac{50}{3w^2}$$

Plugging this into the cost gives,

$$C\,(w) = 60w^2 + 48w\left(\frac{50}{3w^2}\right) = 60w^2 + \frac{800}{w}$$

Now, let's get the first and second (we'll be needing this later...) derivatives,

$$C'\,(w) = 120w - 800w^{-2} = \frac{120w^3 - 800}{w^2} \qquad\qquad C''\,(w) = 120 + 1600w^{-3}$$

Now we need the critical point(s) for the cost function. First, notice that $w = 0$ is not a critical point. Clearly the derivative does not exist at $w = 0$ but then neither does the function and remember that values of w will only be critical points if the function also exists at that point. Note that there is also a physical reason to avoid $w = 0$. We are constructing a box and it would make no sense to have a zero width of the box.

So it looks like the only critical point will come from determining where the numerator is zero.

$$120w^3 - 800 = 0 \qquad \Rightarrow \qquad w = \sqrt[3]{\frac{800}{120}} = \sqrt[3]{\frac{20}{3}} = 1.8821$$

So, we've got a single critical point and we now have to verify that this is in fact the value that will give the absolute minimum cost.

In this case we can't use Method 1 from above. First, the function is not continuous at one of the endpoints, $w = 0$, of our interval of possible values, *i.e.* $w > 0$. Secondly, there is no theoretical upper limit to the width that will give a box with volume of 50 ft^3. If w is very large then we would just need to make h very small.

The second method listed above would work here, but that's going to involve some calculations, not difficult calculations, but more work nonetheless.

The third method however, will work quickly and simply here. First, we know that whatever the value of w that we get it will have to be positive and we can see second derivative above that provided $w > 0$ we will have $C''(w) > 0$ and so in the interval of possible optimal values the cost function will always be concave up and so $w = 1.8821$ must give the absolute minimum cost.

All we need to do now is to find the remaining dimensions.

$$\begin{aligned} w &= 1.8821 \\ l &= 3w = 3\,(1.8821) = 5.6463 \\ h &= \frac{50}{3w^2} = \frac{50}{3(1.8821)^2} = 4.7050 \end{aligned}$$

Also, even though it was not asked for, the minimum cost is : $C\,(1.8821) = \$637.60$.

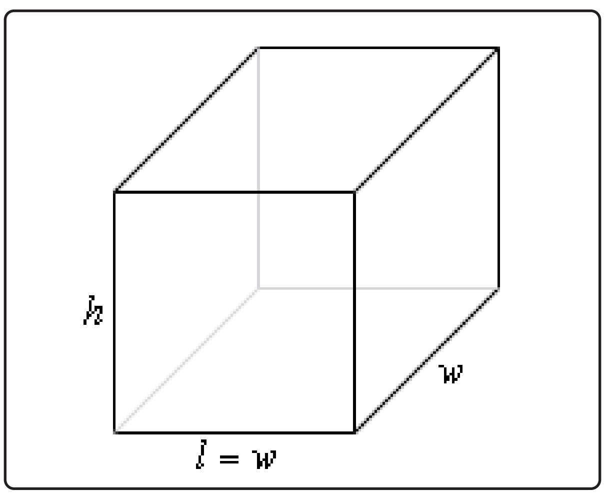
Example 3

We want to construct a box with a square base and we only have 10 m^2 of material to use in construction of the box. Assuming that all the material is used in the construction process determine the maximum volume that the box can have.

Solution

This example is in many ways the exact opposite of the previous example. In this case we want to optimize the volume and the constraint this time is the amount of material used. We don't have a cost here, but if you think about it the cost is nothing more than the amount of material used times a cost and so the amount of material and cost are pretty much tied together. If you can do one you can do the other as well. Note as well that the amount of material used is really just the surface area of the box.

As always, let's start off with a quick sketch of the box.



Now, as mentioned above we want to maximize the volume and the amount of material is the constraint so here are the equations we'll need.

$$\text{Maximize : } V = lwh = w^2 h$$
$$\text{Constraint : } 10 = 2lw + 2wh + 2lh = 2w^2 + 4wh$$

We'll solve the constraint for h and plug this into the equation for the volume.

$$h = \frac{10 - 2w^2}{4w} = \frac{5 - w^2}{2w} \qquad \Rightarrow \qquad V(w) = w^2\left(\frac{5 - w^2}{2w}\right) = \frac{1}{2}\left(5w - w^3\right)$$

Here are the first and second derivatives of the volume function.

$$V'(w) = \frac{1}{2}\left(5 - 3w^2\right) \qquad V''(w) = -3w$$

Note as well here that provided $w > 0$, which from a physical standpoint we know must be true for the width of the box, then the volume function will be concave down and so if we get a single critical point then we know that it will have to be the value that gives the absolute maximum.

Setting the first derivative equal to zero and solving gives us the two critical points,

$$w = \pm\sqrt{\frac{5}{3}} = \pm 1.2910$$

In this case we can exclude the negative critical point since we are dealing with a length of a box and we know that these must be positive. Do not however get into the habit of just excluding any negative critical point. There are problems where negative critical points are perfectly valid possible solutions.

Now, as noted above we got a single critical point, 1.2910, and so this must be the value that gives the maximum volume and since the maximum volume is all that was asked for in the problem statement the answer is then

$$V(1.2910) = 2.1517\,\text{m}^3$$

.

Note that we could also have noted here that if $0 < w < 1.2910$ then $V'(w) > 0$ (using a test point we have $V'(1) = 1 > 0$) and likewise if $w > 1.2910$ then $V'(w) < 0$ (using a test point we have $V'(2) = -\frac{7}{2} < 0$) and so if we are to the left of the critical point the volume is always increasing and if we are to the right of the critical point the volume is always decreasing and so by the Method 2 above we can also see that the single critical point must give the absolute maximum of the volume.

Finally, even though these weren't asked for here are the dimension of the box that gives the maximum volume.

$$l = w = 1.2910 \qquad\qquad h = \frac{5 - 1.2910^2}{2\,(1.2910)} = 1.2910$$

So, it looks like in this case we actually have a perfect cube.

In the last two examples we've seen that many of these optimization problems can be done in both directions so to speak. In both examples we have essentially the same two equations: volume and surface area. However, in Example 2 the volume was the constraint and the cost (which is directly related to the surface area) was the function we were trying to optimize. In Example 3, on the other hand, we were trying to optimize the volume and the surface area was the constraint.

It is important to not get so locked into one way of doing these problems that we can't do it in the opposite direction as needed as well. This is one of the more common mistakes that students make with these kinds of problems. They see one problem and then try to make every other problem that seems to be the same conform to that one solution even if the problem needs to be worked differently. Keep an open mind with these problems and make sure that you understand what is being optimized and what the constraint is before you jump into the solution.

Also, as seen in the last example we used two different methods of verifying that we did get the optimal value. Do not get too locked into one method of doing this verification that you forget about the other methods.

Let's work some another example that this time doesn't involve a rectangle or box.

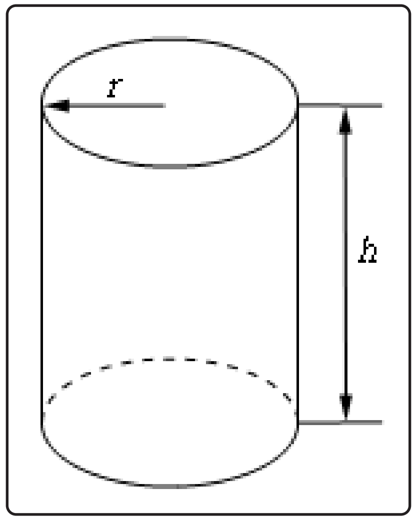
Example 4

A manufacturer needs to make a cylindrical can that will hold 1.5 liters of liquid. Determine the dimensions of the can that will minimize the amount of material used in its construction.

Solution

Before starting the solution let's first address the fact that we are using liters for volume. Because we want length measurements for the radius and height we'll also need the volume to in terms of a length measurement. We can do this using the fact that 1 Liter = 1000 cm^3 and so we can convert 1.5 liters into 1500 cm^3. This will in turn give a radius and height in terms of centimeters.

In this problem the constraint is the volume and we want to minimize the amount of material used. This means that what we want to minimize is the surface area of the can and we'll need to include both the walls of the can as well as the top and bottom "caps". Here is a quick sketch to get us started off.



We'll need the surface area of this can and that will be the surface area of the walls of the can (which is really just a cylinder) and the area of the top and bottom caps (which are just disks, and don't forget that there are two of them).

Note that if you think of a cylinder of height h and radius r as just a bunch of disks/circles of radius r stacked on top of each other the equations for the surface area and volume are pretty simple to remember. The volume is just the area of each of the disks times the height. Similarly, the surface area of the walls of the cylinder is just the circumference of each circle times the height. We also can't forget to add in the area of the two caps, πr^2, to the total surface area.

So, the equation for the volume and surface area of the walls of a cylinder are then,

$$V = \left(\pi r^2\right)(h) = \pi r^2 h \qquad\qquad A = (2\pi r)(h) = 2\pi r h$$

Adding the surface area of the caps of the cylinder to the surface area the equations that we'll need for this problem are,

$$\text{Minimize : } A = 2\pi r h + 2\pi r^2$$
$$\text{Constraint :} 1500 = \pi r^2 h$$

In this case it looks like our best option is to solve the constraint for h and plug this into the area function.

$$h = \frac{1500}{\pi r^2} \qquad \Rightarrow \qquad A(r) = 2\pi r\left(\frac{1500}{\pi r^2}\right) + 2\pi r^2 = 2\pi r^2 + \frac{3000}{r}$$

Notice that this formula will only make sense from a physical standpoint if $r > 0$ which is a good thing as it is not defined at $r = 0$.

Next, let's get the first derivative.

$$A'(r) = 4\pi r - \frac{3000}{r^2} = \frac{4\pi r^3 - 3000}{r^2}$$

From this we can see that we have one critical points : $r = \sqrt[3]{\frac{750}{\pi}} = 6.2035$ (where the derivative is zero). Note that $r = 0$ is not a critical point because the area function does not exist there, which makes sense from a physical standpoint as well given that we know that r must be positive in order to actually have a can.

So, we only have a single critical point to deal with here and notice that 6.2035 is the only value for which the derivative will be zero and hence the only place (with $r > 0$ of course) that the derivative may change sign. It's not difficult, using test points, to check that if $0 < r < 6.2035$ then $A'(r) < 0$ and likewise if $r > 6.2035$ then $A'(r) > 0$. The variant of the First Derivative Test above then tells us that the absolute minimum value of the area (for $r > 0$) must occur at

$$r = 6.2035$$

All we need to do this is determine height of the can and we'll be done.

$$h = \frac{1500}{\pi(6.2035)^2} = 12.4070$$

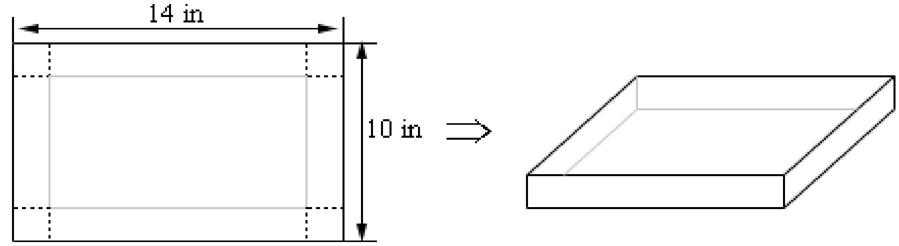
Therefore, if the manufacturer makes the can with a radius of 6.2035 cm and a height of 12.4070 cm the least amount of material will be used to make the can.

As an interesting side problem and extension to the above example you might want to show that for a given volume, L, the minimum material will be used if $h = 2r$ regardless of the volume of the can.

In the examples to this point we've put in quite a bit of discussion in the solution. In the remaining problems we won't be putting in quite as much discussion and leave it to you to fill in any missing details.

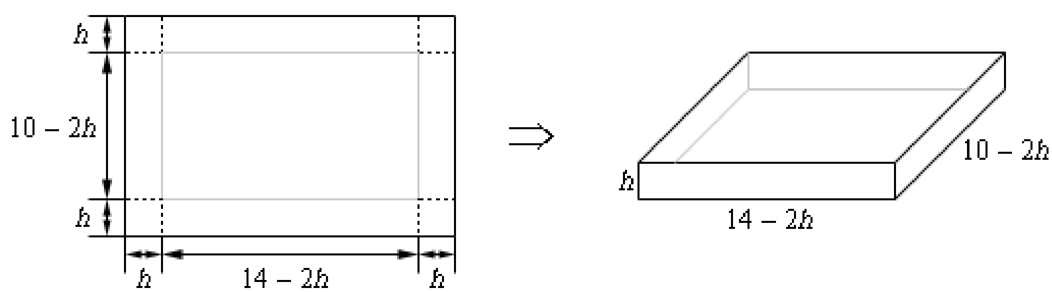
Example 5

We have a piece of cardboard that is 14 inches by 10 inches and we're going to cut out the corners as shown below and fold up the sides to form a box, also shown below. Determine the height of the box that will give a maximum volume.



Solution

Let's let the height of the box be h. So, the width/length of the corners being cut out is also h and so the vertical side will have a "new" height of $10 - 2h$ and the horizontal side will have a "new" width of $14 - 2h$. Here is a sketch with all this information put in,



In this example, for the first time, we've run into a problem where the constraint doesn't really have an equation. The constraint is simply the size of the piece of cardboard and has already been factored into the figure above. This will happen on occasion and so don't get excited about it when it does. This just means that we have one less equation to worry about. In this case we want to maximize the volume. Here is the volume, in terms of h and its first derivative.

$$V(h) = h(14 - 2h)(10 - 2h) = 140h - 48h^2 + 4h^3 \qquad V'(h) = 140 - 96h + 12h^2$$

Setting the first derivative equal to zero and solving gives the following two critical points,

$$h = \frac{12 \pm \sqrt{39}}{3} = 1.9183, \quad 6.0817$$

We now have an apparent problem. We have two critical points and we'll need to determine which one is the value we need. The fact that we have two critical points means that neither the first derivative test or the second derivative test can be used here as they both require a single critical point. This isn't a real problem however. Go back to the figure at the start of the solution and notice that we can quite easily find limits on h. The smallest h can be is $h = 0$ even though this doesn't make much sense as we won't get a box in this case. Also, from the 10 inch side we can see that the largest h can be is $h = 5$ although again, this doesn't make much sense physically.

So, knowing that whatever h is it must be in the range $0 \le h \le 5$ we can see that the second critical point is outside this range and so the only critical point that we need to worry about is 1.9183.

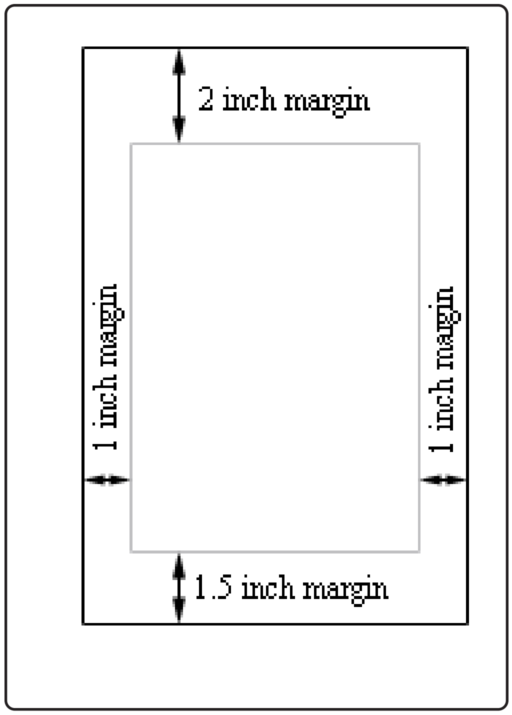
Finally, since the volume is defined and continuous on $0 \le h \le 5$ all we need to do is plug in the critical points and endpoints into the volume to determine which gives the largest volume. Here are those function evaluations.

$$V(0) = 0 \qquad V(1.9183) = 120.1644 \qquad V(5) = 0$$

So, if we take $h = 1.9183$ we get a maximum volume.

Example 6

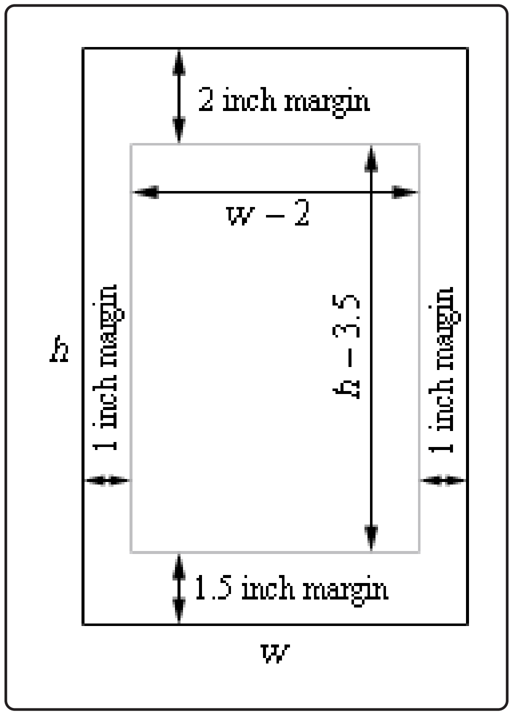
A printer needs to make a poster that will have a total area of 200 in^2 and will have 1 inch margins on the sides, a 2 inch margin on the top and a 1.5 inch margin on the bottom as shown below. What dimensions will give the largest printed area?



Solution

This problem is a little different from the previous problems. Both the constraint and the function we are going to optimize are areas. The constraint is that the overall area of the poster must be 200 in^2 while we want to optimize the printed area (*i.e.* the area of the poster with the margins taken out).

Let's define the height of the poster to be h and the width of the poster to be w. Here is a new sketch of the poster and we can see that once we've taken the margins into account the width of the printed area is $w - 2$ and the height of the printer area is $h - 3.5$.



Here are the equations that we'll be working with.

$$\text{Maximize : } A = (w-2)(h-3.5)$$
$$\text{Constraint : } 200 = wh$$

Solving the constraint for h and plugging into the equation for the printed area gives,

$$A(w) = (w-2)\left(\frac{200}{w} - 3.5\right) = 207 - 3.5w - \frac{400}{w}$$

The first and second derivatives are,

$$A'(w) = -3.5 + \frac{400}{w^2} = \frac{400 - 3.5w^2}{w^2} \qquad A''(w) = -\frac{800}{w^3}$$

From the first derivative we have the following two critical points ($w = 0$ is not a critical point because the area function does not exist there).

$$w = \pm\sqrt{\frac{400}{3.5}} = \pm 10.6904$$

However, since we're dealing with the dimensions of a piece of paper we know that we must have $w > 0$ and so only 10.6904 will make sense.

Also notice that provided $w > 0$ the second derivative will always be negative and so in the range of possible optimal values of the width the area function is always concave down and so we know that the maximum printed area will be at $w = 10.6904$ inches.

The height of the paper that gives the maximum printed area is then,

$$h = \frac{200}{10.6904} = 18.7084 \text{ inches}$$

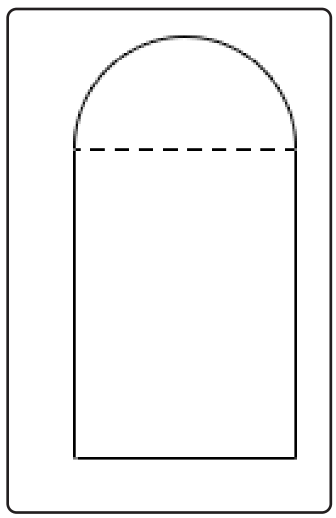
We've worked quite a few examples to this point and we have quite a few more to work. However, this section has gotten quite lengthy so let's continue our examples in the next section. This is being done mostly because these notes are also being presented on the web and this will help to keep the load times on the pages down somewhat.

4.9 More Optimization

Because these notes are also being presented on the web we've broken the optimization examples up into several sections to keep the load times to a minimum. Do not forget the various methods for verifying that we have the optimal value that we looked at in the previous section. In this section we'll just use them without acknowledging so make sure you understand them and can use them. So let's get going on some more examples.

Example 1

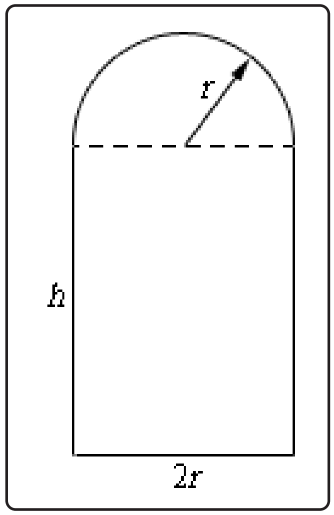
A window is being built and the bottom is a rectangle and the top is a semicircle. If there is 12 meters of framing materials what must the dimensions of the window be to let in the most light?



Solution

Okay, let's ask this question again in slightly easier to understand terms. We want a window in the shape described above to have a maximum area (and hence let in the most light) and have a perimeter of 12 m (because we have 12 m of framing material). Little bit easier to understand in those terms.

Let the radius of the semicircle on the top be r and the height of the rectangle be h. Now, because the semicircle is on top of the window we can think of the width of the rectangular portion at 2 r as shown below.



The perimeter (our constraint) is the lengths of the three sides on the rectangular portion plus half the circumference of a circle of radius r. The area (what we want to maximize) is the area of the rectangle plus half the area of a circle of radius r. Here are the equations we'll be working with in this example.

$$\text{Maximize : } A = 2hr + \frac{1}{2}\pi r^2$$
$$\text{Constraint : } 12 = 2h + 2r + \pi r$$

In this case we'll solve the constraint for h and plug that into the area equation.

$$h = 6 - r - \frac{1}{2}\pi r \qquad \Rightarrow \qquad A\left(r\right) = 2r\left(6 - r - \frac{1}{2}\pi r\right) + \frac{1}{2}\pi r^2 = 12r - 2r^2 - \frac{1}{2}\pi r^2$$

The first and second derivatives are,

$$A'\left(r\right) = 12 - r\left(4 + \pi\right) \qquad\qquad A''\left(r\right) = -4 - \pi$$

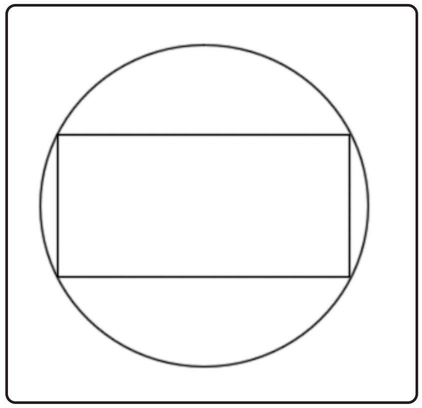
We can see that the only critical point is,

$$r = \frac{12}{4 + \pi} = 1.6803$$

We can also see that the second derivative is always negative (in fact it's a constant) and so we can see that the maximum area must occur at this point. So, for the maximum area the semicircle on top must have a radius of 1.6803 and the rectangle must have the dimensions 3.3606 x 1.6803 $(2r$ x $h)$.

Example 2

Determine the area of the largest rectangle that can be inscribed in a circle of radius 4.



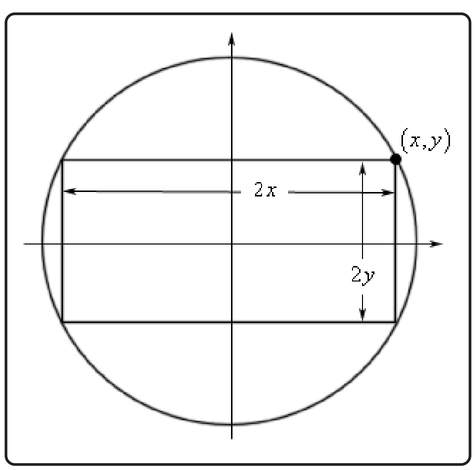
Solution

Huh? This problem type of problem never seems to make sense originally. What we want to do is maximize the area of the largest rectangle that we can fit inside a circle and have all of its corners touching the circle.

To do this problem it's easiest to assume that the circle (and hence the rectangle) is centered at the origin of a standard xy axis system. Doing this we know that the equation of the circle will be

$$x^2 + y^2 = 16$$

and that the right upper corner of the rectangle will have the coordinates (x, y). This means that the width of the rectangle will be $2x$ and the height of the rectangle will be $2y$ as shown below



The area of the rectangle will then be,

$$A = (2x)(2y) = 4xy$$

So, we've got the function we want to maximize (the area), but what is the constraint? Well since the coordinates of the upper right corner must be on the circle we know that x and y must satisfy the equation of the circle. In other words, the equation of the circle is the constraint.

The first thing to do then is to solve the constraint for one of the variables.

$$y = \pm\sqrt{16 - x^2}$$

Since the point that we're looking at is in the first quadrant we know that y must be positive and so we can take the "+" part of this. Plugging this into the area and computing the first derivative gives,

$$\begin{aligned} A(x) &= 4x\sqrt{16 - x^2} \\ A'(x) &= 4\sqrt{16 - x^2} - \frac{4x^2}{\sqrt{16 - x^2}} = \frac{64 - 8x^2}{\sqrt{16 - x^2}} \end{aligned}$$

Before getting the critical points let's notice that we can limit x to the range $0 \le x \le 4$ since we are assuming that x is in the first quadrant and must stay inside the circle. Now the four critical points we get (two from the numerator and two from the denominator) are,

$$\begin{aligned} 16 - x^2 &= 0 & &\Rightarrow & x &= \pm 4 \\ 64 - 8x^2 &= 0 & &\Rightarrow & x &= \pm 2\sqrt{2} \end{aligned}$$

We only want critical points that are in the range of possible optimal values so that means that we have two critical points to deal with : $x = 2\sqrt{2}$ and $x = 4$. Notice however that the second critical point is also one of the endpoints of our interval.

Now, area function is continuous and we have an interval of possible solution with finite endpoints so,

$$A(0) = 0 \qquad A\left(2\sqrt{2}\right) = 32 \qquad A(4) = 0$$

So, we can see that we'll get the maximum area if $x = 2\sqrt{2}$ and the corresponding value of y is,

$$y = \sqrt{16 - \left(2\sqrt{2}\right)^2} = \sqrt{8} = 2\sqrt{2}$$

It looks like the maximum area will be found if the inscribed rectangle is in fact a square.

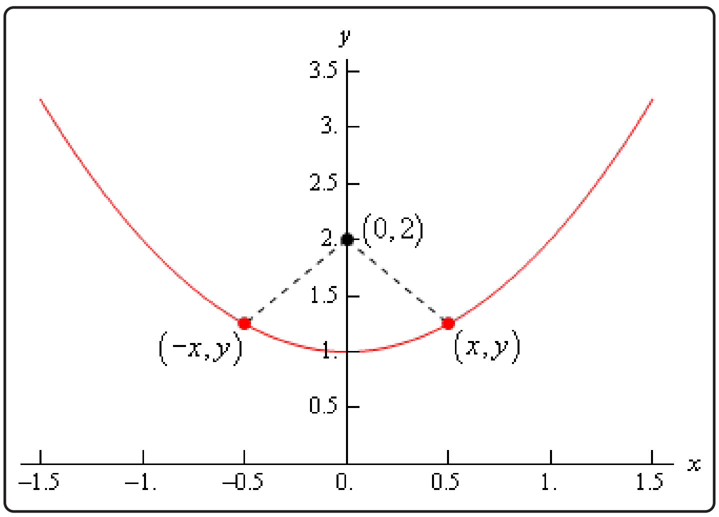
We need to again make a point that was made several times in the previous section. We excluded several critical points in the work above. Do not always expect to do that. There will often be physical reasons to exclude zero and/or negative critical points, however, there will be problems where these are perfectly acceptable values. You should always write down every possible critical point and then exclude any that can't be possible solutions. This keeps you in the habit of finding all the critical points and then deciding which ones you actually need and that in turn will make it less likely that you'll miss one when it is actually needed.

Example 3

Determine the point(s) on $y = x^2 + 1$ that are closest to $(0, 2)$.

Solution

Here's a quick sketch of the situation.



So, we're looking for the shortest length of the dashed line. Notice as well that if the shortest distance isn't at $x = 0$ there will be two points on the graph, as we've shown above, that will give the shortest distance. This is because the parabola is symmetric to the y-axis and the point in question is on the y-axis. This won't always be the case of course so don't always expect two points in these kinds of problems.

In this case we need to minimize the distance between the point $(0, 2)$ and any point that is one the graph (x, y). Or,

$$d = \sqrt{(x-0)^2 + (y-2)^2} = \sqrt{x^2 + (y-2)^2}$$

If you think about the situation here it makes sense that the point that minimizes the distance will also minimize the square of the distance and so since it will be easier to work with we will use the square of the distance and minimize that. If you aren't convinced of this we'll take a closer look at this after this problem. So, the function that we're going to minimize is,

$$D = d^2 = x^2 + (y-2)^2$$

The constraint in this case is the function itself since the point must lie on the graph of the function.

At this point there are two methods for proceeding. One of which will require significantly more work than the other. Let's take a look at both of them.

Solution 1

In this case we will use the constraint in probably the most obvious way. We already have the constraint solved for y so let's plug that into the square of the distance and get the derivatives.

$$\begin{aligned} D(x) &= x^2 + \left(x^2+1-2\right)^2 = x^4 - x^2 + 1 \\ D'(x) &= 4x^3 - 2x = 2x\left(2x^2-1\right) \\ D''(x) &= 12x^2 - 2 \end{aligned}$$

So, it looks like there are three critical points for the square of the distance and notice that this time, unlike pretty much every previous example we've worked, we can't exclude zero or negative numbers. They are perfectly valid possible optimal values this time.

$$x = 0, \qquad x = \pm\frac{1}{\sqrt{2}}$$

Before going any farther, let's check these in the second derivative to see if they are all relative minimums.

$$D''(0) = -2 < 0 \qquad D''\left(\frac{1}{\sqrt{2}}\right) = 4 \qquad\qquad D''\left(-\frac{1}{\sqrt{2}}\right) = 4$$

So, $x = 0$ is a relative maximum and so can't possibly be the minimum distance. That means that we've got two critical points. The question is how we verify that these give the minimum distance and yes we did mean to say that both will give the minimum distance. Recall from our sketch above that if x gives the minimum distance then so will $-x$ and so if gives the minimum distance then the other should as well.

None of the methods we discussed in the previous section will really work here. We don't have an interval of possible solutions with finite endpoints and both the first and second derivative change sign. In this case however, we can still verify that they are the points that give the minimum distance.

First, let's see what we have if we are working on the interval $\left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$. On this interval we can try to use the first method of finding absolute extrema discussed in the previous section. That says to evaluate the function at the endpoints and the critical points and in this case, even though we've excluded it we'll need to include $x = 0$ since it is a critical point in the region. Doing this gives,

$$D\left(-\frac{1}{\sqrt{2}}\right) = \frac{3}{4} \qquad D(0) = 1 \qquad D\left(\frac{1}{\sqrt{2}}\right) = \frac{3}{4}$$

So, we can see that the absolute minimum in the interval must occur at $x = \pm\frac{1}{\sqrt{2}}$.

Next, we can see that if $x < -\frac{1}{\sqrt{2}}$ then $D'(x) < 0$. Or in other words, if $x < -\frac{1}{\sqrt{2}}$ the function is decreasing until it hits $x = -\frac{1}{\sqrt{2}}$ and so must always be larger than the function at $x = -\frac{1}{\sqrt{2}}$.

Similarly, $x > \frac{1}{\sqrt{2}}$ then $D'(x) > 0$ and so the function is always increasing to the right of $x = \frac{1}{\sqrt{2}}$ and so must be larger than the function at $x = \frac{1}{\sqrt{2}}$.

Putting all of this together tells us that we do in fact have an absolute minimum at $x = \pm\frac{1}{\sqrt{2}}$.

All that we need to do is to find the value of y for these points.

$$\begin{aligned} x &= \frac{1}{\sqrt{2}} &&: \quad y = \frac{3}{2} \\ x &= -\frac{1}{\sqrt{2}} &&: \quad y = \frac{3}{2} \end{aligned}$$

So, the points on the graph that are closest to $(0, 2)$ are,

$$\left(\frac{1}{\sqrt{2}}, \frac{3}{2}\right) \qquad \left(-\frac{1}{\sqrt{2}}, \frac{3}{2}\right)$$

This solution method shows how tricky it can be to know that we have absolute extrema when there are multiple critical points and none of the methods discussed in the last section will work. Luckily for us, there is another, easier, method we could have done instead.

Solution 2

The first solution that we worked was actually the long solution. There is a much shorter, and easier, solution to this problem. Instead of plugging y into the square of the distance let's plug in x. From the constraint we get,

$$x^2 = y - 1$$

and notice that the only place x show up in the square of the distance it shows up as x^2 and let's just plug this into the square of the distance. Doing this gives,

$$\begin{aligned} D(y) &= y - 1 + (y-2)^2 = y^2 - 3y + 3 \\ D'(y) &= 2y - 3 \\ D''(y) &= 2 \end{aligned}$$

There is now a single critical point, $y = \frac{3}{2}$, and since the second derivative is always positive we know that this point must give the absolute minimum. So, all that we need to do at this point is find the value(s) of x that go with this value of y.

$$x^2 = \frac{3}{2} - 1 = \frac{1}{2} \qquad \Rightarrow \qquad x = \pm\frac{1}{\sqrt{2}}$$

The points are then,

$$\left(\frac{1}{\sqrt{2}}, \frac{3}{2}\right) \qquad \left(-\frac{1}{\sqrt{2}}, \frac{3}{2}\right)$$

So, for significantly less work we got exactly the same answer.

This previous example had a couple of nice points. First, as pointed out in the problem, we couldn't exclude zero or negative critical points this time as we've done in all the previous examples. Again, be careful to not get into the habit of always excluding them as we do many of the examples we'll work.

Next, some of these problems will have multiple solution methods and sometimes one will be significantly easier than the other. The method you use is up to you and often the difficulty of any particular method is dependent upon the person doing the problem. One person may find one way easier and other person may find a different method easier.

Finally, as we saw in the first solution method sometimes we'll need to use a combination of the optimal value verification methods we discussed in the previous section.

Now, before we move onto the next example let's take a look at the claim above that we could find the location of the point that minimizes the distance by finding the point that minimizes the square of the distance. We'll generalize things a little bit,

Fact

Suppose that we have a positive function, $f(x) > 0$, that exists everywhere then $f(x)$ and $g(x) = \sqrt{f(x)}$ will have the same critical points and the relative extrema will occur at the same points.

This is simple enough to prove so let's do that here. First let's take the derivative of $g(x)$ and see what we can determine about the critical points of $g(x)$.

$$g'(x) = \frac{1}{2}[f(x)]^{-\frac{1}{2}} f'(x) = \frac{f'(x)}{2\sqrt{f(x)}}$$

Let's plug $x = c$ into this to get,

$$g'(c) = \frac{f'(c)}{2\sqrt{f(c)}}$$

By assumption we know that $f(c)$ exists and $f(c) > 0$ and therefore the denominator of this will always exist and will never be zero. We'll need this in several places so we can't forget this.

If $f'(c) = 0$ then because we know that the denominator will not be zero here we must also have $g'(c) = 0$. Likewise, if $g'(c) = 0$ then we must have $f'(c) = 0$. So, $f(x)$ and $g(x)$ will have the same critical points in which the derivatives will be zero.

Next, if $f'(c)$ doesn't exist then $g'(c)$ will also not exist and likewise if $g'(c)$ doesn't exist then because we know that the denominator will not be zero then this means that $f'(c)$ will also not exist. Therefore, $f(x)$ and $g(x)$ will have the same critical points in which the derivatives does not exist.

So, the upshot of all this is that $f(x)$ and $g(x)$ will have the same critical points.

Next, let's notice that because we know that $2\sqrt{f(x)} > 0$ then $f'(x)$ and $g'(x)$ will have the same sign and so if we apply the first derivative test (and recalling that they have the same critical points) to each of these functions we can see that the results will be the same and so the relative extrema will occur at the same points.

Note that we could also use the second derivative test to verify that the critical points will have the same classification if we wanted to. The second derivative is (and you should see if you can use the quotient rule to verify this),

$$g''(x) = \frac{2\sqrt{f(x)}\, f''(x) - \left[f'(x)\right]^2 \left[f(x)\right]^{-\frac{1}{2}}}{4f(x)}$$

Then if $x = c$ is a critical point such that $f'(c) = 0$ (and so we can use the second derivative test) we get,

$$\begin{aligned} g''(c) &= \frac{2\sqrt{f(c)}\, f''(c) - \left[f'(c)\right]^2 \left[f(c)\right]^{-\frac{1}{2}}}{4f(c)} \\ &= \frac{2\sqrt{f(c)}\, f''(c) - [0]^2 \left[f(c)\right]^{-\frac{1}{2}}}{4f(c)} = \frac{\sqrt{f(c)}\, f''(c)}{2f(c)} \end{aligned}$$

Now, because we know that $displaystyle 2\sqrt{f(c)} > 0$ and by assumption $f(c) > 0$ we can see that $f''(c)$ and $g''(c)$ will have the same sign and so will have the same conclusion from the second derivative test.

So, now that we have that out of the way let's work some more examples.

Example 4

A 2 feet piece of wire is cut into two pieces and one piece is bent into a square and the other is bent into an equilateral triangle. Where, if anywhere, should the wire be cut so that the total area enclosed by both is minimum and maximum?

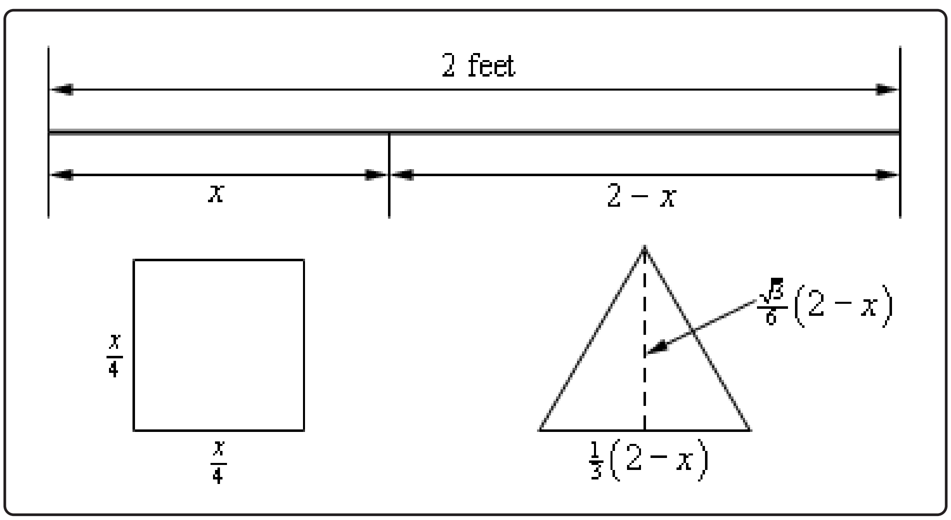
Solution

Before starting the solution recall that an equilateral triangle is a triangle with three equal sides and each of the interior angles are $\frac{\pi}{3}$ (or $60°$).

Also, note the "if anywhere" portion of the problem statement. What this is saying is that it is possible to take the full piece of wire and put all of it into either a square or a triangle. Do not forget about this as it will be important later on in the problem.

Now, this is another problem where the constraint isn't really going to be given by an equation, it is simply that there is 2 ft of wire to work with and this will be taken into account in our work.

So, let's cut the wire into two pieces. The first piece will have length x which we'll bend into a square and each side will have length $\frac{x}{4}$. The second piece will then have length $2 - x$ (we just used the constraint here...) and we'll bend this into an equilateral triangle and each side will have length $\frac{1}{3}(2 - x)$. Here is a sketch of all this.



As noted in the sketch above we also will need the height of the triangle. This is easy to get if you realize that the dashed line divides the equilateral triangle into two other triangles. Let's look at the right one. The hypotenuse is $\frac{1}{3}(2 - x)$ while the lower right angle is $\frac{\pi}{3}$. Finally, the height is then the opposite side to the lower right angle so using basic right triangle trig we arrive at the height of the triangle as follows.

$$\sin\left(\frac{\pi}{3}\right) = \frac{\text{opp}}{\text{hyp}} \qquad \Rightarrow \qquad \text{opp} = \frac{1}{3}(2-x)\sin\left(\frac{\pi}{3}\right) = \frac{1}{3}(2-x)\left(\frac{\sqrt{3}}{2}\right) = \frac{\sqrt{3}}{6}(2-x)$$

So, the total area of both objects is then,

$$A(x) = \left(\frac{x}{4}\right)^2 + \frac{1}{2}\left(\frac{1}{3}(2-x)\right)\left(\frac{\sqrt{3}}{6}(2-x)\right) = \frac{x^2}{16} + \frac{\sqrt{3}}{36}(2-x)^2$$

Here's the first derivative of the area.

$$A'(x) = \frac{x}{8} + \frac{\sqrt{3}}{36}(2)(2-x)(-1) = \frac{x}{8} - \frac{\sqrt{3}}{9} + \frac{\sqrt{3}}{18}x$$

Setting this equal to zero and solving gives the single critical point of,

$$x = \frac{8\sqrt{3}}{9+4\sqrt{3}} = 0.8699$$

Now, let's notice that the problem statement asked for both the minimum and maximum enclosed area and we got a single critical point. This clearly can't be the answer to both, but this is not the problem that it might seem to be.

Let's notice that x must be in the range $0 \le x \le 2$ and since the area function is continuous we use the basic process for finding absolute extrema of a function.

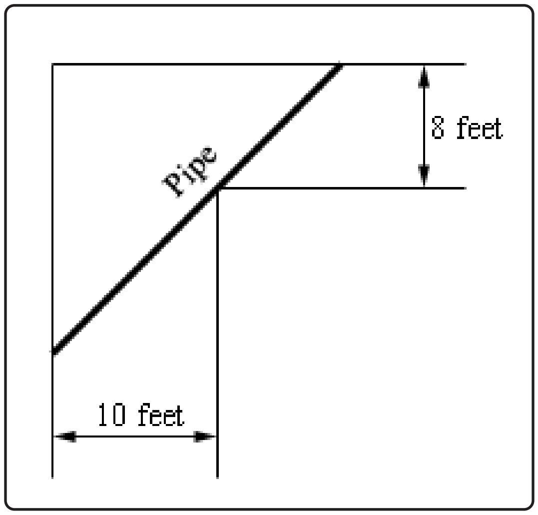
$$A(0) = 0.1925 \qquad A(0.8699) = 0.1087 \qquad A(2) = 0.25$$

So, it looks like the minimum area will arise if we take $x = 0.8699$ while the maximum area will arise if we take the whole piece of wire and bend it into a square.

As the previous problem illustrated we can't get too locked into the answers always occurring at the critical points as they have to this point. That will often happen, but one of the extrema in the previous problem was at an endpoint and that will happen on occasion.

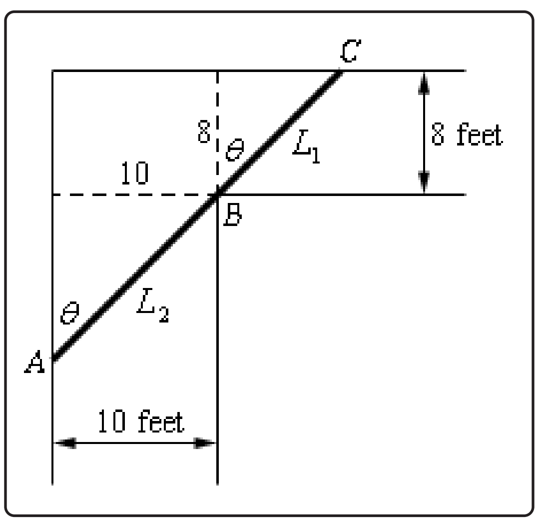
Example 5

A piece of pipe is being carried down a hallway that is 10 feet wide. At the end of the hallway the there is a right-angled turn and the hallway narrows down to 8 feet wide. What is the longest pipe that can be carried (always keeping it horizontal) around the turn in the hallway?



Solution

Let's start off with a sketch of the situation adding in some more information so we can get a grip on what's going on and how we're going to have to go about solving this.



The largest pipe that can go around the turn will do so in the position shown above. One end will be touching the outer wall of the hall way at A and C and the pipe will touch the inner corner at B. Let's assume that the length of the pipe in the small hallway is L_1 while L_2 is the length of the pipe in the large hallway. The pipe then has a length of $L = L_1 + L_2$.

Now, if $\theta = 0$ then the pipe is completely in the wider hallway and we can see that as $\theta \to 0$ the point A will move down the vertical wall and the point C will move along the horizontal wall closer and closer to the corner and as this happens L lengthens and so $L \to \infty$ as $\theta \to 0$.

Likewise, if $\theta = \frac{\pi}{2}$ the pipe is completely in the narrow hallway and as $\theta \to \frac{\pi}{2}$ we also have $L \to \infty$ by a similar line of reasoning above for $\theta \to 0$.

So, because $L \to \infty$ as we near the ends of the interval of possible angles somewhere in the interior of the interval, $0 < \theta < \frac{\pi}{2}$, is an angle that will minimize L and oddly enough that is the length that we're after. The largest pipe that will fit around the turn will in fact be the minimum value of L.

The constraint for this problem is not so obvious and there are actually two of them. The constraints for this problem are the widths of the hallways. We'll use these to get an equation for L in terms of θ and then we'll minimize this new equation.

So, using basic right triangle trig we can see that,

$$L_1 = 8\sec(\theta) \qquad L_2 = 10\csc(\theta) \qquad \Rightarrow \qquad L = 8\sec(\theta) + 10\csc(\theta)$$

So, differentiating L gives,

$$L' = 8\sec(\theta)\tan(\theta) - 10\csc(\theta)\cot(\theta)$$

Setting this equal to zero and solving gives,

$$\begin{aligned} 8\sec(\theta)\tan(\theta) &= 10\csc(\theta)\cot(\theta) \\ \frac{\sec(\theta)\tan(\theta)}{\csc(\theta)\cot(\theta)} &= \frac{10}{8} \\ \frac{\sin(\theta)\tan^2(\theta)}{\cos(\theta)} &= \frac{5}{4} \qquad \Rightarrow \qquad \tan^3(\theta) = 1.25 \end{aligned}$$

Solving for θ gives,

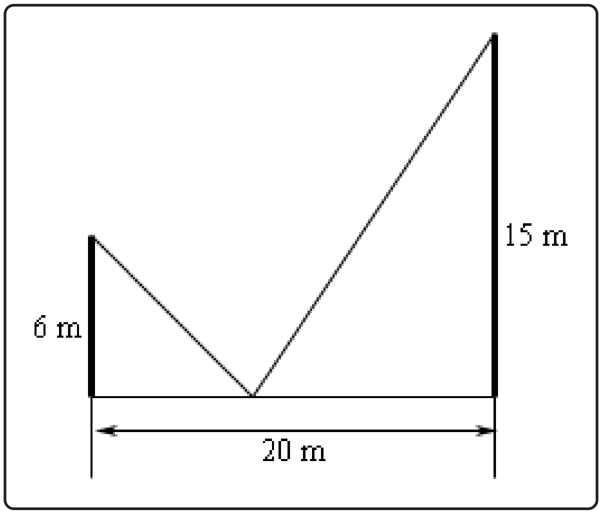
$$\tan(\theta) = \sqrt[3]{1.25} \qquad \Rightarrow \qquad \theta = \tan^{-1}\left(\sqrt[3]{1.25}\right) = 0.8226$$

So, if $\theta = 0.8226$ radians then the pipe will have a minimum length and will just fit around the turn. Anything larger will not fit around the turn and so the largest pipe that can be carried around the turn is,

$$L = 8\sec(0.8226) + 10\csc(0.8226) = 25.4033 \text{ feet}$$

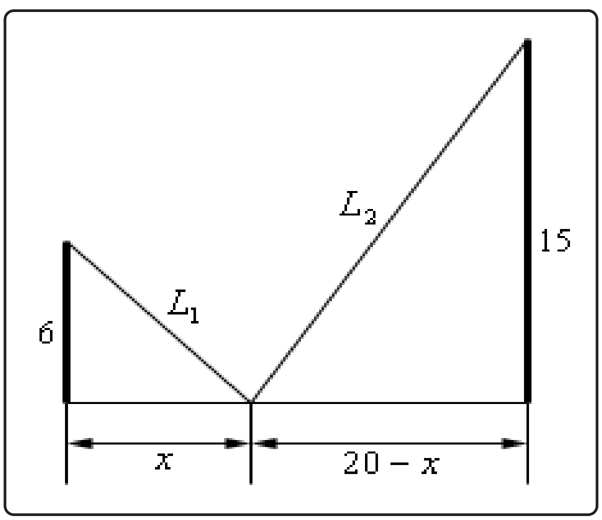
Example 6

Two poles, one 6 meters tall and one 15 meters tall, are 20 meters apart. A length of wire is attached to the top of each pole and it is also staked to the ground somewhere between the two poles. Where should the wire be staked so that the minimum amount of wire is used?



Solution

As always let's start off with a sketch of this situation with some more information added.



The total length of the wire is $L = L_1 + L_2$ and we need to determine the value of x that will minimize this. The constraint in this problem is that the poles must be 20 meters apart and that x must be in the range $0 \le x \le 20$. The first thing that we'll need to do here is to get the length of wire in terms of x, which is fairly simple to do using the Pythagorean Theorem.

$$L_1 = \sqrt{36 + x^2} \qquad L_2 = \sqrt{225 + (20 - x)^2} \qquad L = \sqrt{36 + x^2} + \sqrt{625 - 40x + x^2}$$

Not the nicest function we've had to work with but there it is. Note however, that it is a continuous function and we've got an interval with finite endpoints and so finding the absolute minimum won't require much more work than just getting the critical points of this function. So, let's do that. Here's the derivative.

$$L' = \frac{x}{\sqrt{36 + x^2}} + \frac{x - 20}{\sqrt{625 - 40x + x^2}}$$

Setting this equal to zero gives,

$$\begin{aligned} \frac{x}{\sqrt{36 + x^2}} + \frac{x - 20}{\sqrt{625 - 40x + x^2}} &= 0 \\ x\sqrt{625 - 40x + x^2} &= -(x - 20)\sqrt{36 + x^2} \end{aligned}$$

It's probably been quite a while since you've been asked to solve something like this. To solve this, we'll need to square both sides to get rid of the roots, but this will cause problems as well soon see. Let's first just square both sides and solve that equation.

$$\begin{aligned} x^2\left(625 - 40x + x^2\right) &= (x - 20)^2\left(36 + x^2\right) \\ 625x^2 - 40x^3 + x^4 &= 14400 - 1440x + 436x^2 - 40x^3 + x^4 \\ 189x^2 + 1440x - 14400 &= 0 \\ 9(3x + 40)(7x - 40) &= 0 \qquad \Rightarrow \qquad x = -\frac{40}{3}, \quad x = \frac{40}{7} \end{aligned}$$

Note that if you can't do that factoring don't worry, you can always just use the quadratic formula and you'll get the same answers.

Okay two issues that we need to discuss briefly here. The first solution above (note that we didn't call it a critical point...) doesn't make any sense because it is negative and outside of the range of possible solutions and so we can ignore it.

Secondly, and maybe more importantly, if you were to plug $x = -\frac{40}{3}$ into the derivative you would not get zero and so is not even a critical point. How is this possible? It is a solution after all. We'll recall that we squared both sides of the equation above and it was mentioned at the time that this would cause problems. We'll we've hit those problems. In squaring both sides we've inadvertently introduced a new solution to the equation. When you do something like this you should ALWAYS go back and verify that the solutions that you get are in fact solutions to the original equation. In this case we were lucky and the "bad" solution also happened to be outside the interval of solutions we were interested in but that won't always be the case.

So, if we go back and do a quick verification we can in fact see that the only critical point is $x = \frac{40}{7} = 5.7143$ and this is nicely in our range of acceptable solutions, $0 \le x \le 20$.

Now all that we need to do is plug this critical point and the endpoints of the wire into the length formula and identify the one that gives the minimum value.

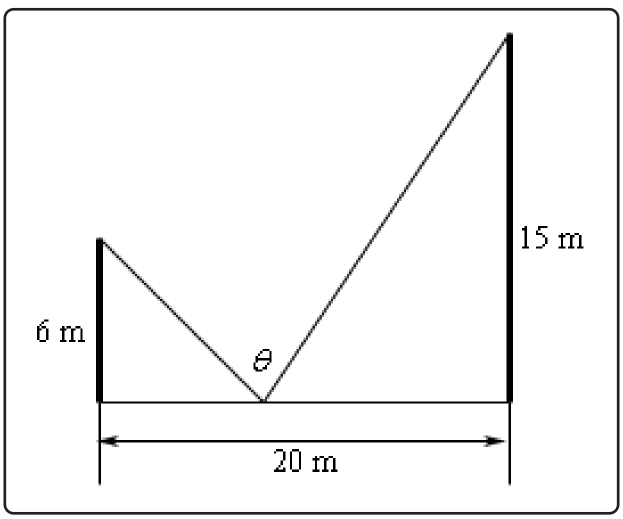
$$L(0) = 31 \qquad L\left(\frac{40}{7}\right) = 29 \qquad L(20) = 35.8806$$

So, we will get the minimum length of wire if we stake it to the ground $\frac{40}{7}$ feet from the smaller pole.

Let's do a modification of the above problem that asks a completely different question.

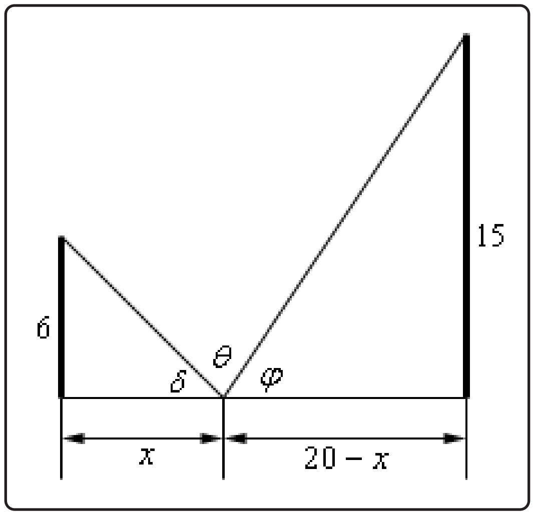
Example 7

Two poles, one 6 meters tall and one 15 meters tall, are 20 meters apart. A length of wire is attached to the top of each pole and it is also staked to the ground somewhere between the two poles. Where should the wire be staked so that the angle formed by the two pieces of wire at the stake is a maximum?



Solution

Here's a sketch for this example with some more information added.



The equation that we're going to need to work with here is not obvious. We can see from the sketch above that,

$$\delta + \theta + \varphi = 180 = \pi$$

Note that we need to make sure that the equation is equal to π because of how we're going to work this problem. Now, basic right triangle trig tells us the following,

$$\begin{aligned} \tan(\delta) &= \frac{6}{x} & \Rightarrow \quad \delta &= \tan^{-1}\left(\frac{6}{x}\right) \\ \tan(\varphi) &= \frac{15}{20-x} & \Rightarrow \quad \varphi &= \tan^{-1}\left(\frac{15}{20-x}\right) \end{aligned}$$

Plugging these into the equation above and solving for θ gives,

$$\theta = \pi - \tan^{-1}\left(\frac{6}{x}\right) - \tan^{-1}\left(\frac{15}{20-x}\right)$$

Note that this is the reason for the π in our equation. The inverse tangents give angles that are in radians and so can't use the 180 that we're used to in this kind of equation.

Next, we'll need the derivative so hopefully you'll recall how to differentiate inverse tangents.

$$\begin{aligned} \theta' &= -\frac{1}{1+\left(\frac{6}{x}\right)^2}\left(-\frac{6}{x^2}\right) - \frac{1}{1+\left(\frac{15}{20-x}\right)^2}\left(\frac{15}{(20-x)^2}\right) \\ &= \frac{6}{x^2+36} - \frac{15}{(20-x)^2+225} \\ &= \frac{6}{x^2+36} - \frac{15}{x^2-40x+625} = \frac{-3\left(3x^2+80x-1070\right)}{\left(x^2+36\right)\left(x^2-40x+625\right)} \end{aligned}$$

Setting this equal to zero and solving give the following two critical points.

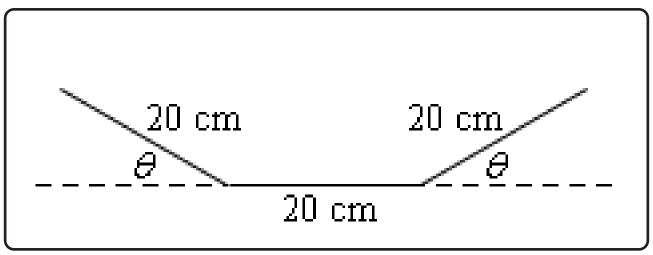
$$x = \frac{-40 \pm \sqrt{4810}}{3} = -36.4514, \quad 9.7847$$

The first critical point is not in the interval of possible solutions and so we can exclude it.

Clearly x must be in the interval $[0, 20]$ and so using test points it's not difficult to show that if $0 \le x < 9.7847$ we have $\theta' > 0$ (and so θ is increasing) and if $9.7847 < x \le 20$ that $\theta' < 0$ (and so θ is decreasing). So by the first derivative test when $x = 9.7847$ we will get the maximum value of θ.

Example 8

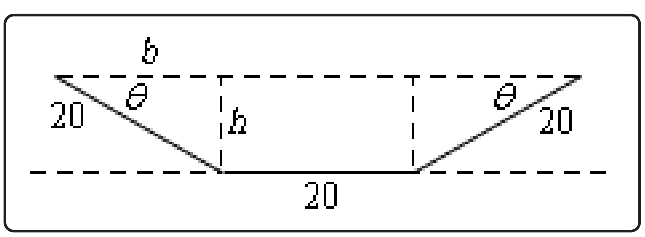
A trough for holding water is formed by taking a piece of sheet metal 60 cm wide and folding the 20 cm on either end up as shown below. Determine the angle θ that will maximize the amount of water that the trough can hold.



Solution

Now, in this case we are being asked to maximize the volume that a trough can hold, but if you think about it the volume of a trough in this shape is nothing more than the cross-sectional area times the length of the trough. So, for a given length in order to maximize the volume all you really need to do is maximize the cross-sectional area.

To get a formula for the cross-sectional area let's redo the sketch above a little.



We can think of the cross-sectional area as a rectangle in the middle with width 20 and height h and two identical triangles on either end with height h, base b and hypotenuse 20. Also note that basic geometry tells us that the angle between the hypotenuse and the base must also be the same angle θ that we had in our original sketch.

Also, basic right triangle trig tells us that the base and height can be written as,

$$b = 20\cos(\theta) \qquad\qquad h = 20\sin(\theta)$$

The cross-sectional area for the whole trough, in terms of θ, is then,

$$A = 20h + 2\left(\frac{1}{2}bh\right) = 400\sin(\theta) + (20\cos(\theta))(20\sin(\theta)) = 400(\sin(\theta) + \sin(\theta)\cos(\theta))$$

The derivative of the area is,

$$\begin{aligned} A'(\theta) &= 400\left(\cos(\theta) + \cos^2(\theta) - \sin^2(\theta)\right) \\ &= 400\left(\cos(\theta) + \cos^2(\theta) - \left(1 - \cos^2(\theta)\right)\right) \\ &= 400\left(2\cos^2(\theta) + \cos(\theta) - 1\right) \\ &= 400(2\cos(\theta) - 1)(\cos(\theta) + 1) \end{aligned}$$

So, we have either,

$$\begin{aligned} 2\cos(\theta) - 1 = 0 \qquad &\Rightarrow \qquad \cos(\theta) = \frac{1}{2} \qquad \Rightarrow \qquad \theta = \frac{\pi}{3} \\ \cos(\theta) + 1 = 0 \qquad &\Rightarrow \qquad \cos(\theta) = -1 \qquad \Rightarrow \qquad \theta = \pi \end{aligned}$$

However, we can see that θ must be in the interval $0 \le \theta \le \frac{\pi}{2}$ or we won't get a trough in the proper shape. Therefore, the second critical point makes no sense and also note that we don't need to add on the standard "$+\,2\pi n$" for the same reason.

Finally, since the equation for the area is continuous all we need to do is plug in the critical point and the end points to find the one that gives the maximum area.

$$A(0) = 0 \qquad A\left(\frac{\pi}{3}\right) = 519.6152 \qquad A\left(\frac{\pi}{2}\right) = 400$$

So, we will get a maximum cross-sectional area, and hence a maximum volume, when $\theta = \frac{\pi}{3}$.

4.10 L'Hospital's Rule and Indeterminate Forms

Back in the chapter on Limits we saw methods for dealing with the following limits.

$$\lim_{x\to 4}\frac{x^2-16}{x-4} \qquad\qquad \lim_{x\to\infty}\frac{4x^2-5x}{1-3x^2}$$

In the first limit if we plugged in $x = 4$ we would get $0/0$ and in the second limit if we "plugged" in infinity we would get $\infty/-\infty$ (recall that as x goes to infinity a polynomial will behave in the same fashion that its largest power behaves). Both of these are called **indeterminate forms**. In both of these cases there are competing interests or rules and it's not clear which will win out.

In the case of $0/0$ we typically think of a fraction that has a numerator of zero as being zero. However, we also tend to think of fractions in which the denominator is going to zero, in the limit, as infinity or might not exist at all. Likewise, we tend to think of a fraction in which the numerator and denominator are the same as one. So, which will win out? Or will neither win out and they all "cancel out" and the limit will reach some other value?

In the case of $\infty/-\infty$ we have a similar set of problems. If the numerator of a fraction is going to infinity we tend to think of the whole fraction going to infinity. Also, if the denominator is going to infinity, in the limit, we tend to think of the fraction as going to zero. We also have the case of a fraction in which the numerator and denominator are the same (ignoring the minus sign) and so we might get -1. Again, it's not clear which of these will win out, if any of them will win out.

With the second limit there is the further problem that infinity isn't really a number and so we really shouldn't even treat it like a number. Much of the time it simply won't behave as we would expect it to if it was a number. To look a little more into this, check out the Types of Infinity section in the Extras appendix at the end of this document.

This is the problem with indeterminate forms. It's just not clear what is happening in the limit. There are other types of indeterminate forms as well. Some other types are,

$$(0)(\pm\infty) \qquad 1^{\infty} \qquad 0^0 \qquad \infty^0 \qquad \infty-\infty$$

These all have competing interests or rules that tell us what should happen and it's just not clear which, if any, of the interests or rules will win out. The topic of this section is how to deal with these kinds of limits.

As already pointed out we do know how to deal with some kinds of indeterminate forms already. For the two limits above we work them as follows.

$$\lim_{x\to 4}\frac{x^2-16}{x-4} = \lim_{x\to 4}(x+4) = 8$$

$$\lim_{x\to\infty}\frac{4x^2-5x}{1-3x^2} = \lim_{x\to\infty}\frac{4-\frac{5}{x}}{\frac{1}{x^2}-3} = -\frac{4}{3}$$

In the first case we simply factored, canceled and took the limit and in the second case we factored out an x^2 from both the numerator and the denominator and took the limit. Notice as well that none of the competing interests or rules in these cases won out! That is often the case.

So, we can deal with some of these. However, what about the following two limits.

$$\lim_{x\to 0}\frac{\sin(x)}{x} \qquad\qquad \lim_{x\to\infty}\frac{\mathbf{e}^x}{x^2}$$

This first is a $0/0$ indeterminate form, but we can't factor this one. The second is an ∞/∞ indeterminate form, but we can't just factor an x^2 out of the numerator. So, nothing that we've got in our bag of tricks will work with these two limits.

This is where the subject of this section comes into play.

L'Hospital's Rule

Suppose that we have one of the following cases,

$$\lim_{x\to a}\frac{f(x)}{g(x)}=\frac{0}{0} \qquad \text{OR} \qquad \lim_{x\to a}\frac{f(x)}{g(x)}=\frac{\pm\infty}{\pm\infty}$$

where a can be any real number, infinity or negative infinity. In these cases we have,

$$\lim_{x\to a}\frac{f(x)}{g(x)}=\lim_{x\to a}\frac{f'(x)}{g'(x)}$$

So, L'Hospital's Rule tells us that if we have an indeterminate form $0/0$ or ∞/∞ all we need to do is differentiate the numerator and differentiate the denominator and then take the limit.

Before proceeding with examples let me address the spelling of "L'Hospital". The more modern spelling is "L'Hôpital". However, when I first learned Calculus my teacher used the spelling that I use in these notes and the first text book that I taught Calculus out of also used the spelling that I use here.

Also, as noted on the Wikipedia page for L'Hospital's Rule,

> In the 17th and 18th centuries, the name was commonly spelled "l'Hospital", and he himself spelled his name that way. However, French spellings have been altered: the silent 's' has been removed and replaced with the circumflex over the preceding vowel. The former spelling is still used in English where there is no circumflex.

So, the spelling that I've used here is an acceptable spelling of his name, albeit not the modern spelling, and because I'm used to spelling it as "L'Hospital" that is the spelling that I'm going to use in these notes.

Let's work some examples.

Example 1

Example 1 Evaluate each of the following limits.

(a) $\displaystyle\lim_{x\to 0}\frac{\sin(x)}{x}$

(b) $\displaystyle\lim_{t\to 1}\frac{5t^4-4t^2-1}{10-t-9t^3}$

(c) $\displaystyle\lim_{x\to\infty}\frac{\mathbf{e}^x}{x^2}$

Solution

(a) $\displaystyle\lim_{x\to 0}\frac{\sin(x)}{x}$

So, we have already established that this is a $0/0$ indeterminate form so let's just apply L'Hospital's Rule.

$$\lim_{x\to 0}\frac{\sin(x)}{x}=\lim_{x\to 0}\frac{\cos(x)}{1}=\frac{1}{1}=1$$

(b) $\displaystyle\lim_{t\to 1}\frac{5t^4-4t^2-1}{10-t-9t^3}$

In this case we also have a $0/0$ indeterminate form and if we were really good at factoring we could factor the numerator and denominator, simplify and take the limit. However, that's going to be more work than just using L'Hospital's Rule.

$$\lim_{t\to 1}\frac{5t^4-4t^2-1}{10-t-9t^3}=\lim_{t\to 1}\frac{20t^3-8t}{-1-27t^2}=\frac{20-8}{-1-27}=-\frac{3}{7}$$

(c) $\displaystyle\lim_{x\to\infty}\frac{\mathbf{e}^x}{x^2}$

This was the other limit that we started off looking at and we know that it's the indeterminate form ∞/∞ so let's apply L'Hospital's Rule.

$$\lim_{x\to\infty}\frac{\mathbf{e}^x}{x^2}=\lim_{x\to\infty}\frac{\mathbf{e}^x}{2x}$$

Now we have a small problem. This new limit is also a ∞/∞ indeterminate form. However, it's not really a problem. We know how to deal with these kinds of limits. Just apply L'Hospital's Rule.

$$\lim_{x\to\infty}\frac{\mathbf{e}^x}{x^2}=\lim_{x\to\infty}\frac{\mathbf{e}^x}{2x}=\lim_{x\to\infty}\frac{\mathbf{e}^x}{2}=\infty$$

Sometimes we will need to apply L'Hospital's Rule more than once.

L'Hospital's Rule works great on the two indeterminate forms $0/0$ and $\pm\infty/\pm\infty$. However, there are many more indeterminate forms out there as we saw earlier. Let's take a look at some of those and see how we deal with those kinds of indeterminate forms.

We'll start with the indeterminate form $(0)(\pm\infty)$.

Example 2

Evaluate the following limit.

$$\lim_{x\to 0^+} x\ln x$$

Solution

Note that we really do need to do the right-hand limit here. We know that the natural logarithm is only defined for positive x and so this is the only limit that makes any sense.

Now, in the limit, we get the indeterminate form $(0)(-\infty)$. L'Hospital's Rule won't work on products, it only works on quotients. However, we can turn this into a fraction if we rewrite things a little.

$$\lim_{x\to 0^+} x\ln x = \lim_{x\to 0^+} \frac{\ln x}{{}^{1}\!/_{x}}$$

The function is the same, just rewritten, and the limit is now in the form $-\infty/\infty$ and we can now use L'Hospital's Rule.

$$\lim_{x\to 0^+} x\ln x = \lim_{x\to 0^+} \frac{\ln x}{{}^{1}\!/_{x}} = \lim_{x\to 0^+} \frac{{}^{1}\!/_{x}}{-{}^{1}\!/_{x^2}}$$

Now, this is a mess, but it cleans up nicely.

$$\lim_{x\to 0^+} x\ln x = \lim_{x\to 0^+} \frac{{}^{1}\!/_{x}}{-{}^{1}\!/_{x^2}} = \lim_{x\to 0^+} (-x) = 0$$

In the previous example we used the fact that we can always write a product of functions as a quotient by doing one of the following.

$$f(x)\,g(x) = \frac{g(x)}{{}^{1}\!/_{f(x)}} \qquad \text{OR} \qquad f(x)\,g(x) = \frac{f(x)}{{}^{1}\!/_{g(x)}}$$

Using these two facts will allow us to turn any limit in the form $(0)(\pm\infty)$ into a limit in the form $0/0$ or $\pm\infty/\pm\infty$. Which one of these two we get after doing the rewrite will depend upon which fact we used to do the rewrite. One of the rewrites will give 0/0 and the other will give $\pm\infty/\pm\infty$. It all depends on which function stays in the numerator and which gets moved down to the denominator.

Let's take a look at another example.

Example 3

Evaluate the following limit.

$$\lim_{x \to -\infty} x\mathbf{e}^x$$

Solution

So, it's in the form $(\infty)(0)$. This means that we'll need to write it as a quotient. Moving the x to the denominator worked in the previous example so let's try that with this problem as well.

$$\lim_{x \to -\infty} x\mathbf{e}^x = \lim_{x \to -\infty} \frac{\mathbf{e}^x}{1/x}$$

Writing the product in this way gives us a product that has the form 0/0 in the limit. So, let's use L'Hospital's Rule on the quotient.

$$\lim_{x \to -\infty} x\mathbf{e}^x = \lim_{x \to -\infty} \frac{\mathbf{e}^x}{1/x} = \lim_{x \to -\infty} \frac{\mathbf{e}^x}{-1/x^2} = \lim_{x \to -\infty} \frac{\mathbf{e}^x}{2/x^3} = \lim_{x \to -\infty} \frac{\mathbf{e}^x}{-6/x^4} = \cdots$$

Hummmm.... This doesn't seem to be getting us anywhere. With each application of L'Hospital's Rule we just end up with another $0/0$ indeterminate form and in fact the derivatives seem to be getting worse and worse. Also note that if we simplified the quotient back into a product we would just end up with either $(\infty)(0)$ or $(-\infty)(0)$ and so that won't do us any good.

This does not mean however that the limit can't be done. It just means that we moved the wrong function to the denominator. Let's move the exponential function instead.

$$\lim_{x \to -\infty} x\mathbf{e}^x = \lim_{x \to -\infty} \frac{x}{1/\mathbf{e}^x} = \lim_{x \to -\infty} \frac{x}{\mathbf{e}^{-x}}$$

Note that we used the fact that,

$$\frac{1}{\mathbf{e}^x} = \mathbf{e}^{-x}$$

to simplify the quotient up a little. This will help us when it comes time to take some derivatives. The quotient is now an indeterminate form of $-\infty/\infty$ and using L'Hospital's Rule gives,

$$\lim_{x \to -\infty} x\mathbf{e}^x = \lim_{x \to -\infty} \frac{x}{\mathbf{e}^{-x}} = \lim_{x \to -\infty} \frac{1}{-\mathbf{e}^{-x}} = 0$$

So, when faced with a product $(0)(\pm\infty)$ we can turn it into a quotient that will allow us to use L'Hospital's Rule. However, as we saw in the last example we need to be careful with how we do that on occasion. Sometimes we can use either quotient and in other cases only one will work.

Let's now take a look at the indeterminate forms,

$$1^\infty \qquad 0^0 \qquad \infty^0$$

These can all be dealt with in the following way so we'll just work one example.

Example 4

Evaluate the following limit.

$$\lim_{x\to\infty} x^{\frac{1}{x}}$$

Solution

In the limit this is the indeterminate form ∞^0. We're actually going to spend most of this problem on a different limit. Let's first define the following.

$$y = x^{\frac{1}{x}}$$

Now, if we take the natural log of both sides we get,

$$\ln(y) = \ln\left(x^{\frac{1}{x}}\right) = \frac{1}{x}\ln(x) = \frac{\ln(x)}{x}$$

Let's now take a look at the following limit.

$$\lim_{x\to\infty} \ln(y) = \lim_{x\to\infty} \frac{\ln(x)}{x} = \lim_{x\to\infty} \frac{1/x}{1} = 0$$

This limit was just a L'Hospital's Rule problem and we know how to do those. So, what did this have to do with our limit? Well first notice that,

$$\mathbf{e}^{\ln(y)} = y$$

and so our limit could be written as,

$$\lim_{x\to\infty} x^{\frac{1}{x}} = \lim_{x\to\infty} y = \lim_{x\to\infty} \mathbf{e}^{\ln(y)}$$

We can now use the limit above to finish this problem.

$$\lim_{x\to\infty} x^{\frac{1}{x}} = \lim_{x\to\infty} y = \lim_{x\to\infty} \mathbf{e}^{\ln(y)} = \mathbf{e}^{\lim\limits_{x\to\infty} \ln(y)} = \mathbf{e}^0 = 1$$

With L'Hospital's Rule we are now able to take the limit of a wide variety of indeterminate forms that we were unable to deal with prior to this section.

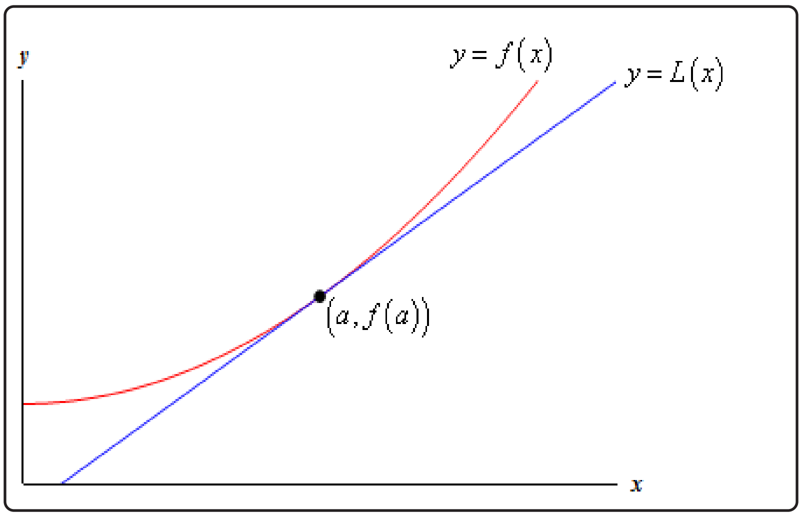
4.11 Linear Approximations

In this section we're going to take a look at an application not of derivatives but of the tangent line to a function. Of course, to get the tangent line we do need to take derivatives, so in some way this is an application of derivatives as well.

Given a function, $f(x)$, we can find its tangent at $x = a$. The equation of the tangent line, which we'll call $L(x)$ for this discussion, is,

$$L(x) = f(a) + f'(a)(x - a)$$

Take a look at the following graph of a function and its tangent line.



From this graph we can see that near $x = a$ the tangent line and the function have nearly the same graph. On occasion we will use the tangent line, $L(x)$, as an approximation to the function, $f(x)$, near $x = a$. In these cases we call the tangent line the **linear approximation** to the function at $x = a$.

So, why would we do this? Let's take a look at an example.

Example 1

Determine the linear approximation for $f(x) = \sqrt[3]{x}$ at $x = 8$. Use the linear approximation to approximate the value of $\sqrt[3]{8.05}$ and $\sqrt[3]{25}$.

Solution

Since this is just the tangent line there really isn't a whole lot to finding the linear approximation.

$$f'(x) = \frac{1}{3}x^{-\frac{2}{3}} = \frac{1}{3\sqrt[3]{x^2}} \qquad f(8) = 2 \qquad f'(8) = \frac{1}{12}$$

The linear approximation is then,

$$L(x) = 2 + \frac{1}{12}(x-8) = \frac{1}{12}x + \frac{4}{3}$$

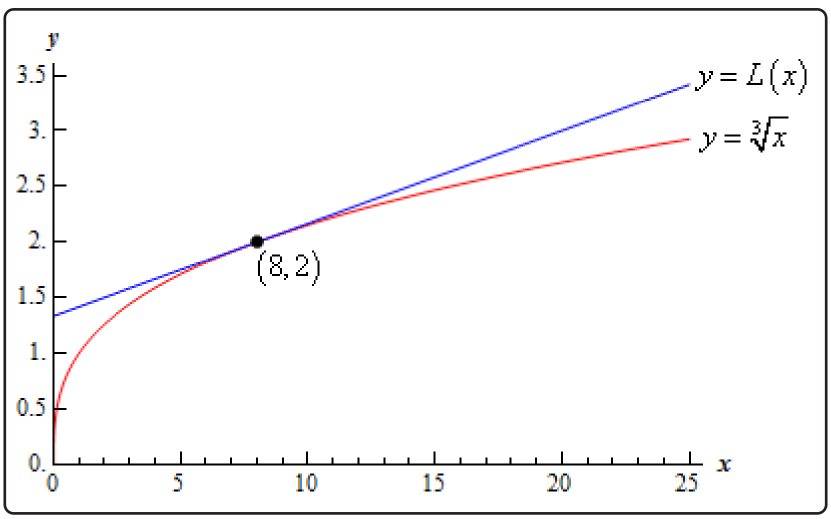
Now, the approximations are nothing more than plugging the given values of x into the linear approximation. For comparison purposes we'll also compute the exact values.

$$L(8.05) = 2.00416667 \qquad \sqrt[3]{8.05} = 2.00415802$$
$$L(25) = 3.41666667 \qquad \sqrt[3]{25} = 2.92401774$$

So, at $x = 8.05$ this linear approximation does a very good job of approximating the actual value. However, at $x = 25$ it doesn't do such a good job.

This shouldn't be too surprising if you think about it. Near $x = 8$ both the function and the linear approximation have nearly the same slope and since they both pass through the point $(8, 2)$ they should have nearly the same value as long as we stay close to $x = 8$. However, as we move away from $x = 8$ the linear approximation is a line and so will always have the same slope while the function's slope will change as x changes and so the function will, in all likelihood, move away from the linear approximation.

Here's a quick sketch of the function and its linear approximation at $x = 8$.



As noted above, the farther from $x = 8$ we get the more distance separates the function itself and its linear approximation.

Linear approximations do a very good job of approximating values of $f(x)$ as long as we stay "near" $x = a$. However, the farther away from $x = a$ we get the worse the approximation is liable to be. The main problem here is that how near we need to stay to $x = a$ in order to get a good approximation will depend upon both the function we're using and the value of $x = a$ that we're using. Also, there will often be no easy way of predicting how far away from $x = a$ we can get and still have a "good" approximation.

Let's take a look at another example that is actually used fairly heavily in some places.

Example 2

Determine the linear approximation for $\sin(\theta)$ at $\theta = 0$.

Solution

Again, there really isn't a whole lot to this example. All that we need to do is compute the tangent line to $\sin\theta$ at $\theta = 0$.

$$f(\theta) = \sin(\theta) \qquad\qquad f'(\theta) = \cos(\theta)$$
$$f(0) = 0 \qquad\qquad f'(0) = 1$$

The linear approximation is,

$$\begin{aligned} L(\theta) &= f(0) + f'(0)(\theta - a) \\ &= 0 + (1)(\theta - 0) \\ &= \theta \end{aligned}$$

So, as long as θ stays small we can say that $\sin(\theta) \approx \theta$.

This is actually a somewhat important linear approximation. In optics this linear approximation is often used to simplify formulas. This linear approximation is also used to help describe the motion of a pendulum and vibrations in a string.

4.12 Differentials

In this section we're going to introduce a notation that we'll be seeing quite a bit in the next chapter. We will also look at an application of this new notation.

Given a function $y = f(x)$ we call dy and dx differentials and the relationship between them is given by,

$$dy = f'(x)\,dx$$

Note that if we are just given $f(x)$ then the differentials are df and dx and we compute them in the same manner.

$$df = f'(x)\,dx$$

Let's compute a couple of differentials.

Example 1

Compute the differential for each of the following.

(a) $y = t^3 - 4t^2 + 7t$

(b) $w = x^2 \sin(2x)$

(c) $f(z) = \mathbf{e}^{3-z^4}$

Solution

Before working any of these we should first discuss just what we're being asked to find here. We defined two differentials earlier and here we're being asked to compute a differential.

So, which differential are we being asked to compute? In this kind of problem we're being asked to compute the differential of the function. In other words, dy for the first problem, dw for the second problem and df for the third problem.

Here are the solutions. Not much to do here other than take a derivative and don't forget to add on the second differential to the derivative.

(a) $dy = \left(3t^2 - 8t + 7\right)dt$

(b) $dw = \left(2x\sin(2x) + 2x^2\cos(2x)\right)dx$

(c) $df = -4z^3\mathbf{e}^{3-z^4}\,dz$

There is a nice application to differentials. If we think of Δx as the change in x then $\Delta y = f(x + \Delta x) - f(x)$ is the change in y corresponding to the change in x. Now, if Δx is small we can assume that $\Delta y \approx dy$. Let's see an illustration of this idea.

Example 2

Compute dy and Δy if $y = \cos\left(x^2 + 1\right) - x$ as x changes from $x = 2$ to $x = 2.03$.

Solution

First let's compute actual the change in y, Δy.

$$\Delta y = \cos\left((2.03)^2 + 1\right) - 2.03 - \left(\cos\left(2^2 + 1\right) - 2\right) = 0.083581127$$

Now let's get the formula for dy.

$$dy = \left(-2x\sin\left(x^2 + 1\right) - 1\right) dx$$

Next, the change in x from $x = 2$ to $x = 2.03$ is $\Delta x = 0.03$ and so we then assume that $dx \approx \Delta x = 0.03$. This gives an approximate change in y of,

$$dy = \left(-2(2)\sin\left(2^2 + 1\right) - 1\right)(0.03) = 0.085070913$$

We can see that in fact we do have that $\Delta y \approx dy$ provided we keep Δx small.

We can use the fact that $\Delta y \approx dy$ in the following way.

Example 3

A sphere was measured and its radius was found to be 45 inches with a possible error of no more that 0.01 inches. What is the maximum possible error in the volume if we use this value of the radius?

Solution

First, recall the equation for the volume of a sphere.

$$V = \frac{4}{3}\pi r^3$$

Now, if we start with $r = 45$ and use $dr \approx \Delta r = 0.01$ then $\Delta V \approx dV$ should give us maximum error.

So, first get the formula for the differential.

$$dV = 4\pi r^2 dr$$

Now compute dV.

$$\Delta V \approx dV = 4\pi(45)^2 (0.01) = 254.47 \text{ in}^3$$

The maximum error in the volume is then approximately 254.47 in^3.

Be careful to not assume this is a large error. On the surface it looks large, however if we compute the actual volume for $r = 45$ we get $V = 381,703.51$ in^3. So, in comparison the error in the volume is,

$$\frac{254.47}{381703.51} \times 100 = 0.067\%$$

That's not much possible error at all!

4.13 Newton's Method

The next application that we'll take a look at in this chapter is an important application that is used in many areas. If you've been following along in the chapter to this point it's quite possible that you've gotten the impression that many of the applications that we've looked at are just made up by us to make you work. This is unfortunate because all of the applications that we've looked at to this point are real applications that really are used in real situations. The problem is often that in order to work more meaningful examples of the applications we would need more knowledge than we generally have about the science and/or physics behind the problem. Without that knowledge we're stuck doing some fairly simplistic examples that often don't seem very realistic at all and that makes it hard to understand that the application we're looking at is a real application.

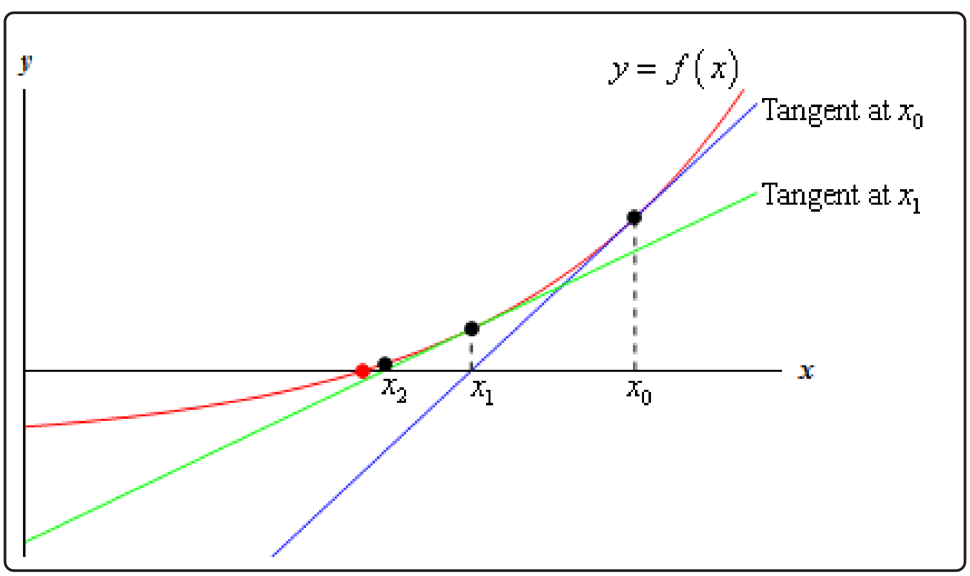
That is going to change in this section. This is an application that we can all understand and we can all understand needs to be done on occasion even if we don't understand the physics/science behind an actual application.

In this section we are going to look at a method for approximating solutions to equations. We all know that equations need to be solved on occasion and in fact we've solved quite a few equations ourselves to this point. In all the examples we've looked at to this point we were able to actually find the solutions, but it's not always possible to do that exactly and/or do the work by hand. That is where this application comes into play. So, let's see what this application is all about.

Let's suppose that we want to approximate the solution to $f(x) = 0$ and let's also suppose that we have somehow found an initial approximation to this solution say, x_0. This initial approximation is probably not all that good, in fact it may be nothing more than a quick guess we made, and so we'd like to find a better approximation. This is easy enough to do. First, we will get the tangent line to $f(x)$ at x_0.

$$y = f(x_0) + f'(x_0)(x - x_0)$$

Now, take a look at the graph below.



The blue line (if you're reading this in color anyway...) is the tangent line at x_0. We can see that this line will cross the x-axis much closer to the actual solution to the equation than x_0 does. Let's call this point where the tangent at x_0 crosses the x-axis x_1 and we'll use this point as our new approximation to the solution.

So, how do we find this point? Well we know it's coordinates, $(x_1, 0)$, and we know that it's on the tangent line so plug this point into the tangent line and solve for x_1 as follows,

$$\begin{aligned} 0 &= f(x_0) + f'(x_0)(x_1 - x_0) \\ x_1 - x_0 &= -\frac{f(x_0)}{f'(x_0)} \\ x_1 &= x_0 - \frac{f(x_0)}{f'(x_0)} \end{aligned}$$

So, we can find the new approximation provided the derivative isn't zero at the original approximation.

Now we repeat the whole process to find an even better approximation. We form up the tangent line to $f(x)$ at x_1 and use its root, which we'll call x_2, as a new approximation to the actual solution. If we do this we will arrive at the following formula.

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

This point is also shown on the graph above and we can see from this graph that if we continue following this process will get a sequence of numbers that are getting very close the actual solution. This process is called Newton's Method.

Here is the general Newton's Method

Newton's Method

If x_n is an approximation a solution of $f(x) = 0$ and if $f'(x_n) \neq 0$ the next approximation is given by,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

This should lead to the question of when do we stop? How many times do we go through this process? One of the more common stopping points in the process is to continue until two successive approximations agree to a given number of decimal places.

Before working any examples we should address two issues. First, we really do need to be solving $f(x) = 0$ in order for Newton's Method to be applied. This isn't really all that much of an issue but we do need to make sure that the equation is in this form prior to using the method.

Secondly, we do need to somehow get our hands on an initial approximation to the solution (*i.e.* we need x_0 somehow). One of the more common ways of getting our hands on x_0 is to sketch the graph of the function and use that to get an estimate of the solution which we then use as x_0. Another common method is if we know that there is a solution to a function in an interval then we can use the midpoint of the interval as x_0.

Let's work an example of Newton's Method.

Example 1

Use Newton's Method to determine an approximation to the solution to $\cos(x) = x$ that lies in the interval $[0, 2]$. Find the approximation to six decimal places.

Solution

First note that we weren't given an initial guess. We were however, given an interval in which to look. We will use this to get our initial guess. As noted above the general rule of thumb in these cases is to take the initial approximation to be the midpoint of the interval. So, we'll use $x_0 = 1$ as our initial guess.

Next, recall that we **must** have the function in the form $f(x) = 0$. Therefore, we first rewrite the equation as,

$$\cos(x) - x = 0$$

We can now write down the general formula for Newton's Method. Doing this will often simplify up the work a little so it's generally not a bad idea to do this.

$$x_{n+1} = x_n - \frac{\cos(x_n) - x_n}{-\sin(x_n) - 1}$$

Let's now get the first approximation.

$$x_1 = 1 - \frac{\cos(1) - 1}{-\sin(1) - 1} = 0.7503638679$$

At this point we should point out that the phrase "six decimal places" does not mean just get x_1 to six decimal places and then stop. Instead it means that we continue until two successive approximations agree to six decimal places.

Given that stopping condition we clearly need to go at least one step farther.

$$x_2 = 0.7503638679 - \frac{\cos(0.7503638679) - 0.7503638679}{-\sin(0.7503638679) - 1} = 0.7391128909$$

Alright, we're making progress. We've got the approximation to 1 decimal place. Let's do another one, leaving the details of the computation to you.

$$x_3 = 0.7390851334$$

We've got it to three decimal places. We'll need another one.

$$x_4 = 0.7390851332$$

And now we've got two approximations that agree to 9 decimal places and so we can stop. We will assume that the solution is approximately $x_4 = 0.7390851332$.

In this last example we saw that we didn't have to do too many computations in order for Newton's Method to give us an approximation in the desired range of accuracy. This will not always be the case. Sometimes it will take many iterations through the process to get to the desired accuracy and on occasion it can fail completely.

The following example is a little silly but it makes the point about the method failing.

Example 2

Use $x_0 = 1$ to find the approximation to the solution to $\sqrt[3]{x} = 0$.

Solution

Yes, it's a silly example. Clearly the solution is $x = 0$, but it does make a very important point. Let's get the general formula for Newton's method.

$$x_{n+1} = x_n - \frac{{x_n}^{\frac{1}{3}}}{\frac{1}{3}{x_n}^{-\frac{2}{3}}} = x_n - 3x_n = -2x_n$$

In fact, we don't really need to do any computations here. These computations get farther and farther away from the solution, $x = 0$,with each iteration. Here are a couple of computations to make the point.

$$\begin{aligned} x_1 &= -2 \\ x_2 &= 4 \\ x_3 &= -8 \\ x_4 &= 16 \\ &\ \ etc. \end{aligned}$$

So, in this case the method fails and fails spectacularly.

So, we need to be a little careful with Newton's method. It will usually quickly find an approximation to an equation. However, there are times when it will take a lot of work or when it won't work at all.

4.14 Business Applications

In the final section of this chapter let's take a look at some applications of derivatives in the business world. For the most part these are really applications that we've already looked at, but they are now going to be approached with an eye towards the business world.

Let's start things out with a couple of optimization problems. We've already looked at more than a few of these in previous sections so there really isn't anything all that new here except for the fact that they are coming out of the business world.

Example 1

An apartment complex has 250 apartments to rent. If they rent x apartments then their monthly profit, in dollars, is given by,

$$P(x) = -8x^2 + 3200x - 80,000$$

How many apartments should they rent in order to maximize their profit?

Solution

All that we're really being asked to do here is to maximize the profit subject to the constraint that x must be in the range $0 \le x \le 250$.

First, we'll need the derivative and the critical point(s) that fall in the range $0 \le x \le 250$.

$$P'(x) = -16x + 3200 \qquad \Rightarrow \qquad 3200 - 16x = 0 \qquad ,\Rightarrow \qquad x = \frac{3200}{16} = 200$$

Since the profit function is continuous and we have an interval with finite bounds we can find the maximum value by simply plugging in the only critical point that we have (which nicely enough in the range of acceptable answers) and the end points of the range.

$$P(0) = -80,000 \qquad P(200) = 240,000 \qquad P(250) = 220,000$$

So, it looks like they will generate the most profit if they only rent out 200 of the apartments instead of all 250 of them.

Note that with these problems you shouldn't just assume that renting all the apartments will generate the most profit. Do not forget that there are all sorts of maintenance costs and that the more tenants renting apartments the more the maintenance costs will be. With this analysis we can see that, for this complex at least, something probably needs to be done to get the maximum profit more towards full capacity. This kind of analysis can help them determine just what they need to do to move towards that goal whether it be raising rent or finding a way to reduce maintenance costs.

Note as well that because most apartment complexes have at least a few units empty after a tenant moves out and the like that it's possible that they would actually like the maximum profit to fall slightly under full capacity to take this into account. Again, another reason to not just assume that maximum profit will always be at the upper limit of the range.

Let's take a quick look at another problem along these lines.

Example 2

A production facility is capable of producing 60,000 widgets in a day and the total daily cost of producing x widgets in a day is given by,

$$C\left(x\right) = 250,000 + 0.08x + \frac{200,000,000}{x}$$

How many widgets per day should they produce in order to minimize production costs?

Solution

Here we need to minimize the cost subject to the constraint that x must be in the range $0 \le x \le 60,000$. Note that in this case the cost function is not continuous at the left endpoint and so we won't be able to just plug critical points and endpoints into the cost function to find the minimum value.

Let's get the first couple of derivatives of the cost function.

$$C'\left(x\right) = 0.08 - \frac{200,000,000}{x^2} \qquad\qquad C''\left(x\right) = \frac{400,000,000}{x^3}$$

The critical points of the cost function are,

$$\begin{aligned} 0.08 - \frac{200,000,000}{x^2} &= 0 \\ 0.08x^2 &= 200,000,000 \\ x^2 &= 2,500,000,000 \qquad \Rightarrow \qquad x = \pm\sqrt{2,500,000,000} = \pm 50,000 \end{aligned}$$

Now, clearly the negative value doesn't make any sense in this setting and so we have a single critical point in the range of possible solutions : 50,000.

Now, as long as $x > 0$ the second derivative is positive and so, in the range of possible solutions the function is always concave up and so producing 50,000 widgets will yield the absolute minimum production cost.

Recall from the Optimization section we discussed how we can use the second derivative to identity the absolute extrema even though all we really get from it is relative extrema.

Now, we shouldn't walk out of the previous two examples with the idea that the only applications to business are just applications we've already looked at but with a business "twist" to them.

There are some very real applications to calculus that are in the business world and at some level that is the point of this section. Note that to really learn these applications and all of their intricacies you'll need to take a business course or two or three. In this section we're just going to scratch the surface and get a feel for some of the actual applications of calculus from the business world and some of the main "buzz" words in the applications.

Let's start off by looking at the following example.

Example 3

The production costs per week for producing x widgets is given by,

$$C\left(x\right) = 500 + 350x - 0.09x^2, \quad 0 \le x \le 1000$$

Answer each of the following questions.

(a) What is the cost to produce the 301^{st} widget?

(b) What is the rate of change of the cost at $x = 300$?

Solution

(a) What is the cost to produce the 301^{st} widget?

We can't just compute $C\left(301\right)$ as that is the cost of producing 301 widgets while we are looking for the actual cost of producing the 301^{st} widget. In other words, what we're looking for here is,

$$C\left(301\right) - C\left(300\right) = 97,695.91 - 97,400.00 = 295.91$$

So, the cost of producing the 301^{st} widget is \$295.91.

(b) What is the rate of change of the cost at $x = 300$?

In this part all we need to do is get the derivative and then compute $C'\left(300\right)$.

$$C'\left(x\right) = 350 - 0.18x \qquad \Rightarrow \qquad C'\left(300\right) = 296.00$$

Okay, so just what did we learn in this example? The cost to produce an additional item is called the **marginal cost** and as we've seen in the above example the marginal cost is approximated by the rate of change of the **cost function**, $C\left(x\right)$. So, we define the **marginal cost function** to be the derivative of the cost function or, $C'(x)$. Let's work a quick example of this.

Example 4

The production costs per day for some widget is given by,

$$C(x) = 2500 - 10x - 0.01x^2 + 0.0002x^3$$

What is the marginal cost when $x = 200$, $x = 300$ and $x = 400$?

Solution

So, we need the derivative and then we'll need to compute some values of the derivative.

$$C'(x) = -10 - 0.02x + 0.0006x^2$$
$$C'(200) = 10 \qquad C'(300) = 38 \qquad C'(400) = 78$$

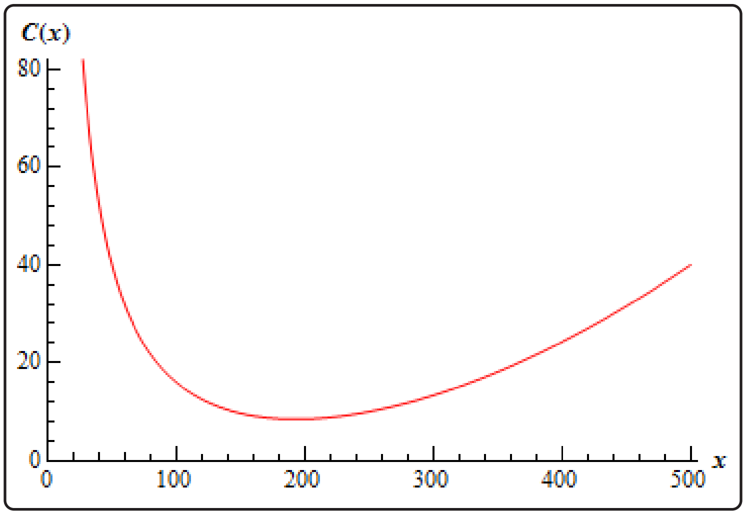
So, in order to produce the 201^{st} widget it will cost approximately \$10. To produce the 301^{st} widget will cost around \$38. Finally, to product the 401^{st} widget it will cost approximately \$78.

Note that it is important to note that $C'(n)$ is the approximate cost of producing the $(n+1)^{\text{st}}$ item and NOT the n^{th} item as it may seem to imply!

Let's now turn our attention to the **average cost** function. If $C(x)$ is the cost function for some item then the average cost function is,

$$\overline{C}(x) = \frac{C(x)}{x}$$

Here is the sketch of the average cost function from Example 4 above.



We can see from this that the average cost function has an absolute minimum. We can also see that this absolute minimum will occur at a critical point when $\overline{C}'(x) = 0$ since it clearly will have a horizontal tangent there.

Now, we could get the average cost function, differentiate that and then find the critical point. However, this average cost function is fairly typical for average cost functions so let's instead differentiate the general formula above using the quotient rule and see what we have.

$$\overline{C}'(x) = \frac{x\,C'(x) - C(x)}{x^2}$$

Now, as we noted above the absolute minimum will occur when $\overline{C}'(x) = 0$ and this will in turn occur when,

$$x\,C'(x) - C(x) = 0 \qquad \Rightarrow \qquad C'(x) = \frac{C(x)}{x} = \overline{C}(x)$$

So, we can see that it looks like for a typical average cost function we will get the minimum average cost when the marginal cost is equal to the average cost.

We should note however that not all average cost functions will look like this and so you shouldn't assume that this will always be the case.

Let's now move onto the revenue and profit functions. First, let's suppose that the price that some item can be sold at if there is a demand for x units is given by $p(x)$. This function is typically called either the **demand function** or the **price function**.

The **revenue function** is then how much money is made by selling x items and is,

$$R(x) = x\,p(x)$$

The **profit function** is then,

$$P(x) = R(x) - C(x) = x\,p(x) - C(x)$$

Be careful to not confuse the demand function, $p(x)$ - lower case p, and the profit function, $P(x)$ - upper case P. Bad notation maybe, but there it is.

Finally, the **marginal revenue function** is $R'(x)$ and the **marginal profit function** is $P'(x)$ and these represent the revenue and profit respectively if one more unit is sold.

Let's take a quick look at an example of using these.

Example 5

The weekly cost to produce x widgets is given by

$$C(x) = 75,000 + 100x - 0.03x^2 + 0.000004x^3 \qquad 0 \le x \le 10000$$

and the demand function for the widgets is given by,

$$p(x) = 200 - 0.005x \qquad 0 \le x \le 10000$$

Determine the marginal cost, marginal revenue and marginal profit when 2500 widgets are sold and when 7500 widgets are sold. Assume that the company sells exactly what they produce.

Solution

Okay, the first thing we need to do is get all the various functions that we'll need. Here are the revenue and profit functions.

$$\begin{aligned} R(x) &= x(200 - 0.005x) = 200x - 0.005x^2 \\ P(x) &= 200x - 0.005x^2 - \left(75,000 + 100x - 0.03x^2 + 0.000004x^3\right) \\ &= -75,000 + 100x + 0.025x^2 - 0.000004x^3 \end{aligned}$$

Now, all the marginal functions are,

$$\begin{aligned} C'(x) &= 100 - 0.06x + 0.000012x^2 \\ R'(x) &= 200 - 0.01x \\ P'(x) &= 100 + 0.05x - 0.000012x^2 \end{aligned}$$

The marginal functions when 2500 widgets are sold are,

$$C'(2500) = 25 \qquad R'(2500) = 175 \qquad P'(2500) = 150$$

The marginal functions when 7500 are sold are,

$$C'(7500) = 325 \qquad R'(7500) = 125 \qquad P'(7500) = -200$$

So, upon producing and selling the 2501^{st} widget it will cost the company approximately $25 to produce the widget and they will see an added $175 in revenue and $150 in profit.

On the other hand, when they produce and sell the 7501^{st} widget it will cost an additional $325 and they will receive an extra $125 in revenue, but lose $200 in profit.

We'll close this section out with a brief discussion on maximizing the profit. If we assume that the maximum profit will occur at a critical point such that $P'(x) = 0$ we can then say the following,

$$P'(x) = R'(x) - C'(x) = 0 \qquad \Rightarrow \qquad R'(x) = C'(x)$$

We then will know that this will be a maximum we also were to know that the profit was always concave down or,

$$P''(x) = R''(x) - C''(x) < 0 \qquad \Rightarrow \qquad R''(x) < C''(x)$$

So, if we know that $R''(x) < C''(x)$ then we will maximize the profit if $R'(x) = C'(x)$ or if the marginal cost equals the marginal revenue.

In this section we took a brief look at some of the ideas in the business world that involve calculus. Again, it needs to be stressed however that there is a lot more going on here and to really see how these applications are done you should really take some business courses. The point of this section was to just give a few ideas on how calculus is used in a field other than the sciences.

5 Integrals

In this chapter we will be looking at the third and final major topic that will be covered in a typical first Calculus course, integrals. As with derivatives this chapter will be devoted almost exclusively to finding and computing integrals. Applications will be given in the following chapter. There are really two types of integrals that we'll be looking at in this chapter : Indefinite Integrals and Definite Integrals. The first half of this chapter is devoted to indefinite integrals and the last half is devoted to definite integrals.

As we investigate indefinite integrals we will see that as long as we understand basic differentiation we shouldn't have a lot of problems with basic indefinite integrals. The reason for this is that indefinite integration is basically "undoing" differentiation. In fact, indefinite integrals are sometimes called anti-derivatives to make this idea clear. Having said that however we will be using the phrase indefinite integral instead of anti-derivative as that is the more common phrase used.

We will also spend a fair amount of time learning the substitution rule for integrals. We will see that it is really just "undoing" the chain rule and so, again, if you understand the chain rule it will help when using the substitution rule. In addition, as we'll see as we go through the rest of the calculus course the substitution rule will come up time and again and so it is very important to make sure that we have that down so we don't have issues with it in later topics.

As we move over to investigating definite integrals we will quickly realize just how important it is to be able to do indefinite integrals. As we will see we will not be able to compute definite integrals unless we can fist compute indefinite integrals.

We will also take a look at an important interpretation of definite integrals. Namely, a definite integral can be interpreted as the net area between the graph of the function and the x-axis.

5.1 Indefinite Integrals

In the past two chapters we've been given a function, $f(x)$, and asking what the derivative of this function was. Starting with this section we are now going to turn things around. We now want to ask what function we differentiated to get the function $f(x)$.

Let's take a quick look at an example to get us started.

Example 1

What function did we differentiate to get the following function.

$$f(x) = x^4 + 3x - 9$$

Solution

Let's actually start by getting the derivative of this function to help us see how we're going to have to approach this problem. The derivative of this function is,

$$f'(x) = 4x^3 + 3$$

The point of this was to remind us of how differentiation works. When differentiating powers of x we multiply the term by the original exponent and then drop the exponent by one.

Now, let's go back and work the problem. In fact, let's just start with the first term. We got x^4 by differentiating a function and since we drop the exponent by one it looks like we must have differentiated x^5. However, if we had differentiated x^5 we would have $5x^4$ and we don't have a 5 in front our first term, so the 5 needs to cancel out after we've differentiated. It looks then like we would have to differentiate $\frac{1}{5}x^5$ in order to get x^4.

Likewise, for the second term, in order to get $3x$ after differentiating we would have to differentiate $\frac{3}{2}x^2$. Again, the fraction is there to cancel out the 2 we pick up in the differentiation.

The third term is just a constant and we know that if we differentiate x we get 1. So, it looks like we had to differentiate $-9x$ to get the last term.

Putting all of this together gives the following function,

$$F(x) = \frac{1}{5}x^5 + \frac{3}{2}x^2 - 9x$$

Our answer is easy enough to check. Simply differentiate $F(x)$.

$$F'(x) = x^4 + 3x - 9 = f(x)$$

So, it looks like we got the correct function. Or did we? We know that the derivative of a constant is zero and so any of the following will also give $f(x)$ upon differentiating.

$$\begin{aligned} F(x) &= \frac{1}{5}x^5 + \frac{3}{2}x^2 - 9x + 10 \\ F(x) &= \frac{1}{5}x^5 + \frac{3}{2}x^2 - 9x - 1954 \\ F(x) &= \frac{1}{5}x^5 + \frac{3}{2}x^2 - 9x + \frac{3469}{123} \\ &\textit{etc.} \end{aligned}$$

In fact, any function of the form,

$$F(x) = \frac{1}{5}x^5 + \frac{3}{2}x^2 - 9x + c, \qquad c \text{ is a constant}$$

will give $f(x)$ upon differentiating.

There were two points to this last example. The first point was to get you thinking about how to do these problems. It is important initially to remember that we are really just asking what we differentiated to get the given function.

The other point is to recognize that there are actually an infinite number of functions that we could use and they will all differ by a constant.

Now that we've worked an example let's get some of the definitions and terminology out of the way.

Definitions

Given a function, $f(x)$, an **anti-derivative** of $f(x)$ is any function $F(x)$ such that

$$F'(x) = f(x)$$

If $F(x)$ is any anti-derivative of $f(x)$ then the most general anti-derivative of $f(x)$ is called an **indefinite integral** and denoted,

$$\int f(x)\,dx = F(x) + c, \qquad c \text{ is an arbitrary constant}$$

In this definition the $\int$ is called the **integral symbol**, $f(x)$ is called the **integrand**, x is called the **integration variable** and the "c" is called the **constant of integration**.

Note that often we will just say integral instead of indefinite integral (or definite integral for that matter when we get to those). It will be clear from the context of the problem that we are talking about an indefinite integral (or definite integral).

The process of finding the indefinite integral is called **integration** or **integrating** $f(x)$. If we need to be specific about the integration variable we will say that we are **integrating** $f(x)$ **with respect to** x.

Let's rework the first problem in light of the new terminology.

Example 2

Evaluate the following indefinite integral.

$$\int x^4 + 3x - 9\,dx$$

Solution

Since this is really asking for the most general anti-derivative we just need to reuse the final answer from the first example.

The indefinite integral is,

$$\int x^4 + 3x - 9\,dx = \frac{1}{5}x^5 + \frac{3}{2}x^2 - 9x + c$$

A couple of warnings are now in order. One of the more common mistakes that students make with integrals (both indefinite and definite) is to drop the dx at the end of the integral. This is required! Think of the integral sign and the dx as a set of parentheses. You already know and are probably quite comfortable with the idea that every time you open a parenthesis you must close it. With integrals, think of the integral sign as an "open parenthesis" and the *dx* as a "close parenthesis".

If you drop the *dx* it won't be clear where the integrand ends. Consider the following variations of the above example.

$$\begin{aligned}
\int x^4 + 3x - 9\,dx &= \frac{1}{5}x^5 + \frac{3}{2}x^2 - 9x + c \\
\int x^4 + 3x\,dx - 9 &= \frac{1}{5}x^5 + \frac{3}{2}x^2 + c - 9 \\
\int x^4\,dx + 3x - 9 &= \frac{1}{5}x^5 + c + 3x - 9
\end{aligned}$$

You only integrate what is between the integral sign and the dx. Each of the above integrals end in a different place and so we get different answers because we integrate a different number of terms each time. In the second integral the "-9" is outside the integral and so is left alone and not integrated. Likewise, in the third integral the "$3x-9$" is outside the integral and so is left alone.

Knowing which terms to integrate is not the only reason for writing the dx down. In the Substitution Rule section we will actually be working with the dx in the problem and if we aren't in the habit of writing it down it will be easy to forget about it and then we will get the wrong answer at that stage.

The moral of this is to make sure and put in the dx! At this stage it may seem like a silly thing to do, but it just needs to be there, if for no other reason than knowing where the integral stops.

On a side note, the dx notation should seem a little familiar to you. We saw things like this a couple of sections ago. We called the dx a differential in that section and yes that is exactly what it is. The dx that ends the integral is nothing more than a differential.

The next topic that we should discuss here is the integration variable used in the integral. Actually, there isn't really a lot to discuss here other than to note that the integration variable doesn't really matter. For instance,

$$\int x^4 + 3x - 9\,dx = \frac{1}{5}x^5 + \frac{3}{2}x^2 - 9x + c$$
$$\int t^4 + 3t - 9\,dt = \frac{1}{5}t^5 + \frac{3}{2}t^2 - 9t + c$$
$$\int w^4 + 3w - 9\,dw = \frac{1}{5}w^5 + \frac{3}{2}w^2 - 9w + c$$

Changing the integration variable in the integral simply changes the variable in the answer. It is important to notice however that when we change the integration variable in the integral we also changed the differential (dx, dt, or dw) to match the new variable. This is more important than we might realize at this point.

Another use of the differential at the end of integral is to tell us what variable we are integrating with respect to. At this stage that may seem unimportant since most of the integrals that we're going to be working with here will only involve a single variable. However, if you are on a degree track that will take you into multi-variable calculus this will be very important at that stage since there will be more than one variable in the problem. You need to get into the habit of writing the correct differential at the end of the integral so when it becomes important in those classes you will already be in the habit of writing it down.

To see why this is important take a look at the following two integrals.

$$\int 2x\,dx \qquad\qquad \int 2t\,dx$$

The first integral is simple enough.

$$\int 2x\,dx = x^2 + c$$

The second integral is also fairly simple, but we need to be careful. The dx tells us that we are integrating x's. That means that we only integrate x's that are in the integrand and all other variables in the integrand are considered to be constants. The second integral is then,

$$\int 2t\,dx = 2tx + c$$

So, it may seem silly to always put in the dx, but it is a vital bit of notation that can cause us to get the incorrect answer if we neglect to put it in.

Now, there are some important properties of integrals that we should take a look at.

Properties of the Indefinite Integral

1. $\displaystyle\int k\,f(x)\,dx = k\int f(x)\,dx$ where k is any number.

   So, we can factor multiplicative constants out of indefinite integrals. See the Proof of Various Integral Formulas section of the Extras appendix to see the proof of this property.

2. $\displaystyle\int -f(x)\,dx = -\int f(x)\,dx$.

   This is really the first property with $k = -1$ and so no proof of this property will be given.

3. $\displaystyle\int f(x) \pm g(x)\,dx = \int f(x)\,dx \pm \int g(x)\,dx$.

   In other words, the integral of a sum or difference of functions is the sum or difference of the individual integrals. This rule can be extended to as many functions as we need. See the Proof of Various Integral Formulas section of the Extras appendix to see the proof of this property.

Notice that when we worked the first example above we used the first and third property in the discussion. We integrated each term individually, put any constants back in and then put everything back together with the appropriate sign.

Not listed in the properties above were integrals of products and quotients. The reason for this is simple. Just like with derivatives each of the following will NOT work.

$$\int f(x)\,g(x)\,dx \neq \int f(x)\,dx \int g(x)\,dx \qquad\qquad \int \frac{f(x)}{g(x)}\,dx \neq \frac{\int f(x)\,dx}{\int g(x)\,dx}$$

With derivatives we had a product rule and a quotient rule to deal with these cases. However, with integrals there are no such rules. When faced with a product and quotient in an integral we will have a variety of ways of dealing with it depending on just what the integrand is.

There is one final topic to be discussed briefly in this section. On occasion we will be given $f'(x)$ and will ask what $f(x)$ was. We can now answer this question easily with an indefinite integral.

$$f(x) = \int f'(x)\,dx$$

Example 3

If $f'(x) = x^4 + 3x - 9$ what was $f(x)$?

Solution

By this point in this section this is a simple question to answer.

$$f(x) = \int f'(x)\, dx = \int x^4 + 3x - 9\, dx = \frac{1}{5}x^5 + \frac{3}{2}x^2 - 9x + c$$

In this section we kept evaluating the same indefinite integral in all of our examples. The point of this section was not to do indefinite integrals, but instead to get us familiar with the notation and some of the basic ideas and properties of indefinite integrals. The next couple of sections are devoted to actually evaluating indefinite integrals.

5.2 Computing Indefinite Integrals

In the previous section we started looking at indefinite integrals and in that section we concentrated almost exclusively on notation, concepts and properties of the indefinite integral. In this section we need to start thinking about how we actually compute indefinite integrals. We'll start off with some of the basic indefinite integrals.

The first integral that we'll look at is the integral of a power of x.

$$\int x^n \, dx = \frac{x^{n+1}}{n+1} + c, \quad n \neq -1$$

The general rule when integrating a power of x we add one onto the exponent and then divide by the new exponent. It is clear (hopefully) that we will need to avoid $n = -1$ in this formula. If we allow $n = -1$ in this formula we will end up with division by zero. We will take care of this case in a bit.

Next is one of the easier integrals but always seems to cause problems for people.

$$\int k \, dx = kx + c, \qquad c \text{ and } k \text{ are constants}$$

If you remember that all we're asking is what did we differentiate to get the integrand this is pretty simple, but it does seem to cause problems on occasion.

Let's now take a look at the trig functions.

$$\int \sin(x) \, dx = -\cos(x) + c \qquad \int \cos(x) \, dx = \sin(x) + c$$

$$\int \sec^2(x) \, dx = \tan(x) + c \qquad \int \sec(x)\tan(x) \, dx = \sec(x) + c$$

$$\int \csc^2(x) \, dx = -\cot(x) + c \qquad \int \csc(x)\cot(x) \, dx = -\csc(x) + c$$

Notice that we only integrated two of the six trig functions here. The remaining four integrals are really integrals that give the remaining four trig functions. Also, be careful with signs here. It is easy to get the signs for derivatives and integrals mixed up. Again, remember that we're asking what function we differentiated to get the integrand.

We will be able to integrate the remaining four trig functions in a couple of sections, but they all require the Substitution Rule

Now, let's take care of exponential and logarithm functions.

$$\int \mathbf{e}^x \, dx = \mathbf{e}^x + c \qquad \int a^x \, dx = \frac{a^x}{\ln(a)} + c \qquad \int \frac{1}{x} \, dx = \int x^{-1} \, dx = \ln|x| + c$$

Integrating logarithms requires a topic that is usually taught in Calculus II and so we won't be integrating a logarithm in this class. Also note the third integrand can be written in a couple of ways and don't forget the absolute value bars in the x in the answer to the third integral.

Finally, let's take care of the inverse trig and hyperbolic functions.

$$\int \frac{1}{x^2+1}\,dx = \tan^{-1}(x) + c \qquad \int \frac{1}{\sqrt{1-x^2}}\,dx = \sin^{-1}(x) + c$$

$$\int \sinh(x)\,dx = \cosh(x) + c \qquad \int \cosh(x)\,dx = \sinh(x) + c$$

$$\int \operatorname{sech}^2(x)\,dx = \tanh(x) + c \qquad \int \operatorname{sech}(x)\tanh(x)\,dx = -\operatorname{sech}(x) + c$$

$$\int \operatorname{csch}^2(x)\,dx = -\coth(x) + c \qquad \int \operatorname{csch}(x)\coth(x)\,dx = -\operatorname{csch}(x) + c$$

As with logarithms integrating inverse trig functions requires a topic usually taught in Calculus II and so we won't be integrating them in this class. There is also a different answer for the second integral above. Recalling that since all we are asking here is what function did we differentiate to get the integrand the second integral could also be,

$$\int \frac{1}{\sqrt{1-x^2}}\,dx = -\cos^{-1}(x) + c$$

Traditionally we use the first form of this integral.

Okay, now that we've got most of the basic integrals out of the way let's do some indefinite integrals. In all of these problems remember that we can always check our answer by differentiating and making sure that we get the integrand.

Example 1

Evaluate each of the following indefinite integrals.

(a) $\displaystyle\int 5t^3 - 10t^{-6} + 4\,dt$

(b) $\displaystyle\int x^8 + x^{-8}\,dx$

(c) $\displaystyle\int 3\sqrt[4]{x^3} + \frac{7}{x^5} + \frac{1}{6\sqrt{x}}\,dx$

(d) $\displaystyle\int dy$

(e) $\displaystyle\int \left(w + \sqrt[3]{w}\right)\left(4 - w^2\right)dw$

(f) $\displaystyle\int \frac{4x^{10} - 2x^4 + 15x^2}{x^3}\,dx$

Solution

Okay, in all of these remember the basic rules of indefinite integrals. First, to integrate sums and differences all we really do is integrate the individual terms and then put the terms back together with the appropriate signs. Next, we can ignore any coefficients until we are done with integrating that particular term and then put the coefficient back in. Also, do not forget the "+c" at the end it is important and must be there.

So, let's evaluate some integrals.

(a) $\displaystyle\int 5t^3 - 10t^{-6} + 4\,dt$

There's not really a whole lot to do here other than use the first two formulas from the beginning of this section. Remember that when integrating powers (that aren't -1 of course) we just add one onto the exponent and then divide by the new exponent.

$$\begin{aligned}\int 5t^3 - 10t^{-6} + 4\,dt &= 5\left(\frac{1}{4}\right)t^4 - 10\left(\frac{1}{-5}\right)t^{-5} + 4t + c\\ &= \frac{5}{4}t^4 + 2t^{-5} + 4t + c\end{aligned}$$

Be careful when integrating negative exponents. Remember to add one onto the exponent. One of the more common mistakes that students make when integrating negative exponents is to "add one" and end up with an exponent of "-7" instead of the correct exponent of "-5".

(b) $\displaystyle\int x^8 + x^{-8}\,dx$

This is here just to make sure we get the point about integrating negative exponents.

$$\int x^8 + x^{-8}\,dx = \frac{1}{9}x^9 - \frac{1}{7}x^{-7} + c$$

(c) $\displaystyle\int 3\sqrt[4]{x^3} + \frac{7}{x^5} + \frac{1}{6\sqrt{x}}\,dx$

In this case there isn't a formula for explicitly dealing with radicals or rational expressions. However, just like with derivatives we can write all these terms so they are in the numerator and they all have an exponent. This should always be your first step when faced with this kind of integral just as it was when differentiating.

$$\begin{aligned}\int 3\sqrt[4]{x^3} + \frac{7}{x^5} + \frac{1}{6\sqrt{x}}\,dx &= \int 3x^{\frac{3}{4}} + 7x^{-5} + \frac{1}{6}x^{-\frac{1}{2}}\,dx\\ &= 3\frac{1}{{}^7\!/_4}x^{\frac{7}{4}} - \frac{7}{4}x^{-4} + \frac{1}{6}\left(\frac{1}{{}^1\!/_2}\right)x^{\frac{1}{2}} + c\\ &= \frac{12}{7}x^{\frac{7}{4}} - \frac{7}{4}x^{-4} + \frac{1}{3}x^{\frac{1}{2}} + c\end{aligned}$$

When dealing with fractional exponents we usually don't "divide by the new exponent". Doing this is equivalent to multiplying by the reciprocal of the new exponent and so that is what we will usually do.

(d) $\int dy$

Don't make this one harder than it is...

$$\int dy = \int 1\, dy = y + c$$

In this case we are really just integrating a one!

(e) $\int \left(w + \sqrt[3]{w}\right)\left(4 - w^2\right) dw$

We've got a product here and as we noted in the previous section there is no rule for dealing with products. However, in this case we don't need a rule. All that we need to do is multiply things out (taking care of the radicals at the same time of course) and then we will be able to integrate.

$$\begin{aligned}\int \left(w + \sqrt[3]{w}\right)\left(4 - w^2\right)\, dw &= \int 4w - w^3 + 4w^{\frac{1}{3}} - w^{\frac{7}{3}} dw \\ &= 2w^2 - \frac{1}{4}w^4 + 3w^{\frac{4}{3}} - \frac{3}{10}w^{\frac{10}{3}} + c\end{aligned}$$

(f) $\int \frac{4x^{10} - 2x^4 + 15x^2}{x^3}\, dx$

As with the previous part it's not really a problem that we don't have a rule for quotients for this integral. In this case all we need to do is break up the quotient and then integrate the individual terms.

$$\begin{aligned}\int \frac{4x^{10} - 2x^4 + 15x^2}{x^3}\, dx &= \int \frac{4x^{10}}{x^3} - \frac{2x^4}{x^3} + \frac{15x^2}{x^3} dx \\ &= \int 4x^7 - 2x + \frac{15}{x} dx \\ &= \frac{1}{2}x^8 - x^2 + 15 \ln|x| + c\end{aligned}$$

Be careful to not think of the third term as x to a power for the purposes of integration. Using that rule on the third term will NOT work. The third term is simply a logarithm. Also, don't get excited about the 15. The 15 is just a constant and so it can be factored out of the integral. In other words, here is what we did to integrate the third term.

$$\int \frac{15}{x}\, dx = 15 \int \frac{1}{x}\, dx = 15 \ln|x| + c$$

Always remember that you can't integrate products and quotients in the same way that we integrate sums and differences. At this point the only way to integrate products and quotients is to multiply the product out or break up the quotient. Eventually we'll see some other products and quotients that can be dealt with in other ways. However, there will never be a single rule that will work for all products and there will never be a single rule that will work for all quotients. Every product and quotient is different and will need to be worked on a case by case basis.

The first set of examples focused almost exclusively on powers of x (or whatever variable we used in each example). It's time to do some examples that involve other functions.

Example 2

Evaluate each of the following integrals.

(a) $\displaystyle\int 3\mathbf{e}^x + 5\cos(x) - 10\sec^2(x)\,dx$

(b) $\displaystyle\int 2\sec(w)\tan(w) + \frac{1}{6w}\,dw$

(c) $\displaystyle\int \frac{23}{y^2+1} + 6\csc(y)\cot(y) + \frac{9}{y}\,dy$

(d) $\displaystyle\int \frac{3}{\sqrt{1-x^2}} + 6\sin(x) + 10\sinh(x)\,dx$

(e) $\displaystyle\int \frac{7-6\sin^2(\theta)}{\sin^2(\theta)}\,d\theta$

Solution

Most of the problems in this example will simply use the formulas from the beginning of this section. More complicated problems involving most of these functions will need to wait until we reach the Substitution Rule.

(a) $\displaystyle\int 3\mathbf{e}^x + 5\cos(x) - 10\sec^2(x)\,dx$

There isn't anything to this one other than using the formulas.

$$\int 3\mathbf{e}^x + 5\cos(x) - 10\sec^2(x)\,dx = 3\mathbf{e}^x + 5\sin(x) - 10\tan(x) + c$$

(b) $\displaystyle\int 2\sec(w)\tan(w) + \frac{1}{6w}\,dw$

Let's be a little careful with this one. First break it up into two integrals and note the rewritten integrand on the second integral.

$$\begin{aligned}\int 2\sec(w)\tan(w) + \frac{1}{6w}\,dw &= \int 2\sec(w)\tan(w)\,dw + \int \frac{1}{6}\frac{1}{w}\,dw \\ &= \int 2\sec(w)\tan(w)\,dw + \frac{1}{6}\int \frac{1}{w}\,dw\end{aligned}$$

Rewriting the second integrand will help a little with the integration at this early stage. We can think of the 6 in the denominator as a $\frac{1}{6}$ out in front of the term and then since this is a constant it can be factored out of the integral. The answer is then,

$$\int 2\sec(w)\tan(w) + \frac{1}{6w}\,dw = 2\sec(w) + \frac{1}{6}\ln|w| + c$$

Note that we didn't factor the 2 out of the first integral as we factored the $\frac{1}{6}$ out of the second. In fact, we will generally not factor the $\frac{1}{6}$ out either in later problems. It was only done here to make sure that you could follow what we were doing.

(c) $\displaystyle\int \frac{23}{y^2+1} + 6\csc(y)\cot(y) + \frac{9}{y}\,dy$

In this one we'll just use the formulas from above and don't get excited about the coefficients. They are just multiplicative constants and so can be ignored while we integrate each term and then once we're done integrating a given term we simply put the coefficients back in.

$$\int \frac{23}{y^2+1} + 6\csc(y)\cot(y) + \frac{9}{y}\,dy = 23\tan^{-1}(y) - 6\csc(y) + 9\ln|y| + c$$

(d) $\displaystyle\int \frac{3}{\sqrt{1-x^2}} + 6\sin(x) + 10\sinh(x)\,dx$

Again, there really isn't a whole lot to do with this one other than to use the appropriate formula from above while taking care of coefficients.

$$\int \frac{3}{\sqrt{1-x^2}} + 6\sin(x) + 10\sinh(x)\,dx = 3\sin^{-1}(x) - 6\cos(x) + 10\cosh(x) + c$$

(e) $\displaystyle\int \frac{7-6\sin^2(\theta)}{\sin^2(\theta)}\,d\theta$

This one can be a little tricky if you aren't ready for it. As discussed previously, at this point the only way we have of dealing with quotients is to break it up.

$$\begin{aligned}\int \frac{7-6\sin^2(\theta)}{\sin^2(\theta)}\,d\theta &= \int \frac{7}{\sin^2(\theta)} - 6\,d\theta \\ &= \int 7\csc^2(\theta) - 6\,d\theta\end{aligned}$$

Notice that upon breaking the integral up we further simplified the integrand by recalling the definition of cosecant. With this simplification we can do the integral.

$$\int \frac{7-6\sin^2(\theta)}{\sin^2(\theta)}\,d\theta = -7\cot(\theta) - 6\theta + c$$

As shown in the last part of this example we can do some fairly complicated looking quotients at this point if we remember to do simplifications when we see them. In fact, this is something that you should always keep in mind. In almost any problem that we're doing here don't forget to simplify where possible. In almost every case this can only help the problem and will rarely complicate the problem.

In the next problem we're going to take a look at a product and this time we're not going to be able to just multiply the product out. However, if we recall the comment about simplifying a little this problem becomes fairly simple.

Example 3

Integrate $\displaystyle\int \sin\left(\frac{t}{2}\right)\cos\left(\frac{t}{2}\right)\,dt$.

Solution

There are several ways to do this integral and most of them require the next section. However, there is a way to do this integral using only the material from this section. All that is required is to remember the trig formula that we can use to simplify the integrand up a little. Recall the following double angle formula.

$$\sin(2t) = 2\sin(t)\cos(t)$$

A small rewrite of this formula gives,

$$\sin(t)\cos(t) = \frac{1}{2}\sin(2t)$$

If we now replace all the t's with $\frac{t}{2}$ we get,

$$\sin\left(\frac{t}{2}\right)\cos\left(\frac{t}{2}\right) = \frac{1}{2}\sin(t)$$

Using this formula, the integral becomes something we can do.

$$\begin{aligned}\int \sin\left(\frac{t}{2}\right)\cos\left(\frac{t}{2}\right)\,dt &= \int \frac{1}{2}\sin(t)\,dt \\ &= -\frac{1}{2}\cos(t) + c\end{aligned}$$

As noted earlier there is another method for doing this integral. In fact, there are two alternate methods. To see all three check out the section on Constant of Integrationin the Extras appendix but be aware that the other two do require the material covered in the next section.

The formula/simplification in the previous problem is a nice "trick" to remember. It can be used on occasion to greatly simplify some problems.

There is one more set of examples that we should do before moving out of this section.

Example 4

Given the following information determine the function $f(x)$.

(a) $f'(x) = 4x^3 - 9 + 2\sin(x) + 7\mathbf{e}^x, \ f(0) = 15$

(b) $f''(x) = 15\sqrt{x} + 5x^3 + 6, \quad f(1) = -\frac{5}{4}, \ f(4) = 404$

In both of these we will need to remember that

$$f(x) = \int f'(x)\,dx$$

Also note that because we are giving values of the function at specific points we are also going to be determining what the constant of integration will be in these problems.

Solution

(a) $f'(x) = 4x^3 - 9 + 2\sin(x) + 7\mathbf{e}^x, \ f(0) = 15$

The first step here is integrating to determine the most general possible $f(x)$.

$$\begin{aligned} f(x) &= \int 4x^3 - 9 + 2\sin(x) + 7\mathbf{e}^x \, dx \\ &= x^4 - 9x - 2\cos(x) + 7\mathbf{e}^x + c \end{aligned}$$

Now we have a value of the function so let's plug in $x = 0$ and determine the value of the constant of integration c.

$$\begin{aligned} 15 = f(0) &= 0^4 - 9(0) - 2\cos(0) + 7\mathbf{e}^0 + c \\ &= -2 + 7 + c \\ &= 5 + c \end{aligned}$$

So, from this it looks like $c = 10$. This means that the function is,

$$f(x) = x^4 - 9x - 2\cos(x) + 7\mathbf{e}^x + 10$$

(b) $f''(x) = 15\sqrt{x} + 5x^3 + 6, \quad f(1) = -\frac{5}{4}, \quad f(4) = 404$

This one is a little different form the first one. In order to get the function we will need the first derivative and we have the second derivative. We can however, use an integral to get the first derivative from the second derivative, just as we used an integral to get the function from the first derivative.

So, let's first get the most general possible first derivative by integrating the second derivative.

$$\begin{aligned} f'(x) &= \int f''(x) \, dx \\ &= \int 15x^{\frac{1}{2}} + 5x^3 + 6 \, dx \\ &= 15\left(\frac{2}{3}\right)x^{\frac{3}{2}} + \frac{5}{4}x^4 + 6x + c \\ &= 10x^{\frac{3}{2}} + \frac{5}{4}x^4 + 6x + c \end{aligned}$$

Don't forget the constant of integration!

We can now find the most general possible function by integrating the first derivative which we found above.

$$\begin{aligned} f(x) &= \int 10x^{\frac{3}{2}} + \frac{5}{4}x^4 + 6x + c \, dx \\ &= 4x^{\frac{5}{2}} + \frac{1}{4}x^5 + 3x^2 + cx + d \end{aligned}$$

Do not get excited about integrating the c. It's just a constant and we know how to integrate constants. Also, there will be no reason to think the constants of integration from the integration in each step will be the same and so we'll need to call each constant of integration something different, d in this case.

Now, plug in the two values of the function that we've got.

$$-\frac{5}{4} = f(1) = 4 + \frac{1}{4} + 3 + c + d = \frac{29}{4} + c + d$$
$$404 = f(4) = 4(32) + \frac{1}{4}(1024) + 3(16) + c(4) + d = 432 + 4c + d$$

This gives us a system of two equations in two unknowns that we can solve.

$$\begin{aligned} -\frac{5}{4} &= \frac{29}{4} + c + d \\ 404 &= 432 + 4c + d \end{aligned} \qquad \Rightarrow \qquad \begin{aligned} c &= -\frac{13}{2} \\ d &= -2 \end{aligned}$$

The function is then,

$$f(x) = 4x^{\frac{5}{2}} + \frac{1}{4}x^5 + 3x^2 - \frac{13}{2}x - 2$$

Don't remember how to solve systems? Check out the Solving Systems portion of the Algebra/Trig Review.

In this section we've started the process of integration. We've seen how to do quite a few basic integrals and we also saw a quick application of integrals in the last example.

There are many new formulas in this section that we'll now have to know. However, if you think about it, they aren't really new formulas. They are really nothing more than derivative formulas that we should already know written in terms of integrals. If you remember that you should find it easier to remember the formulas in this section.

Always remember that integration is asking nothing more than what function did we differentiate to get the integrand. If you can remember that many of the basic integrals that we saw in this section and many of the integrals in the coming sections aren't too bad.

5.3 Substitution Rule for Indefinite Integrals

After the last section we now know how to do the following integrals.

$$\int \sqrt[4]{x}\,dx \qquad \int \frac{1}{t^3}\,dt \qquad \int \cos(w)\,dw \qquad \int \mathbf{e}^y\,dy$$

All of the integrals we've done to this point have required that we just had an x, or a t, or a w, *etc.* and not more complicated terms such as,

$$\int 18x^2\,\sqrt[4]{6x^3+5}\,dx \qquad \int \frac{2t^3+1}{\left(t^4+2t\right)^3}\,dt$$
$$\int \left(1-\frac{1}{w}\right)\cos\left(w-\ln w\right)\,dw \qquad \int (8y-1)\,\mathbf{e}^{4y^2-y}\,dy$$

All of these look considerably more difficult than the first set. However, they aren't too bad once you see how to do them. Let's start with the first one.

$$\int 18x^2\,\sqrt[4]{6x^3+5}\,dx$$

In this case let's notice that if we let

$$u = 6x^3+5$$

and we compute the differential (you remember how to compute these right?) for this we get,

$$du = 18x^2dx$$

Now, let's go back to our integral and notice that we can eliminate every x that exists in the integral and write the integral completely in terms of u using both the definition of u and its differential.

$$\begin{aligned}\int 18x^2\,\sqrt[4]{6x^3+5}\,dx &= \int \left(6x^3+5\right)^{\frac{1}{4}}\,\left(18x^2dx\right)\\ &= \int u^{\frac{1}{4}}\,du\end{aligned}$$

In the process of doing this we've taken an integral that looked very difficult and with a quick substitution we were able to rewrite the integral into a very simple integral that we can do.

Evaluating the integral gives,

$$\int 18x^2\,\sqrt[4]{6x^3+5}\,dx = \int u^{\frac{1}{4}}\,du = \frac{4}{5}u^{\frac{5}{4}}+c = \frac{4}{5}\left(6x^3+5\right)^{\frac{5}{4}}+c$$

As always, we can check our answer with a quick derivative if we'd like to and don't forget to "back substitute" and get the integral back into terms of the original variable.

What we've done in the work above is called the **Substitution Rule**. Here is the substitution rule in general.

Substitution Rule

$$\int f\left(g\left(x\right)\right)\,g'\left(x\right)\,dx = \int f\left(u\right)\,du, \qquad \text{where, } u = g\left(x\right)$$

A natural question at this stage is how to identify the correct substitution. Unfortunately, the answer is it depends on the integral. However, there is a general rule of thumb that will work for many of the integrals that we're going to be running across.

When faced with an integral we'll ask ourselves what we know how to integrate. With the integral above we can quickly recognize that we know how to integrate

$$\int \sqrt[4]{x}\,dx$$

However, we didn't have just the root we also had stuff in front of the root and (more importantly in this case) stuff under the root. Since we can only integrate roots if there is just an x under the root a good first guess for the substitution is then to make u be the stuff under the root.

Another way to think of this is to ask yourself if you were to differentiate the integrand (we're not of course, but just for a second pretend that we were) is there a chain rule and what is the inside function for the chain rule. If there is a chain rule (for a derivative) then there is a pretty good chance that the inside function will be the substitution that will allow us to do the integral.

We will have to be careful however. There are times when using this general rule can get us in trouble or overly complicate the problem. We'll eventually see problems where there are more than one "inside function" and/or integrals that will look very similar and yet use completely different substitutions. The reality is that the only way to really learn how to do substitutions is to just work lots of problems and eventually you'll start to get a feel for how these work and you'll find it easier and easier to identify the proper substitutions.

Now, with that out of the way we should ask the following question. How, do we know if we got the correct substitution? Well, upon computing the differential and actually performing the substitution every x in the integral (including the x in the dx) must disappear in the substitution process and the only letters left should be u's (including a du) and we should be left with an integral that we can do.

If there are x's left over or we have an integral that cannot be evaluated then there is a pretty good chance that we chose the wrong substitution. Unfortunately, however there is at least one case (we'll be seeing an example of this in the next section) where the correct substitution will actually leave some x's and we'll need to do a little more work to get it to work.

Again, it cannot be stressed enough at this point that the only way to really learn how to do substitutions is to just work lots of problems. There are lots of different kinds of problems and after working many problems you'll start to get a real feel for these problems and after you work enough of them you'll reach the point where you'll be able to do simple substitutions in your head without having to actually write anything down.

As a final note we should point out that often (in fact in almost every case) the differential will not appear exactly in the integrand as it did in the example above and sometimes we'll need to do some manipulation of the integrand and/or the differential to get all the x's to disappear in the substitution.

Let's work some examples so we can get a better idea on how the substitution rule works.

Example 1

Evaluate each of the following integrals.

(a) $\displaystyle\int \left(1 - \frac{1}{w}\right) \cos\left(w - \ln(w)\right) dw$

(b) $\displaystyle\int 3\left(8y - 1\right) \mathbf{e}^{4y^2 - y}\, dy$

(c) $\displaystyle\int x^2 \left(3 - 10x^3\right)^4 dx$

(d) $\displaystyle\int \frac{x}{\sqrt{1 - 4x^2}}\, dx$

Solution

(a) $\displaystyle\int \left(1 - \frac{1}{w}\right) \cos\left(w - \ln(w)\right) dw$

In this case it looks like we have a cosine with an inside function and so let's use that as the substitution.

$$u = w - \ln(w) \qquad\qquad du = \left(1 - \frac{1}{w}\right) dw$$

So, as with the first example we worked the stuff in front of the cosine appears exactly in the differential. The integral is then,

$$\begin{aligned}\int \left(1 - \frac{1}{w}\right) \cos\left(w - \ln(w)\right) dw &= \int \cos\left(u\right)\, du \\ &= \sin\left(u\right) + c \\ &= \sin\left(w - \ln(w)\right) + c\end{aligned}$$

Don't forget to go back to the original variable in the problem.

(b) $\displaystyle\int 3\,(8y-1)\,\mathbf{e}^{4y^2-y}\,dy$

Again, it looks like we have an exponential function with an inside function (*i.e.* the exponent) and it looks like the substitution should be,

$$u = 4y^2 - y \qquad\qquad du = (8y-1)\,dy$$

Now, with the exception of the 3 the stuff in front of the exponential appears exactly in the differential. Recall however that we can factor the 3 out of the integral and so it won't cause any problems. The integral is then,

$$\begin{aligned}\int 3\,(8y-1)\,\mathbf{e}^{4y^2-y}\,dy &= 3\int \mathbf{e}^u\,du\\ &= 3\mathbf{e}^u + c\\ &= 3\mathbf{e}^{4y^2-y} + c\end{aligned}$$

(c) $\displaystyle\int x^2\left(3-10x^3\right)^4\,dx$

In this case it looks like the following should be the substitution.

$$u = 3 - 10x^3 \qquad\qquad du = -30x^2\,dx$$

Okay, now we have a small problem. We've got an x^2 out in front of the parenthesis but we don't have a "-30". This is not really the problem it might appear to be at first. We will simply rewrite the differential as follows.

$$x^2\,dx = -\frac{1}{30}\,du$$

With this we can now substitute the $x^2\,dx$ away. In the process we will pick up a constant, but that isn't a problem since it can always be factored out of the integral.

We can now do the integral.

$$\begin{aligned}\int x^2\left(3-10x^3\right)^4\,dx &= \int \left(3-10x^3\right)^4 x^2\,dx\\ &= \int u^4\left(-\frac{1}{30}\right)\,du\\ &= -\frac{1}{30}\left(\frac{1}{5}\right)u^5 + c\\ &= -\frac{1}{150}\left(3-10x^3\right)^5 + c\end{aligned}$$

Note that in most problems when we pick up a constant as we did in this example we will generally factor it out of the integral in the same step that we substitute it in.

(d) $\displaystyle\int \frac{x}{\sqrt{1-4x^2}}\,dx$

In this example don't forget to bring the root up to the numerator and change it into fractional exponent form. Upon doing this we can see that the substitution is,

$$u = 1 - 4x^2 \qquad du = -8x\,dx \qquad \Rightarrow \qquad xdx = -\frac{1}{8}du$$

The integral is then,

$$\begin{aligned}\int \frac{x}{\sqrt{1-4x^2}}\,dx &= \int x\left(1-4x^2\right)^{-\frac{1}{2}}\,dx\\ &= -\frac{1}{8}\int u^{-\frac{1}{2}}\,du\\ &= -\frac{1}{4}u^{\frac{1}{2}} + c\\ &= -\frac{1}{4}\left(1-4x^2\right)^{\frac{1}{2}} + c\end{aligned}$$

In the previous set of examples the substitution was generally pretty clear. There was exactly one term that had an "inside function" and so there wasn't really much in the way of options for the substitution. Let's take a look at some more complicated problems to make sure we don't come to expect all substitutions are like those in the previous set of examples.

Example 2

Evaluate each of the following integrals.

(a) $\displaystyle\int \sin\left(1-x\right)\left(2-\cos\left(1-x\right)\right)^4\,dx$

(b) $\displaystyle\int \cos\left(3z\right)\sin^{10}\left(3z\right)\,dz$

(c) $\displaystyle\int \sec^2\left(4t\right)\left(3-\tan\left(4t\right)\right)^3\,dt$

Solution

(a) $\displaystyle\int \sin\left(1-x\right)\left(2-\cos\left(1-x\right)\right)^4\,dx$

In this problem there are two "inside functions". There is the $1-x$ that is inside the two trig functions and there is also the term that is raised to the 4^{th} power.

There are two ways to proceed with this problem. The first idea that many students have is substitute the $1-x$ away. There is nothing wrong with doing this but it doesn't eliminate the problem of the term to the 4^{th} power. That's still there and if we used this idea we would then need to do a second substitution to deal with that.

The second (and much easier) way of doing this problem is to just deal with the stuff raised to the 4^{th} power and see what we get. The substitution in this case would be,

$$u = 2 - \cos(1-x) \qquad du = -\sin(1-x)\,dx \qquad \Rightarrow \qquad \sin(1-x)\,dx = -du$$

Two things to note here. First, don't forget to correctly deal with the "$-$". A common mistake at this point is to lose it. Secondly, notice that the $1-x$ turns out to not really be a problem after all. Because the $1-x$ was "buried" in the substitution that we actually used it was also taken care of at the same time. The integral is then,

$$\begin{aligned}\int \sin(1-x)\left(2-\cos(1-x)\right)^4 dx &= -\int u^4\,du \\ &= -\frac{1}{5}u^5 + c \\ &= -\frac{1}{5}\left(2-\cos(1-x)\right)^5 + c\end{aligned}$$

As seen in this example sometimes there will seem to be two substitutions that will need to be done however, if one of them is buried inside of another substitution then we'll only really need to do one. Recognizing this can save a lot of time in working some of these problems.

(b) $\displaystyle\int \cos(3z)\sin^{10}(3z)\,dz$

This one is a little tricky at first. We can see the correct substitution by recalling that,

$$\sin^{10}(3z) = \left(\sin(3z)\right)^{10}$$

Using this it looks like the correct substitution is,

$$u = \sin(3z) \qquad du = 3\cos(3z)\,dz \qquad \Rightarrow \qquad \cos(3z)\,dz = \frac{1}{3}du$$

Notice that we again had two apparent substitutions in this integral but again the $3z$ is buried in the substitution we're using and so we didn't need to worry about it.

Here is the integral.

$$\begin{aligned}\int \cos(3z)\sin^{10}(3z)\,dz &= \frac{1}{3}\int u^{10}\,du \\ &= \frac{1}{3}\left(\frac{1}{11}\right)u^{11} + c \\ &= \frac{1}{33}\sin^{11}(3z) + c\end{aligned}$$

Note that the one third in front of the integral came about from the substitution on the differential and we just factored it out to the front of the integral. This is what we will usually do with these constants.

(c) $\displaystyle\int \sec^2(4t)\,(3-\tan(4t))^3\,dt$

In this case we've got a $4t$, a secant squared as well as a term cubed. However, it looks like if we use the following substitution the first two issues are going to be taken care of for us.

$$u = 3 - \tan(4t) \qquad du = -4\sec^2(4t)\,dt \qquad \Rightarrow \qquad \sec^2(4t)\,dt = -\frac{1}{4}du$$

The integral is now,

$$\begin{aligned}\int \sec^2(4t)\,(3-\tan(4t))^3\,dt &= -\frac{1}{4}\int u^3\,du\\ &= -\frac{1}{16}u^4 + c\\ &= -\frac{1}{16}(3-\tan(4t))^4 + c\end{aligned}$$

The most important thing to remember in substitution problems is that after the substitution all the original variables need to disappear from the integral. After the substitution the only variables that should be present in the integral should be the new variable from the substitution (usually u). Note as well that this includes the variables in the differential!

This next set of examples, while not particularly difficult, can cause trouble if we aren't paying attention to what we're doing.

Example 3

Evaluate each of the following integrals.

(a) $\displaystyle\int \frac{3}{5y+4}\,dy$

(b) $\displaystyle\int \frac{3y}{5y^2+4}\,dy$

(c) $\displaystyle\int \frac{3y}{(5y^2+4)^2}\,dy$

(d) $\displaystyle\int \frac{3}{5y^2+4}\,dy$

Solution

(a) $\displaystyle\int \frac{3}{5y+4}\,dy$

We haven't seen a problem quite like this one yet. Let's notice that if we take the denominator and differentiate it we get just a constant and the only thing that we have in the numerator is also a constant. This is a pretty good indication that we can use the denominator for our substitution so,

$$u = 5y + 4 \qquad du = 5\,dy \qquad \Rightarrow \qquad dy = \frac{1}{5}du$$

The integral is now,

$$\begin{aligned}\int \frac{3}{5y+4}\,dy &= \frac{3}{5}\int \frac{1}{u}\,du \\ &= \frac{3}{5}\ln|u| + c \\ &= \frac{3}{5}\ln|5y+4| + c\end{aligned}$$

Remember that we can just factor the 3 in the numerator out of the integral and that makes the integral a little clearer in this case.

(b) $\displaystyle\int \frac{3y}{5y^2+4}\,dy$

The integral is very similar to the previous one with a couple of minor differences but notice that again if we differentiate the denominator we get something that is different from the numerator by only a multiplicative constant. Therefore, we'll again take the denominator as our substitution.

$$u = 5y^2 + 4 \qquad du = 10y\,dy \qquad \Rightarrow \qquad y\,dy = \frac{1}{10}du$$

The integral is,

$$\begin{aligned}\int \frac{3y}{5y^2+4}\,dy &= \frac{3}{10}\int \frac{1}{u}\,du \\ &= \frac{3}{10}\ln|u| + c \\ &= \frac{3}{10}\ln\left|5y^2+4\right| + c\end{aligned}$$

(c) $\displaystyle\int \frac{3y}{(5y^2+4)^2}\,dy$

Now, this one is almost identical to the previous part except we added a power onto the denominator. Notice however that if we ignore the power and differentiate what's left we get the same thing as the previous example so we'll use the same substitution here.

$$u = 5y^2 + 4 \qquad du = 10y\,dy \qquad \Rightarrow \qquad y\,dy = \frac{1}{10}du$$

The integral in this case is,

$$\begin{aligned}\int \frac{3y}{(5y^2+4)^2}\,dy &= \frac{3}{10}\int u^{-2}\,du \\ &= -\frac{3}{10}u^{-1} + c \\ &= -\frac{3}{10}\left(5y^2+4\right)^{-1} + c = -\frac{3}{10\left(5y^2+4\right)} + c\end{aligned}$$

Be careful in this case to not turn this into a logarithm. After working problems like the first two in this set a common error is to turn every rational expression into a logarithm. If there is a power on the whole denominator then there is a good chance that it isn't a logarithm.

The idea that we used in the last three parts to determine the substitution is not a bad idea to remember. If we've got a rational expression try differentiating the denominator (ignoring any powers that are on the whole denominator) and if the result is the numerator or only differs from the numerator by a multiplicative constant then we can usually use that as our substitution.

(d) $\displaystyle\int \frac{3}{5y^2+4}\,dy$

Now, this part is completely different from the first three and yet seems similar to them as well. In this case if we differentiate the denominator we get a y that is not in the numerator and so we can't use the denominator as our substitution.

In fact, because we have y^2 in the denominator and no y in the numerator is an indication of how to work this problem. This integral is going to be an inverse tangent when we are done. The key to seeing this is to recall the following formula,

$$\int \frac{1}{1+u^2}\,du = \tan^{-1}u + c$$

We clearly don't have exactly this but we do have something that is similar. The denominator has a squared term plus a constant and the numerator is just a constant. So, with a little work and the proper substitution we should be able to get our integral into a form that will allow us to use this formula.

There is one part of this formula that is really important and that is the "$1+$" in the denominator. The "$1+$" must be there and we've got a "$4+$" but it is easy enough to take care of that. We'll just factor a 4 out of the denominator and at the same time we'll factor the 3 in the numerator out of the integral as well. Doing this gives,

$$\begin{aligned}\int \frac{3}{5y^2+4}\,dy &= \int \frac{3}{4\left(\frac{5y^2}{4}+1\right)}\,dy \\ &= \frac{3}{4}\int \frac{1}{\frac{5y^2}{4}+1}\,dy \\ &= \frac{3}{4}\int \frac{1}{\left(\frac{\sqrt{5}\,y}{2}\right)^2+1}\,dy\end{aligned}$$

Notice that in the last step we rewrote things a little in the denominator. This will help us to see what the substitution needs to be. In order to get this integral into the formula above we need to end up with a u^2 in the denominator. Our substitution will then need to be something that upon squaring gives us $\frac{5y^2}{4}$. With the rewrite we can see what that we'll need to use the following substitution.

$$u = \frac{\sqrt{5}y}{2} \qquad du = \frac{\sqrt{5}}{2}\,dy \qquad \Rightarrow \qquad dy = \frac{2}{\sqrt{5}}du$$

Don't get excited about the root in the substitution, these will show up on occasion. Upon plugging our substitution in we get,

$$\int \frac{3}{5y^2+4}\,dy = \frac{3}{4}\left(\frac{2}{\sqrt{5}}\right)\int \frac{1}{u^2+1}\,du$$

After doing the substitution, and factoring any constants out, we get exactly the integral that gives an inverse tangent and so we know that we did the correct substitution for this integral. The integral is then,

$$\begin{aligned}\int \frac{3}{5y^2+4}\,dy &= \frac{3}{2\sqrt{5}}\int \frac{1}{u^2+1}\,du \\ &= \frac{3}{2\sqrt{5}}\tan^{-1}(u)+c \\ &= \frac{3}{2\sqrt{5}}\tan^{-1}\left(\frac{\sqrt{5}\,y}{2}\right)+c\end{aligned}$$

In this last set of integrals we had four integrals that were similar to each other in many ways and yet all either yielded different answer using the same substitution or used a completely different substitution than one that was similar to it.

This is a fairly common occurrence and so you will need to be able to deal with these kinds of issues.

There are many integrals that on the surface look very similar and yet will use a completely different substitution or will yield a completely different answer when using the same substitution.

Let's take a look at another set of examples to give us more practice in recognizing these kinds of issues. Note however that we won't be putting as much detail into these as we did with the previous examples.

Example 4

Evaluate each of the following integrals.

(a) $\displaystyle\int \frac{2t^3+1}{\left(t^4+2t\right)^3}\,dt$

(b) $\displaystyle\int \frac{2t^3+1}{t^4+2t}\,dt$

(c) $\displaystyle\int \frac{x}{\sqrt{1-4x^2}}\,dx$

(d) $\displaystyle\int \frac{1}{\sqrt{1-4x^2}}\,dx$

Solution

(a) $\displaystyle\int \frac{2t^3+1}{\left(t^4+2t\right)^3}\,dt$

Clearly the derivative of the denominator, ignoring the exponent, differs from the numerator only by a multiplicative constant and so the substitution is,

$$u = t^4 + 2t \qquad du = \left(4t^3+2\right)dt = 2\left(2t^3+1\right)dt \qquad \Rightarrow \qquad \left(2t^3+1\right)dt = \frac{1}{2}du$$

After a little manipulation of the differential we get the following integral.

$$\begin{aligned}\int \frac{2t^3+1}{\left(t^4+2t\right)^3}\,dt &= \frac{1}{2}\int \frac{1}{u^3}\,du \\ &= \frac{1}{2}\int u^{-3}\,du \\ &= \frac{1}{2}\left(-\frac{1}{2}\right)u^{-2}+c \\ &= -\frac{1}{4}\left(t^4+2t\right)^{-2}+c\end{aligned}$$

(b) $\displaystyle\int \frac{2t^3+1}{t^4+2t}\,dt$

The only difference between this problem and the previous one is the denominator. In the previous problem the whole denominator is cubed and in this problem the denominator has no power on it. The same substitution will work in this problem but because we no longer have the power the problem will be different.

So, using the substitution from the previous example the integral is,

$$\begin{aligned}\int \frac{2t^3+1}{t^4+2t}\,dt &= \frac{1}{2}\int \frac{1}{u}\,du\\ &= \frac{1}{2}\ln|u| + c\\ &= \frac{1}{2}\ln\left|t^4+2t\right| + c\end{aligned}$$

So, in this case we get a logarithm from the integral.

(c) $\displaystyle\int \frac{x}{\sqrt{1-4x^2}}\,dx$

Here, if we ignore the root we can again see that the derivative of the stuff under the radical differs from the numerator by only a multiplicative constant and so we'll use that as the substitution.

$$u = 1-4x^2 \qquad du = -8x\,dx \qquad \Rightarrow \qquad x\,dx = -\frac{1}{8}du$$

The integral is then,

$$\begin{aligned}\int \frac{x}{\sqrt{1-4x^2}}\,dx &= -\frac{1}{8}\int u^{-\frac{1}{2}}\,du\\ &= -\frac{1}{8}(2)\,u^{\frac{1}{2}} + c\\ &= -\frac{1}{4}\sqrt{1-4x^2} + c\end{aligned}$$

(d) $\displaystyle\int \frac{1}{\sqrt{1-4x^2}}\,dx$

In this case we are missing the x in the numerator and so the substitution from the last part will do us no good here. This integral is another inverse trig function integral that is similar to the last part of the previous set of problems. In this case we need to following formula.

$$\int \frac{1}{\sqrt{1-u^2}}\,du = \sin^{-1}(u) + c$$

The integral in this problem is nearly this. The only difference is the presence of the coefficient of 4 on the x^2. With the correct substitution this can be dealt with however.

To see what this substitution should be let's rewrite the integral a little. We need to figure out what we squared to get $4x^2$ and that will be our substitution.

$$\int \frac{1}{\sqrt{1-4x^2}}\,dx = \int \frac{1}{\sqrt{1-(2x)^2}}\,dx$$

With this rewrite it looks like we can use the following substitution.

$$u = 2x \qquad du = 2dx \qquad \Rightarrow \qquad dx = \frac{1}{2}du$$

The integral is then,

$$\begin{aligned}\int \frac{1}{\sqrt{1-4x^2}}\,dx &= \frac{1}{2}\int \frac{1}{\sqrt{1-u^2}}\,du\\ &= \frac{1}{2}\sin^{-1}(u) + c\\ &= \frac{1}{2}\sin^{-1}(2x) + c\end{aligned}$$

Since this document is also being presented on the web we're going to put the rest of the substitution rule examples in the next section. With all the examples in one section the section was becoming too large for web presentation.

5.4 More Substitution Rule

In order to allow these pages to be displayed on the web we've broken the substitution rule examples into two sections. The previous section contains the introduction to the substitution rule and some fairly basic examples. The examples in this section tend towards the slightly more difficult side. Also, we'll not be putting quite as much explanation into the solutions here as we did in the previous section.

In the first couple of sets of problems in this section the difficulty is not with the actual integration itself, but with the set up for the integration. Most of the integrals are fairly simple and most of the substitutions are fairly simple. The problems arise in getting the integral set up properly for the substitution(s) to be done. Once you see how these are done it's easy to see what you have to do, but the first time through these can cause problems if you aren't on the lookout for potential problems.

Example 1

Evaluate each of the following integrals.

(a) $\displaystyle\int \mathbf{e}^{2t} + \sec(2t)\tan(2t)\, dt$

(b) $\displaystyle\int \sin(t)\left(4\cos^3(t) + 6\cos^2(t) - 8\right) dt$

(c) $\displaystyle\int x\cos\left(x^2+1\right) + \frac{x}{x^2+1}\, dx$

Solution

(a) $\displaystyle\int \mathbf{e}^{2t} + \sec(2t)\tan(2t)\, dt$

This first integral has two terms in it and both will require the same substitution. This means that we won't have to do anything special to the integral. One of the more common "mistakes" here is to break the integral up and do a separate substitution on each part. This isn't really mistake but will definitely increase the amount of work we'll need to do. So, since both terms in the integral use the same substitution we'll just do everything as a single integral using the following substitution.

$$u = 2t \qquad du = 2dt \qquad \Rightarrow \qquad dt = \frac{1}{2}du$$

The integral is then,

$$\begin{aligned}\int \mathbf{e}^{2t} + \sec(2t)\tan(2t)\,dt &= \frac{1}{2}\int \mathbf{e}^{u} + \sec(u)\tan(u)\,du\\ &= \frac{1}{2}\left(\mathbf{e}^{u} + \sec(u)\right) + c\\ &= \frac{1}{2}\left(\mathbf{e}^{2t} + \sec(2t)\right) + c\end{aligned}$$

Often a substitution can be used multiple times in an integral so don't get excited about that if it happens. Also note that since there was a $\frac{1}{2}$ in front of the whole integral there must also be a $\frac{1}{2}$ in front of the answer from the integral.

(b) $\displaystyle\int \sin(t)\left(4\cos^3(t) + 6\cos^2(t) - 8\right)\,dt$

This integral is similar to the previous one, but it might not look like it at first glance. Here is the substitution for this problem,

$$u = \cos(t) \qquad du = -\sin(t)\,dt \qquad \Rightarrow \qquad \sin(t)\,dt = -du$$

We'll plug the substitution into the problem twice (since there are two cosines) and will only work because there is a sine multiplying everything. Without that sine in front we would not be able to use this substitution.

The integral in this case is,

$$\begin{aligned}\int \sin(t)\left(4\cos^3(t) + 6\cos^2(t) - 8\right)\,dt &= -\int 4u^3 + 6u^2 - 8\,du\\ &= -\left(u^4 + 2u^3 - 8u\right) + c\\ &= -\left(\cos^4(t) + 2\cos^3(t) - 8\cos(t)\right) + c\end{aligned}$$

Again, be careful with the minus sign in front of the whole integral.

(c) $\displaystyle\int x\cos\left(x^2+1\right) + \frac{x}{x^2+1}\,dx$

It should be fairly clear that each term in this integral will use the same substitution, but let's rewrite things a little to make things really clear.

$$\int x\cos\left(x^2+1\right) + \frac{x}{x^2+1}\,dx = \int x\left(\cos\left(x^2+1\right) + \frac{1}{x^2+1}\right)\,dx$$

Since each term had an x in it and we'll need that for the differential we factored that out of both terms to get it into the front. This integral is now very similar to the previous one. Here's the substitution.

$$u = x^2 + 1 \qquad du = 2x\,dx \qquad \Rightarrow \qquad x\,dx = \frac{1}{2}du$$

The integral is,

$$\begin{aligned}\int x\cos\left(x^2+1\right)+\frac{x}{x^2+1}\,dx &= \frac{1}{2}\int \cos(u)+\frac{1}{u}\,du\\ &= \frac{1}{2}\left(\sin(u)+\ln|u|\right)+c\\ &= \frac{1}{2}\left(\sin\left(x^2+1\right)+\ln\left|x^2+1\right|\right)+c\end{aligned}$$

So, as we've seen in the previous set of examples sometimes we can use the same substitution more than once in an integral and doing so will simplify the work.

Example 2

Evaluate each of the following integrals.

(a) $\displaystyle\int x^2+\mathbf{e}^{1-x}\,dx$

(b) $\displaystyle\int x\cos\left(x^2+1\right)+\frac{1}{x^2+1}\,dx$

Solution

(a) $\displaystyle\int x^2+\mathbf{e}^{1-x}\,dx$

In this integral the first term does not need any substitution while the second term does need a substitution. So, to deal with that we'll need to split the integral up as follows,

$$\int x^2+\mathbf{e}^{1-x}\,dx=\int x^2\,dx+\int \mathbf{e}^{1-x}\,dx$$

The substitution for the second integral is then,

$$u=1-x \qquad du=-dx \qquad \Rightarrow \qquad dx=-du$$

The integral is,

$$\begin{aligned}\int x^2+\mathbf{e}^{1-x}\,dx &= \int x^2\,dx-\int \mathbf{e}^{u}\,du\\ &= \frac{1}{3}x^3-\mathbf{e}^{u}+c\\ &= \frac{1}{3}x^3-\mathbf{e}^{1-x}+c\end{aligned}$$

Be careful with this kind of integral. One of the more common mistakes here is do the following "shortcut".

$$\int x^2 + \mathbf{e}^{1-x}\,dx = -\int x^2 + \mathbf{e}^{u}\,du$$

In other words, some students will try do the substitution just the second term without breaking up the integral. There are two issues with this. First, there is a "$-$" in front of the whole integral that shouldn't be there. It should only be on the second term because that is the term getting the substitution. Secondly, and probably more importantly, there are x's in the integral and we have a du for the differential. We can't mix variables like this. When we do integrals all the variables in the integrand must match the variable in the differential.

(b) $\displaystyle\int x\cos\left(x^2+1\right) + \frac{1}{x^2+1}\,dx$

This integral looks very similar to Example 1c above, but it is different. In this integral we no longer have the x in the numerator of the second term and that means that the substitution we'll use for the first term will no longer work for the second term. In fact, the second term doesn't need a substitution at all since it is just an inverse tangent.

The substitution for the first term is then,

$$u = x^2+1 \qquad du = 2x\,dx \qquad \Rightarrow \qquad x\,dx = \frac{1}{2}du$$

Now let's do the integral. Remember to first break it up into two terms before using the substitution.

$$\begin{aligned}\int x\cos\left(x^2+1\right) + \frac{1}{x^2+1}\,dx &= \int x\cos\left(x^2+1\right)\,dx + \int \frac{1}{x^2+1}\,dx \\ &= \frac{1}{2}\int \cos(u)\,du + \int \frac{1}{x^2+1}\,dx \\ &= \frac{1}{2}\sin(u) + \tan^{-1}(x) + c \\ &= \frac{1}{2}\sin\left(x^2+1\right) + \tan^{-1}(x) + c\end{aligned}$$

In this set of examples we saw that sometimes one (or potentially more than one) term in the integrand will not require a substitution. In these cases we'll need to break up the integral into two integrals, one involving the terms that don't need a substitution and another with the term(s) that do need a substitution.

Example 3

Evaluate each of the following integrals.

(a) $\displaystyle\int \mathbf{e}^{-z} + \sec^2\left(\frac{z}{10}\right)\,dz$

(b) $\displaystyle\int \sin w\sqrt{1-2\cos w} + \frac{1}{7w+2}\,dw$

(c) $\displaystyle\int \frac{10x+3}{x^2+16}\,dx$

Solution

(a) $\displaystyle\int \mathbf{e}^{-z} + \sec^2\left(\frac{z}{10}\right)\,dz$

In this integral, unlike any integrals that we've yet done, there are two terms and each will require a different substitution. So, to do this integral we'll first need to split up the integral as follows,

$$\int \mathbf{e}^{-z} + \sec^2\left(\frac{z}{10}\right)\,dz = \int \mathbf{e}^{-z}dz + \int \sec^2\left(\frac{z}{10}\right)\,dz$$

Here are the substitutions for each integral.

$$\begin{array}{llll} u = -z & du = -dz & \Rightarrow & dz = -du \\ v = \dfrac{z}{10} & dv = \dfrac{1}{10}dz & \Rightarrow & dz = 10dv \end{array}$$

Notice that we used different letters for each substitution to avoid confusion when we go to plug back in for u and v.

Here is the integral.

$$\begin{aligned} \int \mathbf{e}^{-z} + \sec^2\left(\frac{z}{10}\right)\,dz &= -\int \mathbf{e}^{u}du + 10\int \sec^2(v)\,dv \\ &= -\mathbf{e}^{u} + 10\tan(v) + c \\ &= -\mathbf{e}^{-z} + 10\tan\left(\frac{z}{10}\right) + c \end{aligned}$$

(b) $\displaystyle\int \sin w\sqrt{1-2\cos w} + \frac{1}{7w+2}\,dw$

As with the last problem this integral will require two separate substitutions. Let's first break up the integral.

$$\int \sin w\sqrt{1-2\cos w} + \frac{1}{7w+2}\,dw = \int \sin w(1-2\cos w)^{\frac{1}{2}}\,dw + \int \frac{1}{7w+2}\,dw$$

Here are the substitutions for this integral.

$$\begin{array}{llll} u = 1-2\cos(w) & du = 2\sin(w)\,dw & \Rightarrow & \sin(w)\,dw = \frac{1}{2}du \\ v = 7w+2 & dv = 7dw & \Rightarrow & dw = \frac{1}{7}dv \end{array}$$

The integral is then,

$$\begin{aligned} \int \sin w\sqrt{1-2\cos w} + \frac{1}{7w+2}\,dw &= \frac{1}{2}\int u^{\frac{1}{2}}\,du + \frac{1}{7}\int \frac{1}{v}\,dv \\ &= \frac{1}{2}\left(\frac{2}{3}\right)u^{\frac{3}{2}} + \frac{1}{7}\ln|v| + c \\ &= \frac{1}{3}(1-2\cos w)^{\frac{3}{2}} + \frac{1}{7}\ln|7w+2| + c \end{aligned}$$

(c) $\displaystyle\int \frac{10x+3}{x^2+16}\,dx$

The last problem in this set can be tricky. If there was just an x in the numerator we could do a quick substitution to get a natural logarithm. Likewise, if there wasn't an x in the numerator we would get an inverse tangent after a quick substitution.

To get this integral into a form that we can work with we will first need to break it up as follows.

$$\begin{aligned} \int \frac{10x+3}{x^2+16}\,dx &= \int \frac{10x}{x^2+16}\,dx + \int \frac{3}{x^2+16}\,dx \\ &= \int \frac{10x}{x^2+16}\,dx + \frac{1}{16}\int \frac{3}{\frac{x^2}{16}+1}\,dx \end{aligned}$$

We now have two integrals each requiring a different substitution. The substitutions for each of the integrals above are,

$$\begin{array}{llll} u = x^2+16 & du = 2x\,dx & \Rightarrow & x\,dx = \frac{1}{2}du \\ v = \frac{x}{4} & dv = \frac{1}{4}dx & \Rightarrow & dx = 4dv \end{array}$$

The integral is then,

$$\begin{aligned}\int \frac{10x+3}{x^2+16}\,dx &= 5\int \frac{1}{u}\,du + \frac{3}{4}\int \frac{1}{v^2+1}\,dv \\ &= 5\ln|u| + \frac{3}{4}\tan^{-1}(v) + c \\ &= 5\ln\left|x^2+16\right| + \frac{3}{4}\tan^{-1}\left(\frac{x}{4}\right) + c\end{aligned}$$

We've now seen a set of integrals in which we need to do more than one substitution. In these cases we will need to break up the integral into separate integrals and do separate substitutions for each.

We now need to move onto a different set of examples that can be a little tricky. Once you've seen how to do these they aren't too bad but doing them for the first time can be difficult if you aren't ready for them.

Example 4

Evaluate each of the following integrals.

(a) $\displaystyle\int \tan(x)\,dx$

(b) $\displaystyle\int \sec(y)\,dy$

(c) $\displaystyle\int \frac{\cos\left(\sqrt{x}\right)}{\sqrt{x}}\,dx$

(d) $\displaystyle\int \mathbf{e}^{t+\mathbf{e}^t}\,dt$

(e) $\displaystyle\int 2x^3\sqrt{x^2+1}\,dx$

Solution

(a) $\displaystyle\int \tan(x)\,dx$

The first question about this problem is probably why is it here? Substitution rule problems generally require more than a single function. The key to this problem is to realize that there really are two functions here. All we need to do is remember the definition of tangent and we can write the integral as,

$$\int \tan(x)\,dx = \int \frac{\sin(x)}{\cos(x)}\,dx$$

Written in this way we can see that the following substitution will work for us,

$$u = \cos(x) \qquad du = -\sin(x)\,dx \qquad \Rightarrow \qquad \sin(x)\,dx = -du$$

The integral is then,

$$\begin{aligned}\int \tan(x)\,dx &= -\int \frac{1}{u}\,du\\ &= -\ln|u| + c\\ &= -\ln|\cos(x)| + c\end{aligned}$$

Now, while this is a perfectly serviceable answer that minus sign in front is liable to cause problems if we aren't careful. So, let's rewrite things a little. Recalling a property of logarithms we can move the minus sign in front to an exponent on the cosine and then do a little simplification.

$$\begin{aligned}\int \tan(x)\,dx &= -\ln|\cos(x)| + c\\ &= \ln|\cos(x)|^{-1} + c\\ &= \ln\frac{1}{|\cos(x)|} + c\\ &= \ln|\sec(x)| + c\end{aligned}$$

This is the formula that is typically given for the integral of tangent.

Note that we could integrate cotangent in a similar manner.

(b) $\int \sec(y)\,dy$

This problem also at first appears to not belong in the substitution rule problems. This is even more of a problem upon noticing that we can't just use the definition of the secant function to write this in a form that will allow the use of the substitution rule.

This problem is going to require a technique that isn't used terribly often at this level but is a useful technique to be aware of. Sometimes we can make an integral doable by multiplying the top and bottom by a common term. This will not always work and even when it does it is not always clear what we should multiply by but when it works it is very useful.

Here is how we'll use this idea for this problem.

$$\int \sec(y)\,dy = \int \frac{\sec(y)}{1}\frac{\left(\sec(y)+\tan(y)\right)}{\left(\sec(y)+\tan(y)\right)}\,dy$$

First, we will think of the secant as a fraction and then multiply the top and bottom of the fraction by the same term. It is probably not clear why one would want to do this here but doing this will actually allow us to use the substitution rule. To see how this will work let's simplify the integrand somewhat.

$$\int \sec(y)\,dy = \int \frac{\sec^2(y)+\tan(y)\sec(y)}{\sec(y)+\tan(y)}\,dy$$

We can now use the following substitution.

$$u = \sec(y)+\tan(y) \qquad du = \left(\sec(y)\tan(y)+\sec^2(y)\right)dy$$

The integral is then,

$$\begin{aligned}\int \sec(y)\,dy &= \int \frac{1}{u}\,du\\ &= \ln|u| + c\\ &= \ln\left|\sec(y)+\tan(y)\right| + c\end{aligned}$$

Sometimes multiplying the top and bottom of a fraction by a carefully chosen term will allow us to work a problem. It does however take some thought sometimes to determine just what the term should be.

We can use a similar process for integrating cosecant.

(c) $\displaystyle\int \frac{\cos\left(\sqrt{x}\right)}{\sqrt{x}}\,dx$

This next problem has a subtlety to it that can get us in trouble if we aren't paying attention. Because of the root in the cosine it makes some sense to use the following substitution.

$$u = x^{\frac{1}{2}} \qquad du = \frac{1}{2}x^{-\frac{1}{2}}\,dx$$

This is where we need to be careful. Upon rewriting the differential we get,

$$2du = \frac{1}{\sqrt{x}}dx$$

The root that is in the denominator will not become a u as we might have been tempted to do. Instead it will get taken care of in the differential.

The integral is,

$$\begin{aligned}\int \frac{\cos\left(\sqrt{x}\right)}{\sqrt{x}}\,dx &= 2\int \cos\left(u\right)\,du\\ &= 2\sin\left(u\right)+c\\ &= 2\sin\left(\sqrt{x}\right)+c\end{aligned}$$

(d) $\displaystyle\int \mathbf{e}^{t+\mathbf{e}^t}\,dt$

With this problem we need to very carefully pick our substitution. As the problem is written we might be tempted to use the following substitution,

$$u = t+\mathbf{e}^t \qquad du = \left(1+\mathbf{e}^t\right)dt$$

However, this won't work as you can probably see. The differential doesn't show up anywhere in the integrand and we just wouldn't be able to eliminate all the t's with this substitution.

In order to work this problem we will need to rewrite the integrand as follows,

$$\int \mathbf{e}^{t+\mathbf{e}^t}\,dt = \int \mathbf{e}^t\mathbf{e}^{\mathbf{e}^t}\,dt$$

We will now use the substitution,

$$u = \mathbf{e}^t \qquad du = \mathbf{e}^t dt$$

The integral is,

$$\begin{aligned}\int \mathbf{e}^{t+\mathbf{e}^t}\,dt &= \int \mathbf{e}^u\,du\\ &= \mathbf{e}^u + c\\ &= \mathbf{e}^{\mathbf{e}^t}+c\end{aligned}$$

Some substitutions can be really tricky to see and it's not unusual that you'll need to do some simplification and/or rewriting to get a substitution to work.

(e) $\displaystyle\int 2x^3\sqrt{x^2+1}\,dx$

This last problem in this set is different from all the other substitution problems that we've worked to this point. Given the fact that we've got more than an x under the root it makes sense that the substitution pretty much has to be,

$$u = x^2+1 \qquad du = 2x\,dx$$

At first glance it looks like this might not work for the substitution because we have an x^3 in front of the root. However, if we first rewrite $2x^3 = x^2\,(2x)$ we could then move the $2x$ to the end of the integral so at least the du will show up explicitly in the integral. Doing this gives the following,

$$\begin{aligned}\int 2x^3\sqrt{x^2+1}\,dx &= \int x^2\sqrt{x^2+1}\;(2x)\,dx \\ &= \int x^2 u^{\frac{1}{2}}\,du\end{aligned}$$

This is a real problem. Our integrals can't have two variables in them. Normally this would mean that we chose our substitution incorrectly. However, in this case we can rewrite the substitution as follows,

$$x^2 = u - 1$$

and now, we can eliminate the remaining x's from our integral. Doing this gives,

$$\begin{aligned}\int 2x^3\sqrt{x^2+1}\,dx &= \int (u-1)\,u^{\frac{1}{2}}\,du \\ &= \int u^{\frac{3}{2}} - u^{\frac{1}{2}}\,du \\ &= \frac{2}{5}u^{\frac{5}{2}} - \frac{2}{3}u^{\frac{3}{2}} + c \\ &= \frac{2}{5}\left(x^2+1\right)^{\frac{5}{2}} - \frac{2}{3}\left(x^2+1\right)^{\frac{3}{2}} + c\end{aligned}$$

Sometimes, we will need to use a substitution more than once.

This kind of problem doesn't arise all that often and when it does there will sometimes be alternate methods of doing the integral. However, it will often work out that the easiest method of doing the integral is to do what we just did here.

This final set of examples isn't too bad once you see the substitutions and that is the point with this set of problems. These all involve substitutions that we've not seen prior to this and so we need to see some of these kinds of problems.

Example 5

Evaluate each of the following integrals.

(a) $\displaystyle\int \frac{1}{x \ln(x)}\, dx$

(b) $\displaystyle\int \frac{\mathbf{e}^{2t}}{1+\mathbf{e}^{2t}}\, dt$

(c) $\displaystyle\int \frac{\mathbf{e}^{2t}}{1+\mathbf{e}^{4t}}\, dt$

(d) $\displaystyle\int \frac{\sin^{-1}(x)}{\sqrt{1-x^2}}\, dx$

Solution

(a) $\displaystyle\int \frac{1}{x \ln(x)}\, dx$

In this case we know that we can't integrate a logarithm by itself and so it makes some sense (hopefully) that the logarithm will need to be in the substitution. Here is the substitution for this problem.

$$u = \ln(x) \qquad\qquad du = \frac{1}{x}dx$$

So, the x in the denominator of the integrand will get substituted away with the differential. Here is the integral for this problem.

$$\begin{aligned}\int \frac{1}{x \ln(x)}\, dx &= \int \frac{1}{u}\, du \\ &= \ln|u| + c \\ &= \ln|\ln(x)| + c\end{aligned}$$

(b) $\displaystyle\int \frac{\mathbf{e}^{2t}}{1+\mathbf{e}^{2t}}\, dt$

Again, the substitution here may seem a little tricky. In this case the substitution is,

$$u = 1+\mathbf{e}^{2t} \qquad du = 2\mathbf{e}^{2t}dt \qquad \Rightarrow \qquad \mathbf{e}^{2t}dt = \frac{1}{2}du$$

The integral is then,

$$\begin{aligned}\int \frac{\mathbf{e}^{2t}}{1+\mathbf{e}^{2t}}\, dt &= \frac{1}{2}\int \frac{1}{u}\, du \\ &= \frac{1}{2}\ln\left|1+\mathbf{e}^{2t}\right| + c\end{aligned}$$

(c) $\displaystyle\int \frac{\mathbf{e}^{2t}}{1+\mathbf{e}^{4t}}\,dt$

In this case we can't use the same type of substitution that we used in the previous problem. In order to use the substitution in the previous example the exponential in the numerator and the denominator need to be the same and in this case they aren't.

To see the correct substitution for this problem note that,

$$\mathbf{e}^{4t} = \left(\mathbf{e}^{2t}\right)^2$$

Using this, the integral can be written as follows,

$$\int \frac{\mathbf{e}^{2t}}{1+\mathbf{e}^{4t}}\,dt = \int \frac{\mathbf{e}^{2t}}{1+\left(\mathbf{e}^{2t}\right)^2}\,dt$$

We can now use the following substitution.

$$u = \mathbf{e}^{2t} \qquad du = 2\mathbf{e}^{2t}dt \qquad \Rightarrow \qquad \mathbf{e}^{2t}dt = \frac{1}{2}du$$

The integral is then,

$$\begin{aligned}\int \frac{\mathbf{e}^{2t}}{1+\mathbf{e}^{4t}}\,dt &= \frac{1}{2}\int \frac{1}{1+u^2}\,du\\ &= \frac{1}{2}\tan^{-1}(u)+c\\ &= \frac{1}{2}\tan^{-1}\left(\mathbf{e}^{2t}\right)+c\end{aligned}$$

(d) $\displaystyle\int \frac{\sin^{-1}(x)}{\sqrt{1-x^2}}\,dx$

This integral is similar to the first problem in this set. Since we don't know how to integrate inverse sine functions it seems likely that this will be our substitution. If we use this as our substitution we get,

$$u = \sin^{-1}(x) \qquad du = \frac{1}{\sqrt{1-x^2}}dx$$

So, the root in the integral will get taken care of in the substitution process and this will eliminate all the x's from the integral. Therefore, this was the correct substitution.

The integral is,

$$\begin{aligned}\int \frac{\sin^{-1}(x)}{\sqrt{1-x^2}}\,dx &= \int u\,du\\ &= \frac{1}{2}u^2+c\\ &= \frac{1}{2}\left(\sin^{-1}(x)\right)^2+c\end{aligned}$$

Over the last couple of sections we've seen a lot of substitution rule examples. There are a couple of general rules that we will need to remember when doing these problems. First, when doing a substitution remember that when the substitution is done all the x's in the integral (or whatever variable is being used for that particular integral) should all be substituted away. This includes the x in the dx. After the substitution only u's should be left in the integral. Also, sometimes the correct substitution is a little tricky to find and more often than not there will need to be some manipulation of the differential or integrand in order to actually do the substitution.

Also, many integrals will require us to break them up so we can do multiple substitutions so be on the lookout for those kinds of integrals/substitutions.

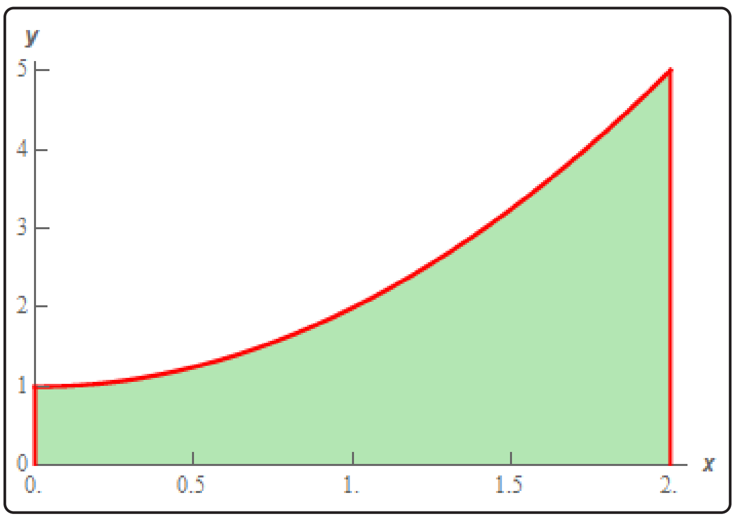
5.5 Area Problem

As noted in the first section of this section there are two kinds of integrals and to this point we've looked at indefinite integrals. It is now time to start thinking about the second kind of integral : Definite Integrals. However, before we do that we're going to take a look at the Area Problem. The area problem is to definite integrals what the tangent and rate of change problems are to derivatives.

The area problem will give us one of the interpretations of a definite integral and it will lead us to the definition of the definite integral.

To start off we are going to assume that we've got a function $f(x)$ that is positive on some interval $[a, b]$. What we want to do is determine the area of the region between the function and the x-axis.

It's probably easiest to see how we do this with an example. So, let's determine the area between $f(x) = x^2 + 1$ on $[0, 2]$. In other words, we want to determine the area of the shaded region below.

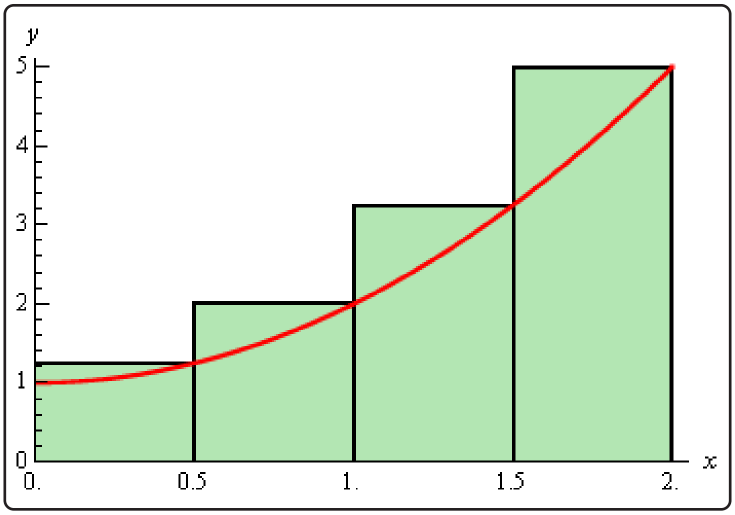


Now, at this point, we can't do this exactly. However, we can estimate the area. We will estimate the area by dividing up the interval into n subintervals each of width,

$$\Delta x = \frac{b-a}{n}$$

Then in each interval we can form a rectangle whose height is given by the function value at a specific point in the interval. We can then find the area of each of these rectangles, add them up and this will be an estimate of the area.

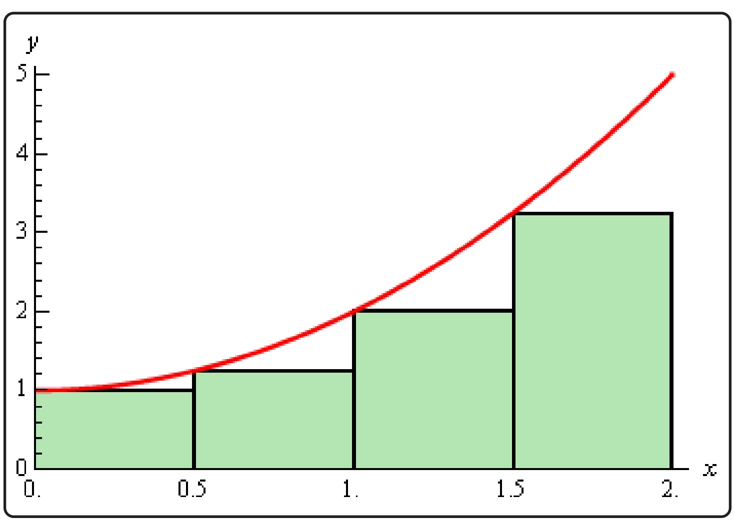
It's probably easier to see this with a sketch of the situation. So, let's divide up the interval into 4 subintervals and use the function value at the right endpoint of each interval to define the height of the rectangle. This gives,



Note that by choosing the height as we did each of the rectangles will over estimate the area since each rectangle takes in more area than the graph each time. Now let's estimate the area. First, the width of each of the rectangles is $\frac{1}{2}$. The height of each rectangle is determined by the function value at the right endpoint and so the height of each rectangle is nothing more that the function value at the right endpoint. Here is the estimated area.

$$\begin{aligned} A_r &= \frac{1}{2}f\left(\frac{1}{2}\right) + \frac{1}{2}f\left(1\right) + \frac{1}{2}f\left(\frac{3}{2}\right) + \frac{1}{2}f\left(2\right) \\ &= \frac{1}{2}\left(\frac{5}{4}\right) + \frac{1}{2}\left(2\right) + \frac{1}{2}\left(\frac{13}{4}\right) + \frac{1}{2}\left(5\right) \\ &= 5.75 \end{aligned}$$

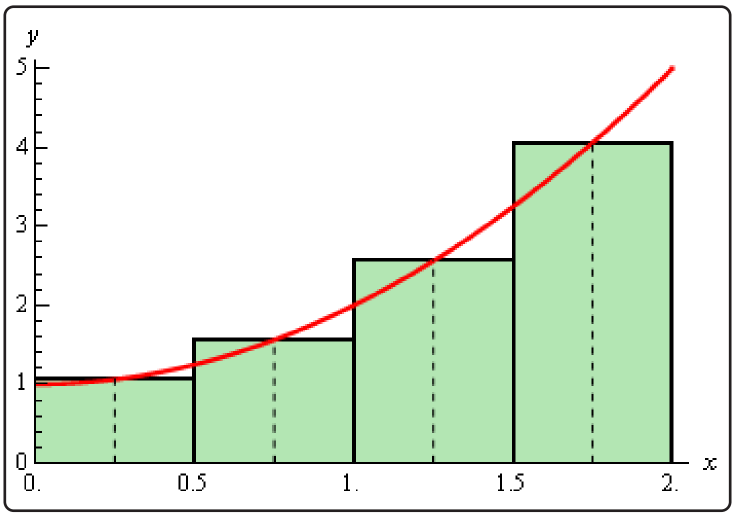
Of course, taking the rectangle heights to be the function value at the right endpoint is not our only option. We could have taken the rectangle heights to be the function value at the left endpoint. Using the left endpoints as the heights of the rectangles will give the following graph and estimated area.



$$
\begin{aligned}
A_l &= \frac{1}{2} f(0) + \frac{1}{2} f\left(\frac{1}{2}\right) + \frac{1}{2} f(1) + \frac{1}{2} f\left(\frac{3}{2}\right) \\
&= \frac{1}{2}(1) + \frac{1}{2}\left(\frac{5}{4}\right) + \frac{1}{2}(2) + \frac{1}{2}\left(\frac{13}{4}\right) \\
&= 3.75
\end{aligned}
$$

In this case we can see that the estimation will be an underestimation since each rectangle misses some of the area each time.

There is one more common point for getting the heights of the rectangles that is often more accurate. Instead of using the right or left endpoints of each sub interval we could take the midpoint of each subinterval as the height of each rectangle. Here is the graph for this case.



So, it looks like each rectangle will over and under estimate the area. This means that the approximation this time should be much better than the previous two choices of points. Here is the estimation for this case.

$$
\begin{aligned}
A_m &= \frac{1}{2} f\left(\frac{1}{4}\right) + \frac{1}{2} f\left(\frac{3}{4}\right) + \frac{1}{2} f\left(\frac{5}{4}\right) + \frac{1}{2} f\left(\frac{7}{4}\right) \\
&= \frac{1}{2}\left(\frac{17}{16}\right) + \frac{1}{2}\left(\frac{25}{16}\right) + \frac{1}{2}\left(\frac{41}{16}\right) + \frac{1}{2}\left(\frac{65}{16}\right) \\
&= 4.625
\end{aligned}
$$

We've now got three estimates. For comparison's sake the exact area is

$$A = \frac{14}{3} = 4.66\overline{6}$$

So, both the right and left endpoint estimation did not do all that great of a job at the estimation. The midpoint estimation however did quite well.

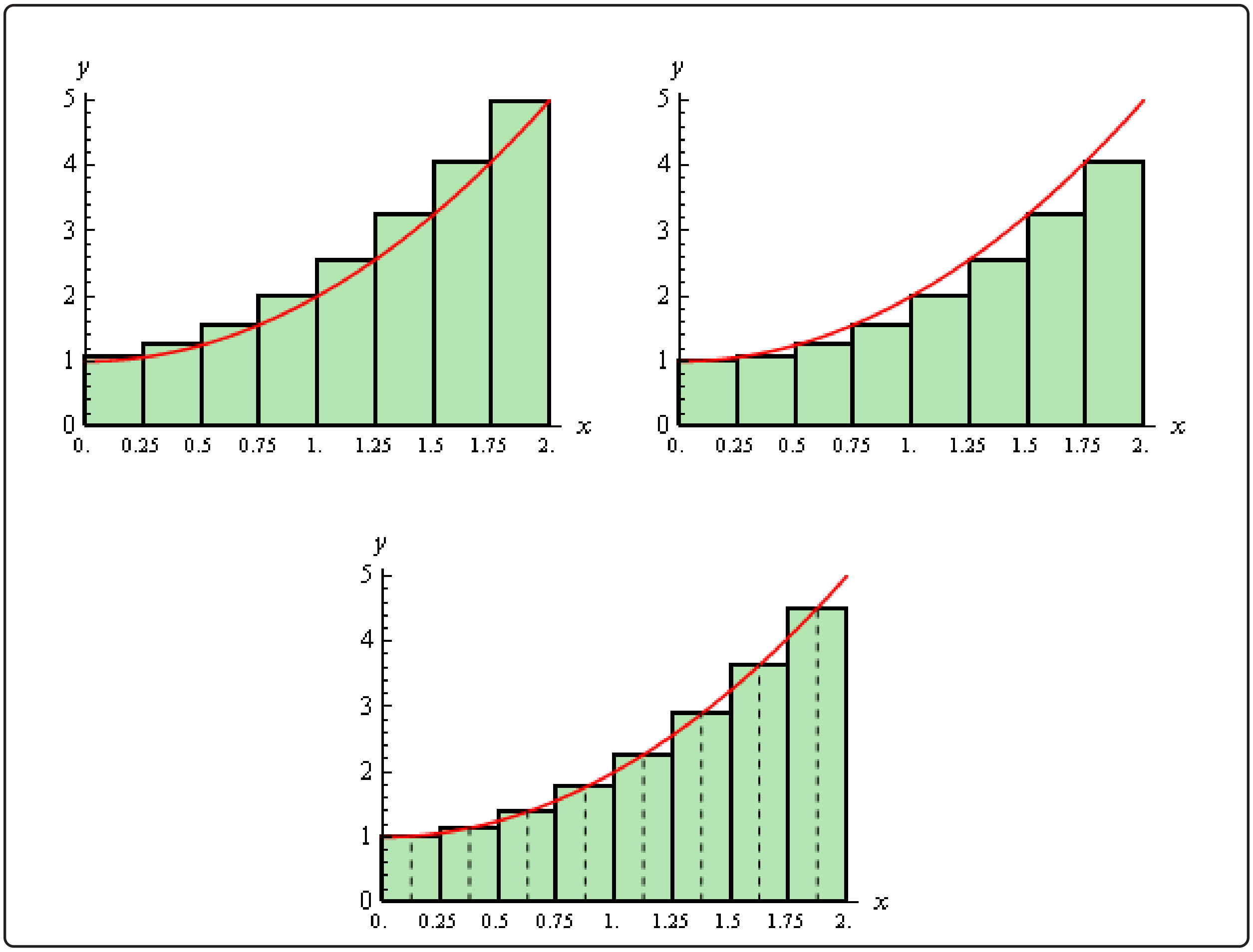
Be careful to not draw any conclusion about how choosing each of the points will affect our estimation. In this case, because we are working with an increasing function choosing the right endpoints will overestimate and choosing left endpoint will underestimate.

If we were to work with a decreasing function we would get the opposite results. For decreasing functions the right endpoints will underestimate and the left endpoints will overestimate.

Also, if we had a function that both increased and decreased in the interval we would, in all likelihood, not even be able to determine if we would get an overestimation or underestimation.

Now, let's suppose that we want a better estimation, because none of the estimations above really did all that great of a job at estimating the area. We could try to find a different point to use for the height of each rectangle but that would be cumbersome and there wouldn't be any guarantee that the estimation would in fact be better. Also, we would like a method for getting better approximations that would work for any function we would chose to work with and if we just pick new points that may not work for other functions.

The easiest way to get a better approximation is to take more rectangles (*i.e.* increase n). Let's double the number of rectangles that we used and see what happens. Here are the graphs showing the eight rectangles and the estimations for each of the three choices for rectangle heights that we used above.



Here are the area estimations for each of these cases.

$$A_r = 5.1875 \qquad A_l = 4.1875 \qquad A_m = 4.65625$$

So, increasing the number of rectangles did improve the accuracy of the estimation as we'd guessed that it would.

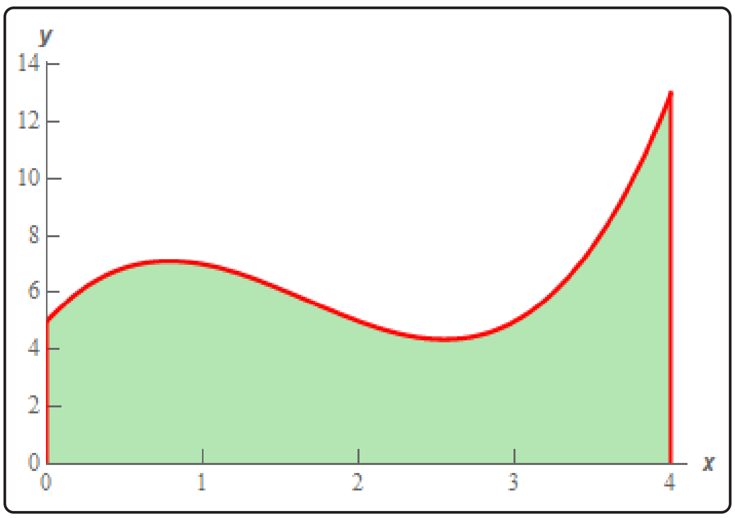
Let's work a slightly more complicated example.

Example 1

Estimate the area between $f(x) = x^3 - 5x^2 + 6x + 5$ and the x-axis on $[0, 4]$ using $n = 5$ subintervals and all three cases above for the heights of each rectangle.

Solution

First, let's get the graph to make sure that the function is positive.



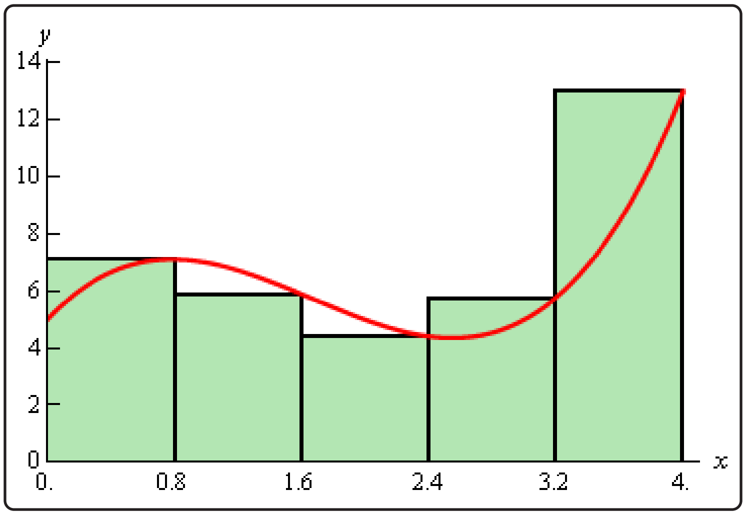
So, the graph is positive and the width of each subinterval will be,

$$\Delta x = \frac{4}{5} = 0.8$$

This means that the endpoints of the subintervals are,

$$0, \ 0.8, \ 1.6, \ 2.4, \ 3.2, \ 4$$

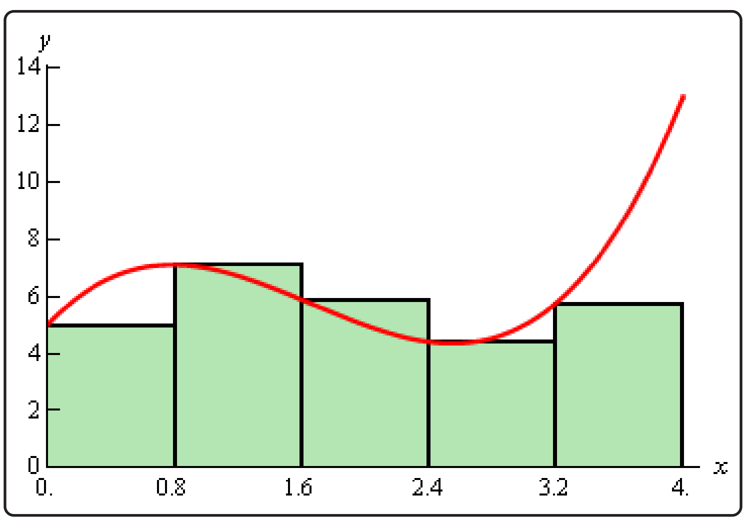
Let's first look at using the right endpoints for the function height. Here is the graph for this case.



Notice, that unlike the first area we looked at, the choosing the right endpoints here will both over and underestimate the area depending on where we are on the curve. This will often be the case with a more general curve that the one we initially looked at. The area estimation using the right endpoints of each interval for the rectangle height is,

$$\begin{aligned} A_r &= 0.8f(0.8) + 0.8f(1.6) + 0.8f(2.4) + 0.8f(3.2) + 0.8f(4) \\ &= 28.96 \end{aligned}$$

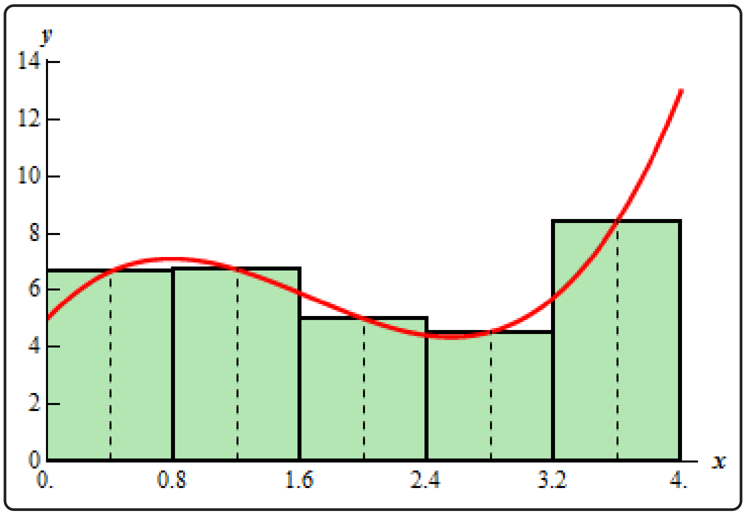
Now let's take a look at left endpoints for the function height. Here is the graph.



The area estimation using the left endpoints of each interval for the rectangle height is,

$$\begin{aligned} A_l &= 0.8f(0) + 0.8f(0.8) + 0.8f(1.6) + 0.8f(2.4) + 0.8f(3.2) \\ &= 22.56 \end{aligned}$$

Finally, let's take a look at the midpoints for the heights of each rectangle. Here is the graph,



The area estimation using the midpoint is then,

$$\begin{aligned} A_m &= 0.8f\,(0.4) + 0.8f\,(1.2) + 0.8f\,(2) + 0.8f\,(2.8) + 0.8f\,(3.6) \\ &= 25.12 \end{aligned}$$

For comparison purposes the exact area is,

$$A = \frac{76}{3} = 25.33\overline{3}$$

So, again the midpoint did a better job than the other two. While this will be the case more often than not, it won't always be the case and so don't expect this to always happen.

Now, let's move on to the general case. Let's start out with $f\,(x) \geq 0$ on $[a, b]$ and we'll divide the interval into n subintervals each of length,

$$\Delta x = \frac{b-a}{n}$$

Note that the subintervals don't have to be equal length, but it will make our work significantly easier. The endpoints of each subinterval are,

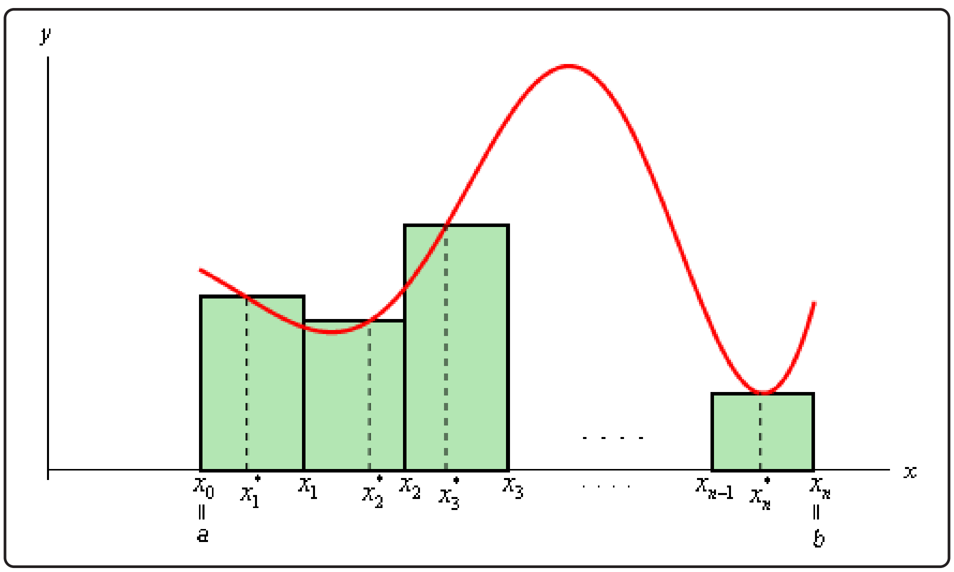
$$\begin{aligned} x_0 &= a \\ x_1 &= a + \Delta x \\ x_2 &= a + 2\Delta x \\ &\ \vdots \\ x_i &= a + i\Delta x \\ &\ \vdots \\ x_{n-1} &= a + (n-1)\,\Delta x \\ x_n &= a + n\Delta x = b \end{aligned}$$

Next in each interval,

$$[x_0, x_1]\,,\ [x_1, x_2]\,, \ldots, [x_{i-1}, x_i]\,, \ldots, [x_{n-1}, x_n]$$

we choose a point $x_1^*, x_2^*, \ldots, x_i^*, \ldots x_n^*$. These points will define the height of the rectangle in each subinterval. Note as well that these points do not have to occur at the same point in each subinterval. However, they are usually the left end point of the interval, right end point of the interval or the midpoint of the interval.

Here is a sketch of this situation.



The area under the curve on the given interval is then approximately,

$$A \approx f\left(x_1^*\right)\Delta x + f\left(x_2^*\right)\Delta x + \cdots + f\left(x_i^*\right)\Delta x + \cdots + f\left(x_n^*\right)\Delta x$$

We will use **summation notation** or **sigma notation** at this point to simplify up our notation a little. If you need a refresher on summation notation check out the section devoted to this in the Extras appendix.

Using summation notation the area estimation is,

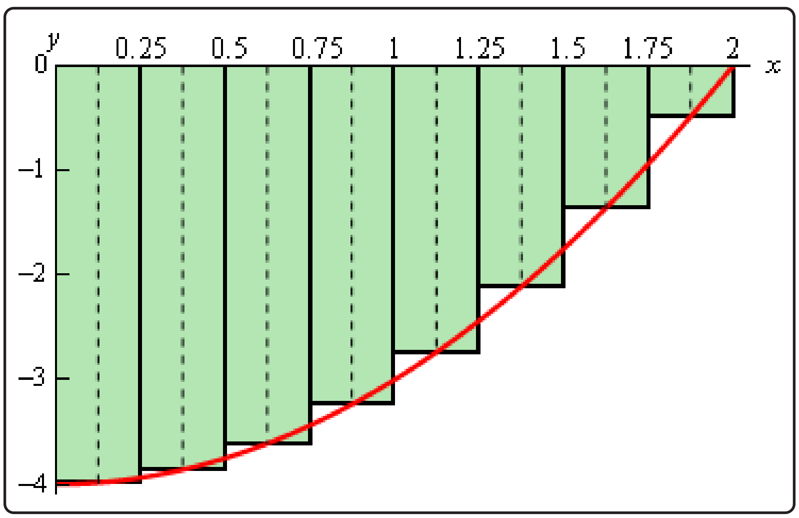
$$A \approx \sum_{i=1}^{n} f\left(x_i^*\right) \Delta x$$

The summation in the above equation is called a **Riemann Sum**.

To get a better estimation we will take n larger and larger. In fact, if we let n go out to infinity we will get the exact area. In other words,

$$A = \lim_{n\to\infty} \sum_{i=1}^{n} f\left(x_i^*\right) \Delta x$$

Before leaving this section let's address one more issue. To this point we've required the function to be positive in our work. Many functions are not positive however. Consider the case of $f(x) = x^2 - 4$ on $[0, 2]$. If we use $n = 8$ and the midpoints for the rectangle height we get the following graph,

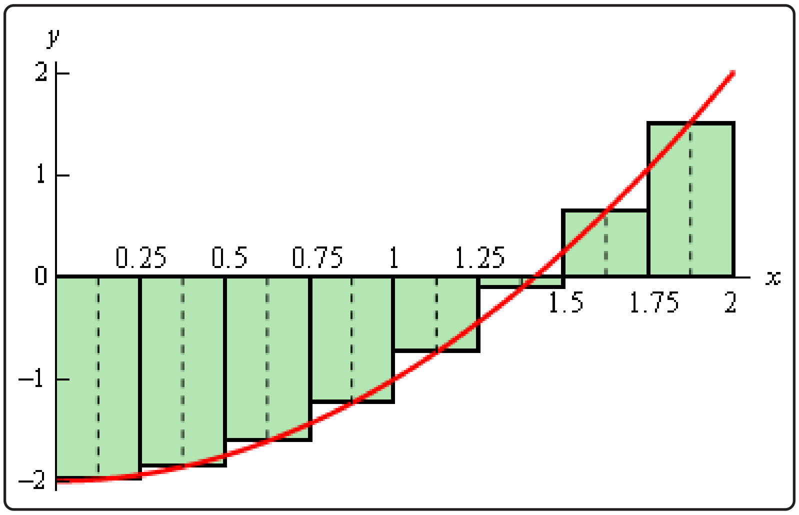


In this case let's notice that the function lies completely below the x-axis and hence is always negative. If we ignore the fact that the function is always negative and use the same ideas above to estimate the area between the graph and the x-axis we get,

$$\begin{aligned} A_m &= \frac{1}{4} f\left(\frac{1}{8}\right) + \frac{1}{4} f\left(\frac{3}{8}\right) + \frac{1}{4} f\left(\frac{5}{8}\right) + \frac{1}{4} f\left(\frac{7}{8}\right) + \frac{1}{4} f\left(\frac{9}{8}\right) + \\ &\qquad \frac{1}{4} f\left(\frac{11}{8}\right) + \frac{1}{4} f\left(\frac{13}{8}\right) + \frac{1}{4} f\left(\frac{15}{8}\right) \\ &= -5.34375 \end{aligned}$$

Our answer is negative as we might have expected given that all the function evaluations are negative.

So, using the technique in this section it looks like if the function is above the x-axis we will get a positive area and if the function is below the x-axis we will get a negative area. Now, what about a function that is both positive and negative in the interval? For example, $f(x) = x^2 - 2$ on $[0, 2]$. Using $n = 8$ and midpoints the graph is,



Some of the rectangles are below the x-axis and so will give negative areas while some are above the x-axis and will give positive areas. Since more rectangles are below the x-axis than above it looks like we should probably get a negative area estimation for this case. In fact that is correct. Here the area estimation for this case.

$$\begin{aligned} A_m &= \frac{1}{4}f\left(\frac{1}{8}\right) + \frac{1}{4}f\left(\frac{3}{8}\right) + \frac{1}{4}f\left(\frac{5}{8}\right) + \frac{1}{4}f\left(\frac{7}{8}\right) + \frac{1}{4}f\left(\frac{9}{8}\right) + \\ &\qquad \frac{1}{4}f\left(\frac{11}{8}\right) + \frac{1}{4}f\left(\frac{13}{8}\right) + \frac{1}{4}f\left(\frac{15}{8}\right) \\ &= -1.34375 \end{aligned}$$

In cases where the function is both above and below the x-axis the technique given in the section will give the *net* area between the function and the x-axis with areas below the x-axis negative and areas above the x-axis positive. So, if the net area is negative then there is more area under the x-axis than above while a positive net area will mean that more of the area is above the x-axis.H

5.6 Definition of the Definite Integral

In this section we will formally define the definite integral and give many of the properties of definite integrals. Let's start off with the definition of a definite integral.

Definite Integral

Given a function $f(x)$ that is continuous on the interval $[a, b]$ we divide the interval into n subintervals of equal width, Δx, and from each interval choose a point, x_i^*. Then the **definite integral of** $f(x)$ **from** a **to** b is

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\,\Delta x$$

The definite integral is defined to be exactly the limit and summation that we looked at in the last section to find the net area between a function and the x-axis. Also note that the notation for the definite integral is very similar to the notation for an indefinite integral. The reason for this will be apparent eventually.

There is also a little bit of terminology that we should get out of the way here. The number "a" that is at the bottom of the integral sign is called the **lower limit** of the integral and the number "b" at the top of the integral sign is called the **upper limit** of the integral. Also, despite the fact that a and b were given as an interval the lower limit does not necessarily need to be smaller than the upper limit. Collectively we'll often call a and b the **interval of integration**.

Let's work a quick example. This example will use many of the properties and facts from the brief review of summation notation in the Extras appendix.

Example 1

Using the definition of the definite integral compute the following.

$$\int_0^2 x^2 + 1\,dx$$

Solution

First, we can't actually use the definition unless we determine which points in each interval that well use for x_i^*. In order to make our life easier we'll use the right endpoints of each interval.

From the previous section we know that for a general n the width of each subinterval is,

$$\Delta x = \frac{2-0}{n} = \frac{2}{n}$$

The subintervals are then,

$$\left[0,\frac{2}{n}\right],\quad \left[\frac{2}{n},\frac{4}{n}\right],\quad \left[\frac{4}{n},\frac{6}{n}\right],\ \ldots,\ \left[\frac{2(i-1)}{n},\frac{2i}{n}\right],\ \ldots,\ \left[\frac{2(n-1)}{n},2\right]$$

As we can see the right endpoint of the i^{th} subinterval is

$$x_i^* = \frac{2i}{n}$$

The summation in the definition of the definite integral is then,

$$\begin{aligned}\sum_{i=1}^{n} f\left(x_i^*\right)\Delta x &= \sum_{i=1}^{n} f\left(\frac{2i}{n}\right)\left(\frac{2}{n}\right)\\ &= \sum_{i=1}^{n}\left(\left(\frac{2i}{n}\right)^2+1\right)\left(\frac{2}{n}\right)\\ &= \sum_{i=1}^{n}\left(\frac{8i^2}{n^3}+\frac{2}{n}\right)\end{aligned}$$

Now, we are going to have to take a limit of this. That means that we are going to need to "evaluate" this summation. In other words, we are going to have to use the formulas given in the summation notation review to eliminate the actual summation and get a formula for this for a general n.

To do this we will need to recognize that n is a constant as far as the summation notation is concerned. As we cycle through the integers from 1 to n in the summation only i changes and so anything that isn't an i will be a constant and can be factored out of the summation. In particular any n that is in the summation can be factored out if we need to.

Here is the summation "evaluation".

$$\begin{aligned}\sum_{i=1}^{n} f\left(x_i^*\right)\Delta x &= \sum_{i=1}^{n}\frac{8i^2}{n^3}+\sum_{i=1}^{n}\frac{2}{n}\\ &= \frac{8}{n^3}\sum_{i=1}^{n} i^2+\frac{1}{n}\sum_{i=1}^{n}2\\ &= \frac{8}{n^3}\left(\frac{n(n+1)(2n+1)}{6}\right)+\frac{1}{n}(2n)\\ &= \frac{4(n+1)(2n+1)}{3n^2}+2\\ &= \frac{14n^2+12n+4}{3n^2}\end{aligned}$$

We can now compute the definite integral.

$$\begin{aligned}\int_0^2 x^2 + 1\,dx &= \lim_{n\to\infty}\sum_{i=1}^{n} f\left(x_i^*\right)\Delta x\\ &= \lim_{n\to\infty}\frac{14n^2+12n+4}{3n^2}\\ &= \frac{14}{3}\end{aligned}$$

We've seen several methods for dealing with the limit in this problem so we'll leave it to you to verify the results.

Wow, that was a lot of work for a fairly simple function. There is a much simpler way of evaluating these and we will get to it eventually. The main purpose to this section is to get the main properties and facts about the definite integral out of the way. We'll discuss how we compute these in practice starting with the next section.

So, let's start taking a look at some of the properties of the definite integral.

Properties

1. $\displaystyle\int_a^b f\left(x\right)\,dx = -\int_b^a f\left(x\right)\,dx.$

   We can interchange the limits on any definite integral, all that we need to do is tack a minus sign onto the integral when we do.

2. $\displaystyle\int_a^a f\left(x\right)\,dx = 0.$

   If the upper and lower limits are the same then there is no work to do, the integral is zero.

3. $\displaystyle\int_a^b cf\left(x\right)\,dx = c\int_a^b f\left(x\right)\,dx$, where c is any number.

   So, as with limits, derivatives, and indefinite integrals we can factor out a constant.

4. $\displaystyle\int_a^b f\left(x\right)\pm g\left(x\right)\,dx = \int_a^b f\left(x\right)\,dx \pm \int_a^b g\left(x\right)\,dx.$

   We can break up definite integrals across a sum or difference.

5. $\displaystyle\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$ where c is any number.

   This property is more important than we might realize at first. One of the main uses of this property is to tell us how we can integrate a function over the adjacent intervals, $[a, c]$ and $[c, b]$. Note however that c doesn't need to be between a and b.

6. $\displaystyle\int_a^b f(x)\,dx = \int_a^b f(t)\,dt.$

   The point of this property is to notice that as long as the function and limits are the same the variable of integration that we use in the definite integral won't affect the answer.

See the Proof of Various Integral Properties section of the Extras appendix for the proof of properties 1 – 4. Property 5 is not easy to prove and so is not shown there. Property 6 is not really a property in the full sense of the word. It is only here to acknowledge that as long as the function and limits are the same it doesn't matter what letter we use for the variable. The answer will be the same.

Let's do a couple of examples dealing with these properties.

Example 2

Use the results from the first example to evaluate each of the following.

(a) $\displaystyle\int_2^0 x^2 + 1\,dx$

(b) $\displaystyle\int_0^2 10x^2 + 10\,dx$

(c) $\displaystyle\int_0^2 t^2 + 1\,dt$

Solution

All of the solutions to these problems will rely on the fact we proved in the first example. Namely that,

$$\int_0^2 x^2 + 1\,dx = \frac{14}{3}$$

(a) $\displaystyle\int_{2}^{0} x^2 + 1\,dx$

In this case the only difference between the two is that the limits have interchanged. So, using the first property gives,

$$\begin{aligned}\int_{2}^{0} x^2 + 1\,dx &= -\int_{0}^{2} x^2 + 1\,dx \\ &= -\frac{14}{3}\end{aligned}$$

(b) $\displaystyle\int_{0}^{2} 10x^2 + 10\,dx$

For this part notice that we can factor a 10 out of both terms and then out of the integral using the third property.

$$\begin{aligned}\int_{0}^{2} 10x^2 + 10\,dx &= \int_{0}^{2} 10\left(x^2 + 1\right)\,dx \\ &= 10\int_{0}^{2} x^2 + 1\,dx \\ &= 10\left(\frac{14}{3}\right) \\ &= \frac{140}{3}\end{aligned}$$

(c) $\displaystyle\int_{0}^{2} t^2 + 1\,dt$

In this case the only difference is the letter used and so this is just going to use property 6.

$$\int_{0}^{2} t^2 + 1\,dt = \int_{0}^{2} x^2 + 1\,dx = \frac{14}{3}$$

Here are a couple of examples using the other properties.

Example 3

Evaluate the following definite integral.

$$\int_{130}^{130} \frac{x^3 - x\sin(x) + \cos(x)}{x^2+1}\,dx$$

Solution

There really isn't anything to do with this integral once we notice that the limits are the same. Using the second property this is,

$$\int_{130}^{130} \frac{x^3 - x\sin(x) + \cos(x)}{x^2+1}\,dx = 0$$

Example 4

Given that $\int_{6}^{-10} f(x)\,dx = 23$ and $\int_{-10}^{6} g(x)\,dx = -9$ determine the value of

$$\int_{-10}^{6} 2f(x) - 10g(x)\,dx$$

Solution

We will first need to use the fourth property to break up the integral and the third property to factor out the constants.

$$\begin{aligned}\int_{-10}^{6} 2f(x) - 10g(x)\,dx &= \int_{-10}^{6} 2f(x)\,dx - \int_{-10}^{6} 10g(x)\,dx \\ &= 2\int_{-10}^{6} f(x)\,dx - 10\int_{-10}^{6} g(x)\,dx\end{aligned}$$

Now notice that the limits on the first integral are interchanged with the limits on the given integral so switch them using the first property above (and adding a minus sign of course). Once this is done we can plug in the known values of the integrals.

$$\begin{aligned}\int_{-10}^{6} 2f(x) - 10g(x)\,dx &= -2\int_{6}^{-10} f(x)\,dx - 10\int_{-10}^{6} g(x)\,dx \\ &= -2(23) - 10(-9) \\ &= 44\end{aligned}$$

Example 5

Given that $\displaystyle\int_{12}^{-10} f(x)\,dx = 6$, $\displaystyle\int_{100}^{-10} f(x)\,dx = -2$, and $\displaystyle\int_{100}^{-5} f(x)\,dx = 4$ determine the value of $\displaystyle\int_{-5}^{12} f(x)\,dx$.

Solution

This example is mostly an example of property 5 although there are a couple of uses of property 1 in the solution as well.

We need to figure out how to correctly break up the integral using property 5 to allow us to use the given pieces of information. First, we'll note that there is an integral that has a "-5" in one of the limits. It's not the lower limit, but we can use property 1 to correct that eventually. The other limit is 100 so this is the number c that we'll use in property 5.

$$\int_{-5}^{12} f(x)\,dx = \int_{-5}^{100} f(x)\,dx + \int_{100}^{12} f(x)\,dx$$

We'll be able to get the value of the first integral, but the second still isn't in the list of know integrals. However, we do have second integral that has a limit of 100 in it. The other limit for this second integral is -10 and this will be c in this application of property 5.

$$\int_{-5}^{12} f(x)\,dx = \int_{-5}^{100} f(x)\,dx + \int_{100}^{-10} f(x)\,dx + \int_{-10}^{12} f(x)\,dx$$

At this point all that we need to do is use the property 1 on the first and third integral to get the limits to match up with the known integrals. After that we can plug in for the known integrals.

$$\begin{aligned}\int_{-5}^{12} f(x)\,dx &= -\int_{100}^{-5} f(x)\,dx + \int_{100}^{-10} f(x)\,dx - \int_{12}^{-10} f(x)\,dx \\ &= -4 - 2 - 6 \\ &= -12\end{aligned}$$

There are also some nice properties that we can use in comparing the general size of definite integrals. Here they are.

More Properties

7. $\displaystyle\int_a^b c\,dx = c\,(b-a)$, c is any number.

8. If $f(x) \geq 0$ for $a \leq x \leq b$ then $\displaystyle\int_a^b f(x)\,dx \geq 0$.

9. If $f(x) \geq g(x)$ for $a \leq x \leq b$ then $\displaystyle\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx$.

10. If $m \leq f(x) \leq M$ for $a \leq x \leq b$ then $m\,(b-a) \leq \int_a^b f(x)\,dx \leq M\,(b-a)$.

11. $\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx$

See the Proof of Various Integral Properties section of the Extras appendix for the proof of these properties.

Interpretations of Definite Integral

There are a couple of quick interpretations of the definite integral that we can give here.

First, as we alluded to in the previous section one possible interpretation of the definite integral is to give the net area between the graph of $f(x)$ and the x-axis on the interval $[a,b]$. So, the net area between the graph of $f(x) = x^2 + 1$ and the x-axis on $[0,2]$ is,

$$\int_0^2 x^2 + 1\,dx = \frac{14}{3}$$

If you look back in the last section this was the exact area that was given for the initial set of problems that we looked at in this area.

Another interpretation is sometimes called the Net Change Theorem. This interpretation says that if $f(x)$ is some quantity (so $f'(x)$ is the rate of change of $f(x)$) then,

$$\int_a^b f'(x)\,dx = f(b) - f(a)$$

is the net change in $f(x)$ on the interval $[a,b]$. In other words, compute the definite integral of a rate of change and you'll get the net change in the quantity. We can see that the value of the definite integral, $f(b) - f(a)$, does in fact give us the net change in $f(x)$ and so there really isn't anything to prove with this statement. This is really just an acknowledgment of what the definite integral of a rate of change tells us.

So as a quick example, if $V(t)$ is the volume of water in a tank then,

$$\int_{t_1}^{t_2} V'(t)\,dt = V(t_2) - V(t_1)$$

is the net change in the volume as we go from time t_1 to time t_2.

Likewise, if $s(t)$ is the function giving the position of some object at time t we know that the velocity of the object at any time t is : $v(t) = s'(t)$. Therefore, the displacement of the object from time t_1 to time t_2 is,

$$\int_{t_1}^{t_2} v(t)\ dt = s(t_2) - s(t_1)$$

Note that in this case if $v(t)$ is both positive and negative (*i.e.* the object moves to both the right and left) in the time frame this will NOT give the total distance traveled. It will only give the displacement, *i.e.* the difference between where the object started and where it ended up. To get the total distance traveled by an object we'd have to compute,

$$\int_{t_1}^{t_2} |v(t)|\ dt$$

It is important to note here that the Net Change Theorem only really makes sense if we're integrating a derivative of a function.

Fundamental Theorem of Calculus, Part I

As noted by the title above this is only the first part to the Fundamental Theorem of Calculus. We will give the second part in the next section as it is the key to easily computing definite integrals and that is the subject of the next section.

The first part of the Fundamental Theorem of Calculus tells us how to differentiate certain types of definite integrals and it also tells us about the very close relationship between integrals and derivatives.

Fundamental Theorem of Calculus, Part I

If $f(x)$ is continuous on $[a, b]$ then,

$$g(x) = \int_{a}^{x} f(t)\ dt$$

is continuous on $[a, b]$ and it is differentiable on (a, b) and,

$$g'(x) = f(x)$$

An alternate notation for the derivative portion of this is,

$$\frac{d}{dx}\int_{a}^{x} f(t)\ dt = f(x)$$

To see the proof of this see the Proof of Various Integral Properties section of the Extras appendix.

Let's check out a couple of quick examples using this.

Example 6

Differentiate each of the following.

(a) $g(x) = \displaystyle\int_{-4}^{x} \mathbf{e}^{2t} \cos^2(1 - 5t)\, dt$

(b) $\displaystyle\int_{x^2}^{1} \frac{t^4 + 1}{t^2 + 1}\, dt$

Solution

(a) $g(x) = \displaystyle\int_{-4}^{x} \mathbf{e}^{2t} \cos^2(1 - 5t)\, dt$

This one is nothing more than a quick application of the Fundamental Theorem of Calculus.

$$g'(x) = \mathbf{e}^{2x} \cos^2(1 - 5x)$$

(b) $\displaystyle\int_{x^2}^{1} \frac{t^4 + 1}{t^2 + 1}\, dt$

This one needs a little work before we can use the Fundamental Theorem of Calculus. The first thing to notice is that the Fundamental Theorem of Calculus requires the lower limit to be a constant and the upper limit to be the variable. So, using a property of definite integrals we can interchange the limits of the integral we just need to remember to add in a minus sign after we do that. Doing this gives,

$$\frac{d}{dx} \int_{x^2}^{1} \frac{t^4 + 1}{t^2 + 1}\, dt = \frac{d}{dx}\left(- \int_{1}^{x^2} \frac{t^4 + 1}{t^2 + 1}\, dt \right) = -\frac{d}{dx} \int_{1}^{x^2} \frac{t^4 + 1}{t^2 + 1}\, dt$$

The next thing to notice is that the Fundamental Theorem of Calculus also requires an x in the upper limit of integration and we've got x^2. To do this derivative we're going to need the following version of the chain rule.

$$\frac{d}{dx}\Big(g(u)\Big) = \frac{d}{du}\Big(g(u)\Big) \frac{du}{dx} \qquad \text{where } u = f(x)$$

So, if we let $u = x^2$ we use the chain rule to get,

$$\begin{aligned}\frac{d}{dx}\int_{x^2}^{1}\frac{t^4+1}{t^2+1}\,dt &= -\frac{d}{dx}\int_{1}^{x^2}\frac{t^4+1}{t^2+1}\,dt \\ &= -\frac{d}{du}\int_{1}^{u}\frac{t^4+1}{t^2+1}\,dt\;\frac{du}{dx} \qquad \text{where } u = x^2 \\ &= -\frac{u^4+1}{u^2+1}\left(2x\right) \\ &= -2x\frac{u^4+1}{u^2+1}\end{aligned}$$

The final step is to get everything back in terms of x.

$$\begin{aligned}\frac{d}{dx}\int_{x^2}^{1}\frac{t^4+1}{t^2+1}\,dt &= -2x\frac{\left(x^2\right)^4+1}{\left(x^2\right)^2+1} \\ &= -2x\frac{x^8+1}{x^4+1}\end{aligned}$$

Using the chain rule as we did in the last part of this example we can derive some general formulas for some more complicated problems.

First,

$$\frac{d}{dx}\int_{a}^{u(x)} f\left(t\right)\,dt = u'\left(x\right)f\big(u\left(x\right)\big)$$

This is simply the chain rule for these kinds of problems.

Next, we can get a formula for integrals in which the upper limit is a constant and the lower limit is a function of x. All we need to do here is interchange the limits on the integral (adding in a minus sign of course) and then use the formula above to get,

$$\frac{d}{dx}\int_{v(x)}^{b} f\left(t\right)\,dt = -\frac{d}{dx}\int_{b}^{v(x)} f\left(t\right)\,dt = -v'\left(x\right)f\big(v\left(x\right)\big)$$

Finally, we can also get a version for both limits being functions of x. In this case we'll need to use Property 5 above to break up the integral as follows,

$$\int_{v(x)}^{u(x)} f\left(t\right)\,dt = \int_{v(x)}^{a} f\left(t\right)\,dt + \int_{a}^{u(x)} f\left(t\right)\,dt$$

We can use pretty much any value of a when we break up the integral. The only thing that we need to do is to make sure that $f\left(a\right)$ exists. So, assuming that $f\left(a\right)$ exists after we break up the integral we can then differentiate and use the two formulas above to get,

$$\begin{aligned}\frac{d}{dx}\int_{v(x)}^{u(x)} f\left(t\right)\,dt &= \frac{d}{dx}\left(\int_{v(x)}^{a} f\left(t\right)\,dt + \int_{a}^{u(x)} f\left(t\right)\,dt\right) \\ &= -v'\left(x\right)f\big(v\left(x\right)\big) + u'\left(x\right)f\big(u\left(x\right)\big)\end{aligned}$$

Let's work a quick example.

Example 7

Differentiate the following integral.

$$\int_{\sqrt{x}}^{3x} t^2 \sin\left(1+t^2\right)\, dt$$

Solution

This will use the final formula that we derived above.

$$\begin{aligned}\frac{d}{dx}\int_{\sqrt{x}}^{3x} t^2 \sin\left(1+t^2\right)\, dt &= -\frac{1}{2}x^{-\frac{1}{2}}\left(\sqrt{x}\right)^2 \sin\left(1+\left(\sqrt{x}\right)^2\right) + (3)\left(3x\right)^2 \sin\left(1+\left(3x\right)^2\right) \\ &= -\frac{1}{2}\sqrt{x}\sin\left(1+x\right) + 27x^2\sin\left(1+9x^2\right)\end{aligned}$$

5.7 Computing Definite Integrals

In this section we are going to concentrate on how we actually evaluate definite integrals in practice. To do this we will need the Fundamental Theorem of Calculus, Part II.

Fundamental Theorem of Calculus, Part II

Suppose $f(x)$ is a continuous function on $[a, b]$ and also suppose that $F(x)$ is any anti-derivative for $f(x)$. Then,

$$\int_a^b f(x)\,dx = F(x)\Big|_a^b = F(b) - F(a)$$

To see the proof of this see the Proof of Various Integral Properties section of the Extras appendix.

Recall that when we talk about an anti-derivative for a function we are really talking about the indefinite integral for the function. So, to evaluate a definite integral the first thing that we're going to do is evaluate the indefinite integral for the function. This should explain the similarity in the notations for the indefinite and definite integrals.

Also notice that we require the function to be continuous in the interval of integration. This was also a requirement in the definition of the definite integral. We didn't make a big deal about this in the last section. In this section however, we will need to keep this condition in mind as we do our evaluations.

Next let's address the fact that we can use any anti-derivative of $f(x)$ in the evaluation. Let's take a final look at the following integral.

$$\int_0^2 x^2 + 1\,dx$$

Both of the following are anti-derivatives of the integrand.

$$F(x) = \frac{1}{3}x^3 + x \qquad \text{and} \qquad F(x) = \frac{1}{3}x^3 + x - \frac{18}{31}$$

Using the Fundamental Theorem of Calculus to evaluate this integral with the first anti-derivatives gives,

$$\begin{aligned}\int_0^2 x^2 + 1\,dx &= \left(\frac{1}{3}x^3 + x\right)\Bigg|_0^2 \\ &= \frac{1}{3}(2)^3 + 2 - \left(\frac{1}{3}(0)^3 + 0\right) \\ &= \frac{14}{3}\end{aligned}$$

Much easier than using the definition wasn't it? Let's now use the second anti-derivative to evaluate this definite integral.

$$\begin{aligned}\int_0^2 x^2 + 1\,dx &= \left(\frac{1}{3}x^3 + x - \frac{18}{31}\right)\bigg|_0^2 \\ &= \frac{1}{3}(2)^3 + 2 - \frac{18}{31} - \left(\frac{1}{3}(0)^3 + 0 - \frac{18}{31}\right) \\ &= \frac{14}{3} - \frac{18}{31} + \frac{18}{31} \\ &= \frac{14}{3}\end{aligned}$$

The constant that we tacked onto the second anti-derivative canceled in the evaluation step. So, when choosing the anti-derivative to use in the evaluation process make your life easier and don't bother with the constant as it will only end up canceling in the long run.

Also, note that we're going to have to be very careful with minus signs and parenthesis with these problems. It's very easy to get in a hurry and mess them up.

Let's start our examples with the following set designed to make a couple of quick points that are very important.

Example 1

Evaluate each of the following.

(a) $\displaystyle\int y^2 + y^{-2}\,dy$

(b) $\displaystyle\int_1^2 y^2 + y^{-2}\,dy$

(c) $\displaystyle\int_{-1}^2 y^2 + y^{-2}\,dy$

Solution

(a) $\displaystyle\int y^2 + y^{-2}\,dy$

This is the only indefinite integral in this section and by now we should be getting pretty good with these so we won't spend a lot of time on this part. This is here only to make sure that we understand the difference between an indefinite and a definite integral. The integral is,

$$\int y^2 + y^{-2}\,dy = \frac{1}{3}y^3 - y^{-1} + c$$

(b) $\displaystyle\int_1^2 y^2 + y^{-2}\,dy$

Recall from our first example above that all we really need here is any anti-derivative of the integrand. We just computed the most general anti-derivative in the first part so we can use that if we want to. However, recall that as we noted above any constants we tack on will just cancel in the long run and so we'll use the answer from (a) without the "$+c$".

Here's the integral,

$$\begin{aligned}\int_1^2 y^2 + y^{-2}\,dy &= \left(\frac{1}{3}y^3 - \frac{1}{y}\right)\Bigg|_1^2 \\ &= \frac{1}{3}(2)^3 - \frac{1}{2} - \left(\frac{1}{3}(1)^3 - \frac{1}{1}\right) \\ &= \frac{8}{3} - \frac{1}{2} - \frac{1}{3} + 1 \\ &= \frac{17}{6}\end{aligned}$$

Remember that the evaluation is always done in the order of evaluation at the upper limit minus evaluation at the lower limit. Also, be very careful with minus signs and parenthesis. It's very easy to forget them or mishandle them and get the wrong answer.

Notice as well that, in order to help with the evaluation, we rewrote the indefinite integral a little. In particular we got rid of the negative exponent on the second term. It's generally easier to evaluate the term with positive exponents.

(c) $\displaystyle\int_{-1}^2 y^2 + y^{-2}\,dy$

This integral is here to make a point. Recall that in order for us to do an integral the integrand must be continuous in the range of the limits. In this case the second term will have division by zero at $y = 0$ and since $y = 0$ is in the interval of integration, *i.e.* it is between the lower and upper limit, this integrand is not continuous in the interval of integration and so we can't do this integral.

Note that this problem will not prevent us from doing the integral in (b) since $y = 0$ is not in the interval of integration.

So, what have we learned from this example?

First, in order to do a definite integral the first thing that we need to do is the indefinite integral. So, we aren't going to get out of doing indefinite integrals, they will be in every integral that we'll be doing in the rest of this course so make sure that you're getting good at computing them.

Second, we need to be on the lookout for functions that aren't continuous at any point between the limits of integration. Also, it's important to note that this will only be a problem if the point(s) of discontinuity occur between the limits of integration or at the limits themselves. If the point of discontinuity occurs outside of the limits of integration the integral can still be evaluated.

In the following sets of examples we won't make too much of an issue with continuity problems, or lack of continuity problems, unless it affects the evaluation of the integral. Do not let this convince you that you don't need to worry about this idea. It arises often enough that it can cause real problems if you aren't on the lookout for it.

Finally, note the difference between indefinite and definite integrals. Indefinite integrals are functions while definite integrals are numbers.

Let's work some more examples.

Example 2

Evaluate each of the following.

(a) $\displaystyle\int_{-3}^{1} 6x^2 - 5x + 2\,dx$

(b) $\displaystyle\int_{4}^{0} \sqrt{t}\,(t-2)\;dt$

(c) $\displaystyle\int_{1}^{2} \frac{2w^5 - w + 3}{w^2}\,dw$

(d) $\displaystyle\int_{25}^{-10} dR$

Solution

(a) $\displaystyle\int_{-3}^{1} 6x^2 - 5x + 2\,dx$

There isn't a lot to this one other than simply doing the work.

$$\begin{aligned}\int_{-3}^{1} 6x^2 - 5x + 2\,dx &= \left(2x^3 - \frac{5}{2}x^2 + 2x\right)\Bigg|_{-3}^{1}\\ &= \left(2 - \frac{5}{2} + 2\right) - \left(-54 - \frac{45}{2} - 6\right)\\ &= 84\end{aligned}$$

(b) $\displaystyle\int_4^0 \sqrt{t}\,(t-2)\ dt$

Recall that we can't integrate products as a product of integrals and so we first need to multiply the integrand out before integrating, just as we did in the indefinite integral case.

$$\begin{aligned}\int_4^0 \sqrt{t}\,(t-2)\ dt &= \int_4^0 t^{\frac{3}{2}} - 2t^{\frac{1}{2}}\,dt\\ &= \left(\frac{2}{5}t^{\frac{5}{2}} - \frac{4}{3}t^{\frac{3}{2}}\right)\Bigg|_4^0\\ &= 0 - \left(\frac{2}{5}(4)^{\frac{5}{2}} - \frac{4}{3}(4)^{\frac{3}{2}}\right)\\ &= -\frac{32}{15}\end{aligned}$$

In the evaluation process recall that,

$$(4)^{\frac{5}{2}} = \left((4)^{\frac{1}{2}}\right)^5 = (2)^5 = 32$$

$$(4)^{\frac{3}{2}} = \left((4)^{\frac{1}{2}}\right)^3 = (2)^3 = 8$$

Also, don't get excited about the fact that the lower limit of integration is larger than the upper limit of integration. That will happen on occasion and there is absolutely nothing wrong with this.

(c) $\displaystyle\int_1^2 \frac{2w^5 - w + 3}{w^2}\,dw$

First, notice that we will have a division by zero issue at $w = 0$, but since this isn't in the interval of integration we won't have to worry about it.

Next again recall that we can't integrate quotients as a quotient of integrals and so the first step that we'll need to do is break up the quotient so we can integrate the function.

$$\begin{aligned}\int_1^2 \frac{2w^5 - w + 3}{w^2}\,dw &= \int_1^2 2w^3 - \frac{1}{w} + 3w^{-2}dw\\ &= \left(\frac{1}{2}w^4 - \ln|w| - \frac{3}{w}\right)\Bigg|_1^2\\ &= \left(8 - \ln(2) - \frac{3}{2}\right) - \left(\frac{1}{2} - \ln 1 - 3\right)\\ &= 9 - \ln(2)\end{aligned}$$

Don't get excited about answers that don't come down to a simple integer or fraction. Often times they won't. Also, don't forget that $\ln(1) = 0$.

(d) $\displaystyle\int_{25}^{-10} dR$

This one is actually pretty easy. Recall that we're just integrating 1.

$$\begin{aligned}\int_{25}^{-10} dR &= R\Big|_{25}^{-10}\\ &= -10-25\\ &= -35\end{aligned}$$

The last set of examples dealt exclusively with integrating powers of x. Let's work a couple of examples that involve other functions.

Example 3

Evaluate each of the following.

(a) $\displaystyle\int_0^1 4x - 6\sqrt[3]{x^2}\,dx$

(b) $\displaystyle\int_0^{\frac{\pi}{3}} 2\sin(\theta) - 5\cos(\theta)\,d\theta$

(c) $\displaystyle\int_{\pi/6}^{\pi/4} 5 - 2\sec(z)\tan(z)\,dz$

(d) $\displaystyle\int_{-20}^{-1} \frac{3}{\mathbf{e}^{-z}} - \frac{1}{3z}\,dz$

(e) $\displaystyle\int_{-2}^{3} 5t^6 - 10t + \frac{1}{t}\,dt$

Solution

(a) $\displaystyle\int_0^1 4x - 6\sqrt[3]{x^2}\,dx$

This one is here mostly here to contrast with the next example.

$$\begin{aligned}\int_0^1 4x - 6\sqrt[3]{x^2}\,dx &= \int_0^1 4x - 6x^{\frac{2}{3}}\,dx\\ &= \left(2x^2 - \frac{18}{5}x^{\frac{5}{3}}\right)\Bigg|_0^1\\ &= 2 - \frac{18}{5} - (0)\\ &= -\frac{8}{5}\end{aligned}$$

(b) $\displaystyle\int_0^{\frac{\pi}{3}} 2\sin(\theta) - 5\cos(\theta)\,d\theta$

Be careful with signs with this one. Recall from the indefinite integral sections that it's easy to mess up the signs when integrating sine and cosine.

$$\begin{aligned}\int_0^{\frac{\pi}{3}} 2\sin(\theta) - 5\cos(\theta)\,d\theta &= \left(-2\cos(\theta) - 5\sin(\theta)\right)\Big|_0^{\pi/3}\\ &= -2\cos\left(\frac{\pi}{3}\right) - 5\sin\left(\frac{\pi}{3}\right) - \left(-2\cos(0) - 5\sin(0)\right)\\ &= -1 - \frac{5\sqrt{3}}{2} + 2\\ &= 1 - \frac{5\sqrt{3}}{2}\end{aligned}$$

Compare this answer to the previous answer, especially the evaluation at zero. It's very easy to get into the habit of just writing down zero when evaluating a function at zero. This is especially a problem when many of the functions that we integrate involve only x's raised to positive integers; these evaluate is zero of course. After evaluating many of these kinds of definite integrals it's easy to get into the habit of just writing down zero when you evaluate at zero. However, there are many functions out there that aren't zero when evaluated at zero so be careful.

(c) $\displaystyle\int_{\pi/6}^{\pi/4} 5 - 2\sec(z)\tan(z)\,dz$

Not much to do other than do the integral.

$$\begin{aligned}\int_{\pi/6}^{\pi/4} 5 - 2\sec(z)\tan(z)\,dz &= \left(5z - 2\sec(z)\right)\Big|_{\pi/6}^{\pi/4}\\ &= 5\left(\frac{\pi}{4}\right) - 2\sec\left(\frac{\pi}{4}\right) - \left(5\left(\frac{\pi}{6}\right) - 2\sec\left(\frac{\pi}{6}\right)\right)\\ &= \frac{5\pi}{12} - 2\sqrt{2} + \frac{4}{\sqrt{3}}\end{aligned}$$

For the evaluation, recall that

$$\sec(z) = \frac{1}{\cos(z)}$$

and so if we can evaluate cosine at these angles we can evaluate secant at these angles.

(d) $\displaystyle\int_{-20}^{-1} \frac{3}{\mathbf{e}^{-z}} - \frac{1}{3z}\, dz$

In order to do this one will need to rewrite both of the terms in the integral a little as follows,

$$\int_{-20}^{-1} \frac{3}{\mathbf{e}^{-z}} - \frac{1}{3z}\, dz = \int_{-20}^{-1} 3\mathbf{e}^{z} - \frac{1}{3}\frac{1}{z}\, dz$$

For the first term recall we used the following fact about exponents.

$$x^{-a} = \frac{1}{x^a} \qquad\qquad \frac{1}{x^{-a}} = x^a$$

In the second term, taking the 3 out of the denominator will just make integrating that term easier.

Now the integral.

$$\begin{aligned}\int_{-20}^{-1} \frac{3}{\mathbf{e}^{-z}} - \frac{1}{3z}\, dz &= \left(3\mathbf{e}^{z} - \frac{1}{3}\ln|z|\right)\Bigg|_{-20}^{-1}\\ &= 3\mathbf{e}^{-1} - \frac{1}{3}\ln|-1| - \left(3\mathbf{e}^{-20} - \frac{1}{3}\ln|-20|\right)\\ &= 3\mathbf{e}^{-1} - 3\mathbf{e}^{-20} + \frac{1}{3}\ln(20)\end{aligned}$$

Just leave the answer like this. It's messy, but it's also exact.

Note that the absolute value bars on the logarithm are required here. Without them we couldn't have done the evaluation.

(e) $\displaystyle\int_{-2}^{3} 5t^6 - 10t + \frac{1}{t}\, dt$

This integral can't be done. There is division by zero in the third term at $t = 0$ and $t = 0$ lies in the interval of integration. The fact that the first two terms can be integrated doesn't matter. If even one term in the integral can't be integrated then the whole integral can't be done.

So, we've computed a fair number of definite integrals at this point. Remember that the vast majority of the work in computing them is first finding the indefinite integral. Once we've found that the rest is just some number crunching.

There are a couple of particularly tricky definite integrals that we need to take a look at next. Actually they are only tricky until you see how to do them, so don't get too excited about them. The first one involves integrating a piecewise function.

Example 4

Given,

$$f(x) = \begin{cases} 6 & \text{if } x > 1 \\ 3x^2 & \text{if } x \le 1 \end{cases}$$

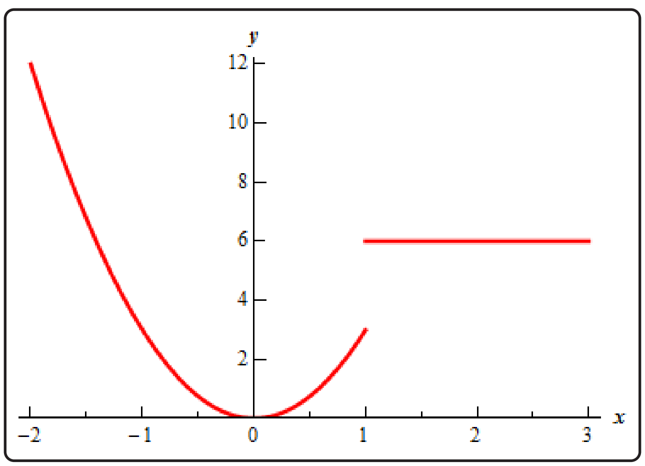
Evaluate each of the following integrals.

(a) $\displaystyle\int_{10}^{22} f(x)\, dx$

(b) $\displaystyle\int_{-2}^{3} f(x)\, dx$

Solution

Let's first start with a graph of this function.



The graph reveals a problem. This function is not continuous at $x = 1$ and we're going to have to watch out for that.

(a) $\displaystyle\int_{10}^{22} f(x)\, dx$

For this integral notice that $x = 1$ is not in the interval of integration and so that is something that we'll not need to worry about in this part.

Also note the limits for the integral lie entirely in the range for the first function. What this means for us is that when we do the integral all we need to do is plug in the first function into the integral.

Here is the integral.

$$\begin{aligned}\int_{10}^{22} f(x)\,dx &= \int_{10}^{22} 6\,dx \\ &= 6x\Big|_{10}^{22} \\ &= 132 - 60 \\ &= 72\end{aligned}$$

(b) $\displaystyle\int_{-2}^{3} f(x)\,dx$

In this part $x = 1$ is between the limits of integration. This means that the integrand is no longer continuous in the interval of integration and that is a show stopper as far we're concerned. As noted above we simply can't integrate functions that aren't continuous in the interval of integration.

Also, even if the function was continuous at $x = 1$ we would still have the problem that the function is actually two different equations depending where we are in the interval of integration.

Let's first address the problem of the function not being continuous at $x = 1$. As we'll see, in this case, if we can find a way around this problem the second problem will also get taken care of at the same time.

In the previous examples where we had functions that weren't continuous we had division by zero and no matter how hard we try we can't get rid of that problem. Division by zero is a real problem and we can't really avoid it. In this case the discontinuity does not stem from problems with the function not existing at $x = 1$. Instead the function is not continuous because it takes on different values on either sides of $x = 1$. We can "remove" this problem by recalling Property 5 from the previous section. This property tells us that we can write the integral as follows,

$$\int_{-2}^{3} f(x)\,dx = \int_{-2}^{1} f(x)\,dx + \int_{1}^{3} f(x)\,dx$$

On each of these intervals the function is continuous. In fact we can say more. In the first integral we will have x between -2 and 1 and this means that we can use the second equation for $f(x)$ and likewise for the second integral x will be between 1 and

3 and so we can use the first function for $f(x)$. The integral in this case is then,

$$\begin{aligned}\int_{-2}^{3} f(x)\,dx &= \int_{-2}^{1} f(x)\,dx + \int_{1}^{3} f(x)\,dx\\ &= \int_{-2}^{1} 3x^2\,dx + \int_{1}^{3} 6\,dx\\ &= x^3\Big|_{-2}^{1} + 6x\Big|_{1}^{3}\\ &= 1-(-8)+(18-6)\\ &= 21\end{aligned}$$

So, to integrate a piecewise function, all we need to do is break up the integral at the break point(s) that happen to occur in the interval of integration and then integrate each piece.

Next, we need to look at is how to integrate an absolute value function.

Example 5

Evaluate the following integral.

$$\int_{0}^{3} |3t-5|\,dt$$

Solution

Recall that the point behind indefinite integration (which we'll need to do in this problem) is to determine what function we differentiated to get the integrand. To this point we've not seen any functions that will differentiate to get an absolute value nor will we ever see a function that will differentiate to get an absolute value.

The only way that we can do this problem is to get rid of the absolute value. To do this we need to recall the definition of absolute value.

$$|x| = \begin{cases} x & \text{if } x \ge 0\\ -x & \text{if } x < 0\end{cases}$$

Once we remember that we can define absolute value as a piecewise function we can use the work from Example 4 as a guide for doing this integral.

What we need to do is determine where the quantity on the inside of the absolute value bars is negative and where it is positive. It looks like if $t > \frac{5}{3}$ the quantity inside the absolute value is positive and if $t < \frac{5}{3}$ the quantity inside the absolute value is negative.

Next, note that $t = \frac{5}{3}$ is in the interval of integration and so, if we break up the integral at this point we get,

$$\int_0^3 |3t-5|\ dt = \int_0^{\frac{5}{3}} |3t-5|\ dt + \int_{\frac{5}{3}}^3 |3t-5|\ dt$$

Now, in the first integral we have $t < \frac{5}{3}$ and so $3t-5 < 0$ in this interval of integration. That means we can drop the absolute value bars if we put in a minus sign. Likewise, in the second integral we have $t > \frac{5}{3}$ which means that in this interval of integration we have $3t-5 > 0$ and so we can just drop the absolute value bars in this integral.

After getting rid of the absolute value bars in each integral we can do each integral. So, doing the integration gives,

$$\begin{aligned}
\int_0^3 |3t-5|\ dt &= \int_0^{\frac{5}{3}} -(3t-5)\ dt + \int_{\frac{5}{3}}^3 3t-5\,dt \\
&= \int_0^{\frac{5}{3}} -3t+5\,dt + \int_{\frac{5}{3}}^3 3t-5\,dt \\
&= \left(-\frac{3}{2}t^2+5t\right)\bigg|_0^{\frac{5}{3}} + \left(\frac{3}{2}t^2-5t\right)\bigg|_{\frac{5}{3}}^3 \\
&= -\frac{3}{2}\left(\frac{5}{3}\right)^2 + 5\left(\frac{5}{3}\right) - (0) + \left(\frac{3}{2}(3)^2 - 5(3) - \left(\frac{3}{2}\left(\frac{5}{3}\right)^2 - 5\left(\frac{5}{3}\right)\right)\right) \\
&= \frac{25}{6} + \frac{8}{3} \\
&= \frac{41}{6}
\end{aligned}$$

Integrating absolute value functions isn't too bad. It's a little more work than the "standard" definite integral, but it's not really all that much more work. First, determine where the quantity inside the absolute value bars is negative and where it is positive. When we've determined that point all we need to do is break up the integral so that in each range of limits the quantity inside the absolute value bars is always positive or always negative. Once this is done we can drop the absolute value bars (adding negative signs when the quantity is negative) and then we can do the integral as we've always done.

Even and Odd Functions

This is the last topic that we need to discuss in this section.

First, recall that an even function is any function which satisfies,

$$f(-x) = f(x)$$

Typical examples of even functions are,

$$f(x) = x^2 \qquad\qquad f(x) = \cos(x)$$

An odd function is any function which satisfies,

$$f(-x) = -f(x)$$

The typical examples of odd functions are,

$$f(x) = x^3 \qquad\qquad f(x) = \sin(x)$$

There are a couple of nice facts about integrating even and odd functions over the interval $[-a, a]$. If $f(x)$ is an even function then,

$$\int_{-a}^{a} f(x)\, dx = 2\int_{0}^{a} f(x)\, dx$$

Likewise, if $f(x)$ is an odd function then,

$$\int_{-a}^{a} f(x)\, dx = 0$$

Note that in order to use these facts the limit of integration must be the same number, but opposite signs!

Example 6

Integrate each of the following.

(a) $\displaystyle\int_{-2}^{2} 4x^4 - x^2 + 1\, dx$

(b) $\displaystyle\int_{-10}^{10} x^5 + \sin(x)\, dx$

Solution

Neither of these are terribly difficult integrals, but we can use the facts on them anyway.

(a) $\displaystyle\int_{-2}^{2} 4x^4 - x^2 + 1\, dx$

In this case the integrand is even and the interval is correct so,

$$\begin{aligned}\int_{-2}^{2} 4x^4 - x^2 + 1\, dx &= 2\int_{0}^{2} 4x^4 - x^2 + 1\, dx \\ &= 2\left(\frac{4}{5}x^5 - \frac{1}{3}x^3 + x\right)\Bigg|_0^2 \\ &= \frac{748}{15}\end{aligned}$$

So, using the fact cut the evaluation in half (in essence since one of the new limits was zero).

(b) $\displaystyle\int_{-10}^{10} x^5 + \sin(x)\ dx$

The integrand in this case is odd and the interval is in the correct form and so we don't even need to integrate. Just use the fact.

$$\int_{-10}^{10} x^5 + \sin(x)\ dx = 0$$

Note that the limits of integration are important here. Take the last integral as an example. A small change to the limits will not give us zero.

$$\int_{-10}^{9} x^5 + \sin(x)\ dx = \cos(10) - \cos(9) - \frac{468559}{6} = -78093.09461$$

The moral here is to be careful and not misuse these facts.

5.8 Substitution Rule for Definite Integrals

We now need to go back and revisit the substitution rule as it applies to definite integrals. At some level there really isn't a lot to do in this section. Recall that the first step in doing a definite integral is to compute the indefinite integral and that hasn't changed. We will still compute the indefinite integral first. This means that we already know how to do these. We use the substitution rule to find the indefinite integral and then do the evaluation.

There are however, two ways to deal with the evaluation step. One of the ways of doing the evaluation is the probably the most obvious at this point, but also has a point in the process where we can get in trouble if we aren't paying attention.

Let's work an example illustrating both ways of doing the evaluation step.

Example 1

Evaluate the following definite integral.

$$\int_{-2}^{0} 2t^2\sqrt{1-4t^3}\,dt$$

Solution

Let's start off looking at the first way of dealing with the evaluation step. We'll need to be careful with this method as there is a point in the process where if we aren't paying attention we'll get the wrong answer.

Solution 1 :

We'll first need to compute the indefinite integral using the substitution rule. Note however, that we will constantly remind ourselves that this is a definite integral by putting the limits on the integral at each step. Without the limits it's easy to forget that we had a definite integral when we've gotten the indefinite integral computed.

In this case the substitution is,

$$u = 1-4t^3 \qquad du = -12t^2dt \qquad \Rightarrow \qquad t^2dt = -\frac{1}{12}du$$

Plugging this into the integral gives,

$$\begin{aligned}\int_{-2}^{0} 2t^2\sqrt{1-4t^3}\,dt &= -\frac{1}{6}\int_{-2}^{0} u^{\frac{1}{2}}\,du\\ &= -\frac{1}{9}u^{\frac{3}{2}}\bigg|_{-2}^{0}\end{aligned}$$

Notice that we didn't do the evaluation yet. This is where the potential problem arises with this solution method. The limits given here are from the original integral and hence are values of t. We have u's in our solution. We can't plug values of t in for u.

Therefore, we will have to go back to t's before we do the substitution. This is the standard step in the substitution process, but it is often forgotten when doing definite integrals. Note as well that in this case, if we don't go back to t's we will have a small problem in that one of the evaluations will end up giving us a complex number.

So, finishing this problem gives,

$$\begin{aligned}\int_{-2}^{0} 2t^2\sqrt{1-4t^3}\,dt &= -\frac{1}{9}\left(1-4t^3\right)^{\frac{3}{2}}\Big|_{-2}^{0}\\ &= -\frac{1}{9}-\left(-\frac{1}{9}(33)^{\frac{3}{2}}\right)\\ &= \frac{1}{9}\left(33\sqrt{33}-1\right)\end{aligned}$$

So, that was the first solution method. Let's take a look at the second method.

Solution 2 :

Note that this solution method isn't really all that different from the first method. In this method we are going to remember that when doing a substitution we want to eliminate all the t's in the integral and write everything in terms of u.

When we say all here we really mean all. In other words, remember that the limits on the integral are also values of t and we're going to convert the limits into u values. Converting the limits is pretty simple since our substitution will tell us how to relate t and u so all we need to do is plug in the original t limits into the substitution and we'll get the new u limits.

Here is the substitution (it's the same as the first method) as well as the limit conversions.

$$\begin{aligned}&u = 1-4t^3 \qquad du = -12t^2 dt \qquad \Rightarrow \qquad t^2 dt = -\frac{1}{12}du\\ &t=-2 \qquad \Rightarrow \qquad u = 1-4(-2)^3 = 33\\ &t=0 \qquad \Rightarrow \qquad u = 1-4(0)^3 = 1\end{aligned}$$

The integral is now,

$$\begin{aligned}\int_{-2}^{0} 2t^2\sqrt{1-4t^3}\,dt &= -\frac{1}{6}\int_{33}^{1} u^{\frac{1}{2}}\,du\\ &= -\frac{1}{9}u^{\frac{3}{2}}\Big|_{33}^{1}\end{aligned}$$

As with the first method let's pause here a moment to remind us what we're doing. In this case, we've converted the limits to u's and we've also got our integral in terms of u's and so here we can just plug the limits directly into our integral. Note that in this case we won't plug our substitution back in. Doing this here would cause problems as we would have t's in the integral and our limits would be u's. Here's the rest of this problem.

$$\begin{aligned}\int_{-2}^{0} 2t^2\sqrt{1-4t^3}\,dt &= -\frac{1}{9}u^{\frac{3}{2}}\bigg|_{33}^{1}\\ &= -\frac{1}{9}-\left(-\frac{1}{9}(33)^{\frac{3}{2}}\right)=\frac{1}{9}\left(33\sqrt{33}-1\right)\end{aligned}$$

We got exactly the same answer and this time didn't have to worry about going back to t's in our answer.

So, we've seen two solution techniques for computing definite integrals that require the substitution rule. Both are valid solution methods and each have their uses. We will be using the second almost exclusively however since it makes the evaluation step a little easier.

Let's work some more examples.

Example 2

Evaluate each of the following.

(a) $\displaystyle\int_{-1}^{5} (1+w)\left(2w+w^2\right)^5 dw$

(b) $\displaystyle\int_{-2}^{-6} \frac{4}{(1+2x)^3}-\frac{5}{1+2x}\,dx$

(c) $\displaystyle\int_{0}^{\frac{1}{2}} \mathbf{e}^{y}+2\cos(\pi y)\;dy$

(d) $\displaystyle\int_{\frac{\pi}{3}}^{0} 3\sin\left(\frac{z}{2}\right)-5\cos(\pi-z)\;dz$

Solution

Since we've done quite a few substitution rule integrals to this time we aren't going to put a lot of effort into explaining the substitution part of things here.

(a) $\displaystyle\int_{-1}^{5} (1+w)\left(2w+w^2\right)^5 dw$

The substitution and converted limits are,

$$u = 2w + w^2 \qquad du = (2+2w)\,dw \qquad \Rightarrow \qquad (1+w)\,dw = \frac{1}{2}du$$
$$w = -1 \quad \Rightarrow \quad u = -1$$
$$w = 5 \quad \Rightarrow \quad u = 35$$

Sometimes a limit will remain the same after the substitution. Don't get excited when it happens and don't expect it to happen all the time.

Here is the integral,

$$\begin{aligned}\int_{-1}^{5} (1+w)\left(2w+w^2\right)^5 dw &= \frac{1}{2}\int_{-1}^{35} u^5\,du \\ &= \left.\frac{1}{12}u^6\right|_{-1}^{35} = 153188802\end{aligned}$$

Don't get excited about large numbers for answers here. Sometimes they are. That's life.

(b) $\displaystyle\int_{-2}^{-6} \frac{4}{(1+2x)^3} - \frac{5}{1+2x}\,dx$

Here is the substitution and converted limits for this problem,

$$u = 1 + 2x \qquad du = 2dx \qquad \Rightarrow \qquad dx = \frac{1}{2}du$$
$$x = -2 \quad \Rightarrow \quad u = -3$$
$$x = -6 \quad \Rightarrow \quad u = -11$$

The integral is then,

$$\begin{aligned}\int_{-2}^{-6} \frac{4}{(1+2x)^3} - \frac{5}{1+2x}\,dx &= \frac{1}{2}\int_{-3}^{-11} 4u^{-3} - \frac{5}{u}\,du \\ &= \left.\frac{1}{2}\left(-2u^{-2} - 5\ln|u|\right)\right|_{-3}^{-11} \\ &= \frac{1}{2}\left(-\frac{2}{121} - 5\ln(11)\right) - \frac{1}{2}\left(-\frac{2}{9} - 5\ln(3)\right) \\ &= \frac{112}{1089} - \frac{5}{2}\ln(11) + \frac{5}{2}\ln(3)\end{aligned}$$

(c) $\displaystyle\int_0^{\frac{1}{2}} \mathbf{e}^y + 2\cos(\pi y)\, dy$

This integral needs to be split into two integrals since the first term doesn't require a substitution and the second does.

$$\int_0^{\frac{1}{2}} \mathbf{e}^y + 2\cos(\pi y)\, dy = \int_0^{\frac{1}{2}} \mathbf{e}^y\, dy + \int_0^{\frac{1}{2}} 2\cos(\pi y)\, dy$$

Here is the substitution and converted limits for the second term.

$$\begin{aligned} &u = \pi y \qquad du = \pi dy \qquad \Rightarrow \qquad dy = \frac{1}{\pi} du \\ &y = 0 \qquad \Rightarrow \qquad u = 0 \\ &y = \frac{1}{2} \qquad \Rightarrow \qquad u = \frac{\pi}{2} \end{aligned}$$

Here is the integral.

$$\begin{aligned} \int_0^{\frac{1}{2}} \mathbf{e}^y + 2\cos(\pi y)\, dy &= \int_0^{\frac{1}{2}} \mathbf{e}^y\, dy + \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \cos(u)\, du \\ &= \mathbf{e}^y\Big|_0^{\frac{1}{2}} + \frac{2}{\pi}\sin(u)\Big|_0^{\frac{\pi}{2}} \\ &= \mathbf{e}^{\frac{1}{2}} - \mathbf{e}^0 + \frac{2}{\pi}\sin\left(\frac{\pi}{2}\right) - \frac{2}{\pi}\sin(0) \\ &= \mathbf{e}^{\frac{1}{2}} - 1 + \frac{2}{\pi} \end{aligned}$$

(d) $\displaystyle\int_{\frac{\pi}{3}}^{0} 3\sin\left(\frac{z}{2}\right) - 5\cos(\pi - z)\, dz$

This integral will require two substitutions. So first split up the integral so we can do a substitution on each term.

$$\int_{\frac{\pi}{3}}^{0} 3\sin\left(\frac{z}{2}\right) - 5\cos(\pi - z)\, dz = \int_{\frac{\pi}{3}}^{0} 3\sin\left(\frac{z}{2}\right) dz - \int_{\frac{\pi}{3}}^{0} 5\cos(\pi - z)\, dz$$

There are the two substitutions for these integrals.

$$\begin{aligned} &u = \frac{z}{2} \qquad du = \frac{1}{2}dz \qquad \Rightarrow \qquad dz = 2du \\ &z = \frac{\pi}{3} \qquad \Rightarrow \qquad u = \frac{\pi}{6} \\ &z = 0 \qquad \Rightarrow \qquad u = 0 \end{aligned}$$

$$\begin{aligned} &v = \pi - z \qquad dv = -dz \qquad \Rightarrow \qquad dz = -dv \\ &z = \frac{\pi}{3} \qquad \Rightarrow \qquad v = \frac{2\pi}{3} \\ &z = 0 \qquad \Rightarrow \qquad v = \pi \end{aligned}$$

Here is the integral for this problem.

$$\begin{aligned}\int_{\frac{\pi}{3}}^{0} 3\sin\left(\frac{z}{2}\right) - 5\cos\left(\pi - z\right)\,dz &= 6\int_{\frac{\pi}{6}}^{0}\sin\left(u\right)\,du + 5\int_{\frac{2\pi}{3}}^{\pi}\cos\left(v\right)\,dv\\ &= -6\cos\left(u\right)\Big|_{\frac{\pi}{6}}^{0} + 5\sin\left(v\right)\Big|_{\frac{2\pi}{3}}^{\pi}\\ &= 3\sqrt{3} - 6 + \left(-\frac{5\sqrt{3}}{2}\right)\\ &= \frac{\sqrt{3}}{2} - 6\end{aligned}$$

The next set of examples is designed to make sure that we don't forget about a very important point about definite integrals.

Example 3

Evaluate each of the following.

(a) $\displaystyle\int_{-5}^{5}\frac{4t}{2-8t^2}\,dt$

(b) $\displaystyle\int_{3}^{5}\frac{4t}{2-8t^2}\,dt$

Solution

(a) $\displaystyle\int_{-5}^{5}\frac{4t}{2-8t^2}\,dt$

Be careful with this integral. The denominator is zero at $t = \pm\frac{1}{2}$ and both of these are in the interval of integration. Therefore, this integrand is not continuous in the interval and so the integral can't be done.

Be careful with definite integrals and be on the lookout for division by zero problems. In the previous section they were easy to spot since all the division by zero problems that we had there were where the variable was itself zero. Once we move into substitution problems however they will not always be so easy to spot so make sure that you first take a quick look at the integrand and see if there are any continuity problems with the integrand and if they occur in the interval of integration.

(b) $\displaystyle\int_3^5 \frac{4t}{2-8t^2}\,dt$

Now, in this case the integral can be done because the two points of discontinuity, $t = \pm\frac{1}{2}$, are both outside of the interval of integration. The substitution and converted limits in this case are,

$$u = 2 - 8t^2 \qquad du = -16t\,dt \qquad \Rightarrow \qquad tdt = -\frac{1}{16}du$$
$$t = 3 \quad \Rightarrow \quad u = -70$$
$$t = 5 \quad \Rightarrow \quad u = -198$$

The integral is then,

$$\begin{aligned}\int_3^5 \frac{4t}{2-8t^2}\,dt &= -\frac{4}{16}\int_{-70}^{-198}\frac{1}{u}\,du\\ &= -\frac{1}{4}\ln|u|\Big|_{-70}^{-198}\\ &= -\frac{1}{4}\left(\ln(198) - \ln(70)\right)\end{aligned}$$

Let's work another set of examples. These are a little tougher (at least in appearance) than the previous sets.

Example 4

Evaluate each of the following.

(a) $\displaystyle\int_0^{\ln(1+\pi)} \mathbf{e}^x \cos\left(1-\mathbf{e}^x\right)dx$

(b) $\displaystyle\int_{\mathbf{e}^2}^{\mathbf{e}^6} \frac{[\ln t]^4}{t}\,dt$

(c) $\displaystyle\int_{\frac{\pi}{12}}^{\frac{\pi}{9}} \frac{\sec(3P)\tan(3P)}{\sqrt[3]{2+\sec(3P)}}\,dP$

(d) $\displaystyle\int_{-\pi}^{\frac{\pi}{2}} \cos(x)\cos(\sin(x))\,dx$

(e) $\displaystyle\int_{\frac{1}{50}}^{2} \frac{\mathbf{e}^{\frac{2}{w}}}{w^2}\,dw$

Solution

(a) $\displaystyle\int_0^{\ln(1+\pi)} \mathbf{e}^x \cos(1-\mathbf{e}^x)\,dx$

The limits are a little unusual in this case, but that will happen sometimes so don't get too excited about it. Here is the substitution.

$$\begin{aligned} &u = 1-\mathbf{e}^x && du = -\mathbf{e}^x dx \\ &x = 0 && \Rightarrow \quad u = 1-\mathbf{e}^0 = 1-1 = 0 \\ &x = \ln(1+\pi) && \Rightarrow \quad u = 1-\mathbf{e}^{\ln(1+\pi)} = 1-(1+\pi) = -\pi \end{aligned}$$

The integral is then,

$$\begin{aligned} \int_0^{\ln(1+\pi)} \mathbf{e}^x \cos(1-\mathbf{e}^x)\,dx &= -\int_0^{-\pi} \cos u\,du \\ &= -\sin(u)\Big|_0^{-\pi} \\ &= -(\sin(-\pi)-\sin 0) = 0 \end{aligned}$$

(b) $\displaystyle\int_{\mathbf{e}^2}^{\mathbf{e}^6} \frac{[\ln t]^4}{t}\,dt$

Here is the substitution and converted limits for this problem.

$$\begin{aligned} &u = \ln t && du = \frac{1}{t}dt \\ &t = \mathbf{e}^2 && \Rightarrow \quad u = \ln \mathbf{e}^2 = 2 \\ &t = \mathbf{e}^6 && \Rightarrow \quad u = \ln \mathbf{e}^6 = 6 \end{aligned}$$

The integral is,

$$\begin{aligned} \int_{\mathbf{e}^2}^{\mathbf{e}^6} \frac{[\ln t]^4}{t}\,dt &= \int_2^6 u^4\,du \\ &= \frac{1}{5}u^5\Big|_2^6 \\ &= \frac{7744}{5} \end{aligned}$$

(c) $\displaystyle\int_{\frac{\pi}{12}}^{\frac{\pi}{9}} \frac{\sec(3P)\tan(3P)}{\sqrt[3]{2+\sec(3P)}}\,dP$

Here is the substitution and converted limits and don't get too excited about the substitution. It's a little messy in the case, but that can happen on occasion.

$$u = 2 + \sec(3P) \qquad du = 3\sec(3P)\tan(3P)\,dP \quad \Rightarrow \quad \sec(3P)\tan(3P)\,dP = \frac{1}{3}du$$
$$P = \frac{\pi}{12} \quad \Rightarrow \quad u = 2 + \sec\left(\frac{\pi}{4}\right) = 2 + \sqrt{2}$$
$$P = \frac{\pi}{9} \quad \Rightarrow \quad u = 2 + \sec\left(\frac{\pi}{3}\right) = 4$$

Here is the integral,

$$\begin{aligned}\int_{\frac{\pi}{12}}^{\frac{\pi}{9}} \frac{\sec(3P)\tan(3P)}{\sqrt[3]{2+\sec(3P)}}\,dP &= \frac{1}{3}\int_{2+\sqrt{2}}^{4} u^{-\frac{1}{3}}\,du \\ &= \frac{1}{2}u^{\frac{2}{3}}\bigg|_{2+\sqrt{2}}^{4} \\ &= \frac{1}{2}\left(4^{\frac{2}{3}} - \left(2+\sqrt{2}\right)^{\frac{2}{3}}\right)\end{aligned}$$

So, not only was the substitution messy, but we also have a messy answer, but again that's life on occasion.

(d) $\displaystyle\int_{-\pi}^{\frac{\pi}{2}} \cos(x)\cos(\sin(x))\,dx$

This problem not as bad as it looks. Here is the substitution and converted limits.

$$u = \sin x \qquad du = \cos x\,dx$$
$$x = \frac{\pi}{2} \quad \Rightarrow \quad u = \sin\frac{\pi}{2} = 1$$
$$x = -\pi \quad \Rightarrow \quad u = \sin(-\pi) = 0$$

The cosine in the very front of the integrand will get substituted away in the differential and so this integrand actually simplifies down significantly. Here is the integral.

$$\begin{aligned}\int_{-\pi}^{\frac{\pi}{2}} \cos(x)\cos(\sin(x))\,dx &= \int_0^1 \cos u\,du \\ &= \sin(u)\Big|_0^1 \\ &= \sin(1) - \sin(0) \\ &= \sin(1)\end{aligned}$$

Don't get excited about these kinds of answers. On occasion we will end up with trig function evaluations like this.

(e) $\displaystyle\int_{\frac{1}{50}}^{2} \frac{\mathbf{e}^{\frac{2}{w}}}{w^2}\,dw$

This is also a tricky substitution (at least until you see it). Here it is,

$$\begin{aligned}
&u = \frac{2}{w} \qquad du = -\frac{2}{w^2}dw \qquad \Rightarrow \qquad \frac{1}{w^2}dw = -\frac{1}{2}du\\
&w = 2 \qquad \Rightarrow \qquad u = 1\\
&w = \frac{1}{50} \qquad \Rightarrow \qquad u = 100
\end{aligned}$$

Here is the integral.

$$\begin{aligned}
\int_{\frac{1}{50}}^{2} \frac{\mathbf{e}^{\frac{2}{w}}}{w^2}\,dw &= -\frac{1}{2}\int_{100}^{1} \mathbf{e}^{u}\,du\\
&= -\frac{1}{2}\mathbf{e}^{u}\Big|_{100}^{1}\\
&= -\frac{1}{2}\left(\mathbf{e}^{1} - \mathbf{e}^{100}\right)
\end{aligned}$$

In this last set of examples we saw some tricky substitutions and messy limits, but these are a fact of life with some substitution problems and so we need to be prepared for dealing with them when they happen.

6 Applications of Integrals

The previous chapter dealt exclusively with the computation of definite and indefinite integrals as well as some discussion of their properties and interpretations. It is now time to start looking at some applications of integrals. Note as well that we should probably say applications of **definite** integrals as that is really what we'll be looking at in this section.

In addition, we should note that there are a lot of different applications of (definite) integrals out there. We will look at the ones that can easily be done with the knowledge we have at our disposal at this point. Once we have covered the next chapter, Integration Techniques, we will be able to take a look at a few more applications of integrals. At this point we would not be able to compute many of the integrals that arise in those later applications.

In this chapter we'll take a look at using integrals to compute the average value of a function and the work required to move an object over a given distance. In addition we will take a look at a couple of geometric applications of integrals. In particular we will use integrals to compute the area that is between two curves and note that this application should not be too surprising given one of the major interpretations of the definite integral. We will also see how to compute the volume of some solids. We will compute the volume of solids of revolution, *i.e.* a solid obtained by rotating a curve about a given axis. In addition, we will compute the volume of some slightly more general solids in which the cross sections can be easily described with nice 2D geometric formulas (*i.e.* rectangles, triangles, circles, *etc.*).

6.1 Average Function Value

The first application of integrals that we'll take a look at is the average value of a function. The following fact tells us how to compute this.

Average Function Value

The average value of a continuous function $f(x)$ over the interval $[a, b]$ is given by,

$$f_{avg} = \frac{1}{b-a}\int_a^b f(x)\,dx$$

To see a justification of this formula see the Proof of Various Integral Properties section of the Extras appendix.

Let's work a couple of quick examples.

Example 1

Determine the average value of each of the following functions on the given interval.

(a) $f(t) = t^2 - 5t + 6\cos(\pi t)$ on $\left[-1, \frac{5}{2}\right]$

(b) $R(z) = \sin(2z)\,\mathbf{e}^{1-\cos(2z)}$ on $[-\pi, \pi]$

Solution

(a) $f(t) = t^2 - 5t + 6\cos(\pi t)$ on $\left[-1, \frac{5}{2}\right]$

There's really not a whole lot to do in this problem other than just use the formula.

$$\begin{aligned} f_{avg} &= \frac{1}{\frac{5}{2}-(-1)}\int_{-1}^{\frac{5}{2}} t^2 - 5t + 6\cos(\pi t)\,dt \\ &= \frac{2}{7}\left(\frac{1}{3}t^3 - \frac{5}{2}t^2 + \frac{6}{\pi}\sin(\pi t)\right)\Bigg|_{-1}^{\frac{5}{2}} \\ &= \frac{12}{7\pi} - \frac{13}{6} \\ &= -1.620993 \end{aligned}$$

You caught the substitution needed for the third term right?

So, the average value of this function of the given interval is -1.620993.

(b) $R(z) = \sin(2z)\,\mathbf{e}^{1-\cos(2z)}$ on $[-\pi, \pi]$

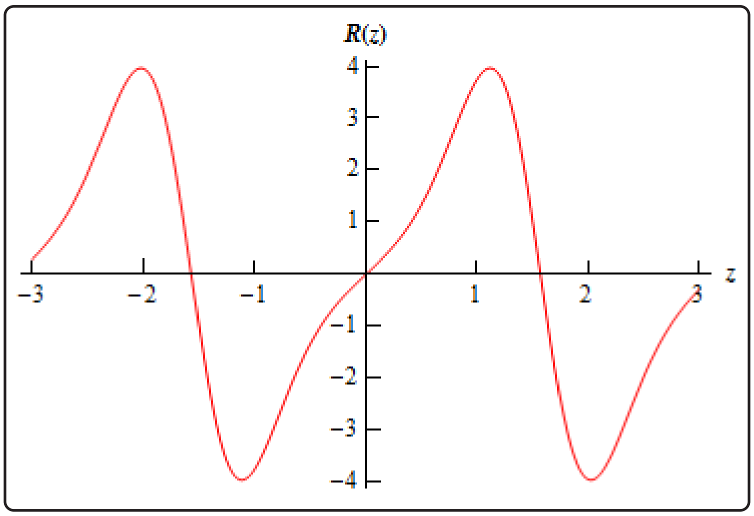
Again, not much to do here other than use the formula. Note that the integral will need the following substitution.

$$u = 1 - \cos(2z)$$

Here is the average value of this function,

$$\begin{aligned} R_{avg} &= \frac{1}{\pi - (-\pi)} \int_{-\pi}^{\pi} \sin(2z)\,\mathbf{e}^{1-\cos(2z)}\,dz \\ &= \frac{1}{4\pi}\mathbf{e}^{1-\cos(2z)}\Big|_{-\pi}^{\pi} \\ &= 0 \end{aligned}$$

So, in this case the average function value is zero. Do not get excited about getting zero here. It will happen on occasion. In fact, if you look at the graph of the function on this interval it's not too hard to see that this is the correct answer.



There is also a theorem that is related to the average function value.

The Mean Value Theorem for Integrals

If $f(x)$ is a continuous function on $[a, b]$ then there is a number c in $[a, b]$ such that,

$$\int_a^b f(x)\,dx = f(c)(b-a)$$

Note that this is very similar to the Mean Value Theorem that we saw in the Derivatives Applications chapter. See the Proof of Various Integral Properties section of the Extras appendix for the proof.

Note that one way to think of this theorem is the following. First rewrite the result as,

$$\frac{1}{b-a}\int_a^b f(x)\,dx = f(c)$$

and from this we can see that this theorem is telling us that there is a number $a < c < b$ such that $f_{avg} = f(c)$. Or, in other words, if $f(x)$ is a continuous function then somewhere in $[a, b]$ the function will take on its average value.

Let's take a quick look at an example using this theorem.

Example 2

Determine the number c that satisfies the Mean Value Theorem for Integrals for the function $f(x) = x^2 + 3x + 2$ on the interval $[1, 4]$.

Solution

First let's notice that the function is a polynomial and so is continuous on the given interval. This means that we can use the Mean Value Theorem. So, let's do that.

$$\begin{aligned}
\int_1^4 x^2 + 3x + 2\,dx &= \left(c^2 + 3c + 2\right)(4-1) \\
\left(\frac{1}{3}x^3 + \frac{3}{2}x^2 + 2x\right)\Big|_1^4 &= 3\left(c^2 + 3c + 2\right) \\
\frac{99}{2} &= 3c^2 + 9c + 6 \\
0 &= 3c^2 + 9c - \frac{87}{2}
\end{aligned}$$

This is a quadratic equation that we can solve. Using the quadratic formula we get the following two solutions,

$$\begin{aligned}
c &= \frac{-3+\sqrt{67}}{2} = 2.593 \\
c &= \frac{-3-\sqrt{67}}{2} = -5.593
\end{aligned}$$

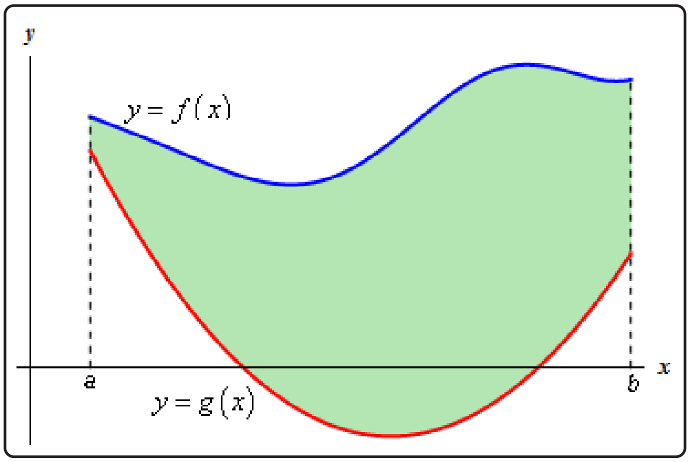
Clearly the second number is not in the interval and so that isn't the one that we're after. The first however is in the interval and so that's the number we want.

Note that it is possible for both numbers to be in the interval so don't expect only one to be in the interval.

6.2 Area Between Curves

In this section we are going to look at finding the area between two curves. There are actually two cases that we are going to be looking at.

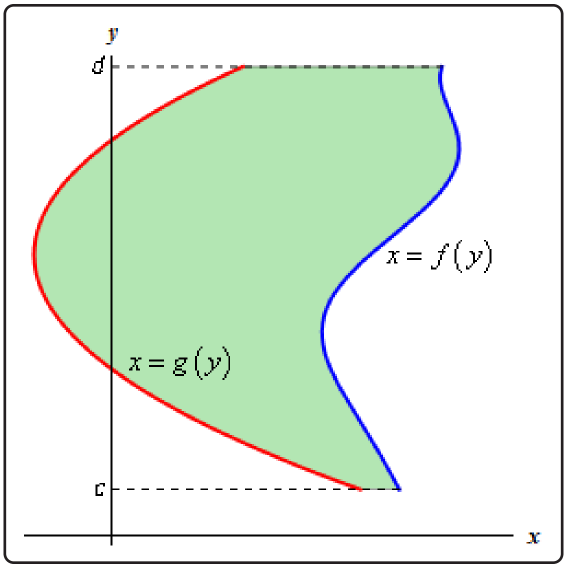
In the first case we want to determine the area between $y = f(x)$ and $y = g(x)$ on the interval $[a, b]$. We are also going to assume that $f(x) \geq g(x)$. Take a look at the following sketch to get an idea of what we're initially going to look at.



In the Area and Volume Formulas section of the Extras appendix we derived the following formula for the area in this case.

$$A = \int_a^b f(x) - g(x)\, dx \tag{6.1}$$

The second case is almost identical to the first case. Here we are going to determine the area between $x = f(y)$ and $x = g(y)$ on the interval $[c, d]$ with $f(y) \geq g(y)$.



In this case the formula is,

$$A = \int_c^d f(y) - g(y)\, dy \tag{6.2}$$

Now Equation 6.1 and Equation 6.2 are perfectly serviceable formulas, however, it is sometimes easy to forget that these always require the first function to be the larger of the two functions. So, instead of these formulas we will instead use the following "word" formulas to make sure that we remember that the area is always the "larger" function minus the "smaller" function.

In the first case we will use,

Area Between Curves, Case 1

$$A = \int_a^b \left(\begin{array}{c}\text{upper}\\ \text{function}\end{array}\right) - \left(\begin{array}{c}\text{lower}\\ \text{function}\end{array}\right) dx, \qquad a \le x \le b \tag{6.3}$$

In the second case we will use,

Area Between Curves, Case 2

$$A = \int_c^d \left(\begin{array}{c}\text{right}\\ \text{function}\end{array}\right) - \left(\begin{array}{c}\text{left}\\ \text{function}\end{array}\right) dy, \qquad c \le y \le d \tag{6.4}$$

Using these formulas will always force us to think about what is going on with each problem and to make sure that we've got the correct order of functions when we go to use the formula.

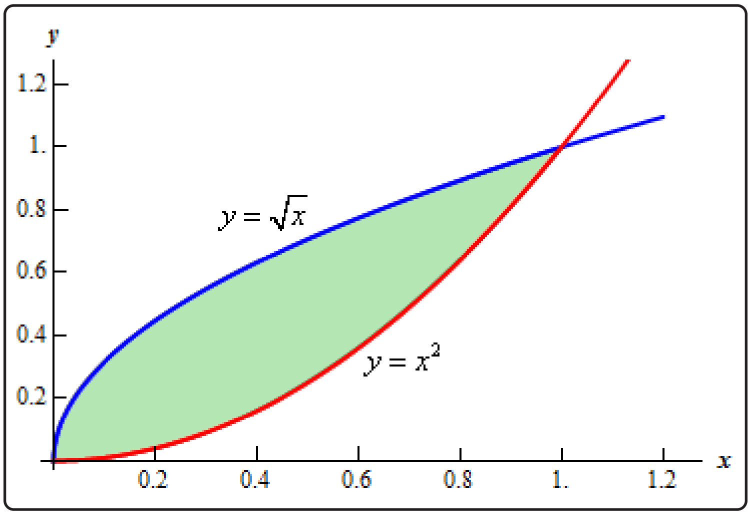
Let's work an example.

Example 1

Determine the area of the region enclosed by $y = x^2$ and $y = \sqrt{x}$.

Solution

First of all, just what do we mean by "area enclosed by". This means that the region we're interested in must have one of the two curves on every boundary of the region. So, here is a graph of the two functions with the enclosed region shaded.



Note that we don't take any part of the region to the right of the rightmost intersection point of these two graphs. In this region there is no boundary on the right side and so this region is not part of the enclosed area. Remember that one of the given functions must be on the boundary of the enclosed region.

Also, from this graph it's clear that the upper function will be dependent on the range of x's that we use. Because of this you should always sketch of a graph of the region. Without a sketch it's often easy to mistake which of the two functions is the larger. In this case most would probably say that $y = x^2$ is the upper function and they would be right for the vast majority of the x's. However, in this case it is the lower of the two functions.

The limits of integration for this will be the intersection points of the two curves. In this case it's pretty easy to see that they will intersect at $x = 0$ and $x = 1$ so these are the limits of integration.

So, the integral that we'll need to compute to find the area is,

$$\begin{aligned} A &= \int_a^b \left(\begin{array}{c}\text{upper}\\ \text{function}\end{array}\right) - \left(\begin{array}{c}\text{lower}\\ \text{function}\end{array}\right) dx \\ &= \int_0^1 \sqrt{x} - x^2 \, dx \\ &= \left.\left(\frac{2}{3}x^{\frac{3}{2}} - \frac{1}{3}x^3\right)\right|_0^1 \\ &= \frac{1}{3} \end{aligned}$$

Before moving on to the next example, there are a couple of important things to note.

First, in almost all of these problems a graph is pretty much required. Often the bounding region, which will give the limits of integration, is difficult to determine without a graph.

Also, it can often be difficult to determine which of the functions is the upper function and which is the lower function without a graph. This is especially true in cases like the last example where the answer to that question actually depended upon the range of x's that we were using.

Finally, unlike the area under a curve that we looked at in the previous chapter the area between two curves will always be positive. If we get a negative number or zero we can be sure that we've made a mistake somewhere and will need to go back and find it.

Note as well that sometimes instead of saying region enclosed by we will say region bounded by. They mean the same thing.

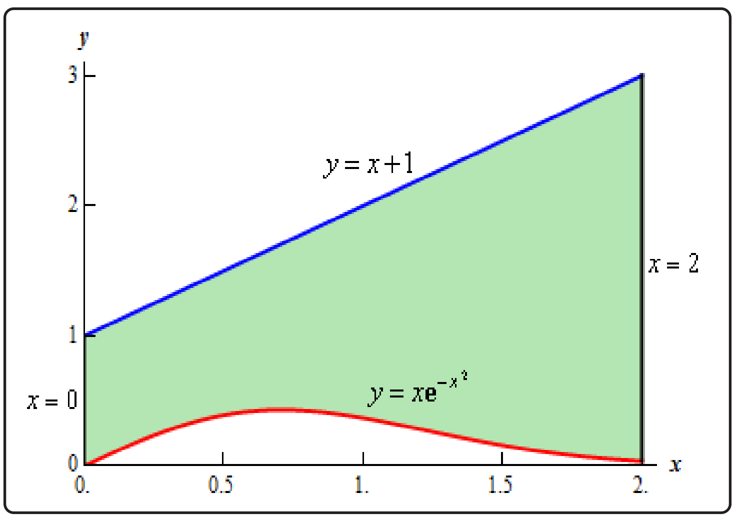
Let's work some more examples.

Example 2

Determine the area of the region bounded by $y = x\mathbf{e}^{-x^2}$, $y = x + 1$, $x = 2$, and the y-axis.

Solution

In this case the last two pieces of information, $x = 2$ and the y-axis, tell us the right and left boundaries of the region. Also, recall that the y-axis is given by the line $x = 0$. Here is the graph with the enclosed region shaded in.



Here, unlike the first example, the two curves don't meet. Instead we rely on two vertical lines to bound the left and right sides of the region as we noted above

Here is the integral that will give the area.

$$\begin{aligned} A &= \int_a^b \left(\begin{array}{c}\text{upper}\\ \text{function}\end{array}\right) - \left(\begin{array}{c}\text{lower}\\ \text{function}\end{array}\right) dx \\ &= \int_0^2 x + 1 - x\mathbf{e}^{-x^2}\, dx \\ &= \left(\frac{1}{2}x^2 + x + \frac{1}{2}\mathbf{e}^{-x^2}\right)\Big|_0^2 \\ &= \frac{7}{2} + \frac{\mathbf{e}^{-4}}{2} = 3.5092 \end{aligned}$$

Example 3

Determine the area of the region bounded by $y = 2x^2 + 10$ and $y = 4x + 16$.

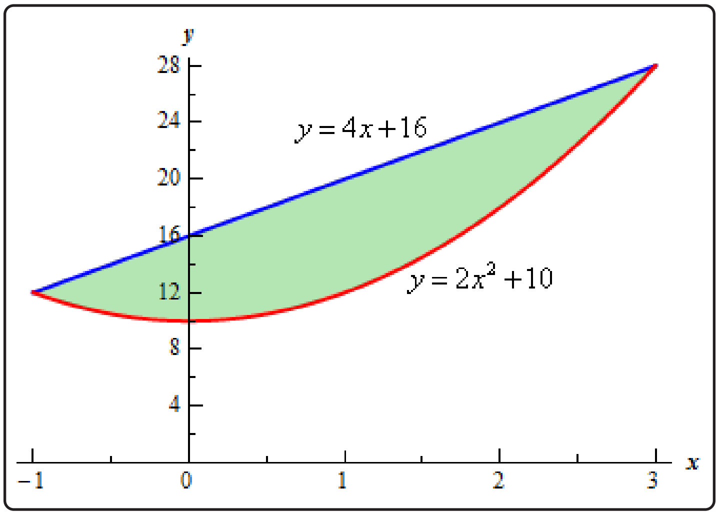
Solution

In this case the intersection points (which we'll need eventually) are not going to be easily identified from the graph so let's go ahead and get them now. Note that for most of these problems you'll not be able to accurately identify the intersection points from the graph and so you'll need to be able to determine them by hand. In this case we can get the intersection points by setting the two equations equal.

$$\begin{aligned} 2x^2 + 10 &= 4x + 16 \\ 2x^2 - 4x - 6 &= 0 \\ 2(x+1)(x-3) &= 0 \end{aligned}$$

So, it looks like the two curves will intersect at $x = -1$ and $x = 3$. If we need them we can get the y values corresponding to each of these by plugging the values back into either of the equations. We'll leave it to you to verify that the coordinates of the two intersection points on the graph are $(-1, 12)$ and $(3, 28)$.

Note as well that if you aren't good at graphing knowing the intersection points can help in at least getting the graph started. Here is a graph of the region.



With the graph we can now identify the upper and lower function and so we can now find the enclosed area.

$$A = \int_a^b \begin{pmatrix} \text{upper} \\ \text{function} \end{pmatrix} - \begin{pmatrix} \text{lower} \\ \text{function} \end{pmatrix} dx$$

$$= \int_{-1}^{3} 4x + 16 - \left(2x^2 + 10\right) dx$$

$$= \int_{-1}^{3} -2x^2 + 4x + 6\,dx$$

$$= \left(-\frac{2}{3}x^3 + 2x^2 + 6x\right)\Big|_{-1}^{3}$$

$$= \frac{64}{3}$$

Be careful with parenthesis in these problems. One of the more common mistakes students make with these problems is to neglect parenthesis on the second term.

Example 4

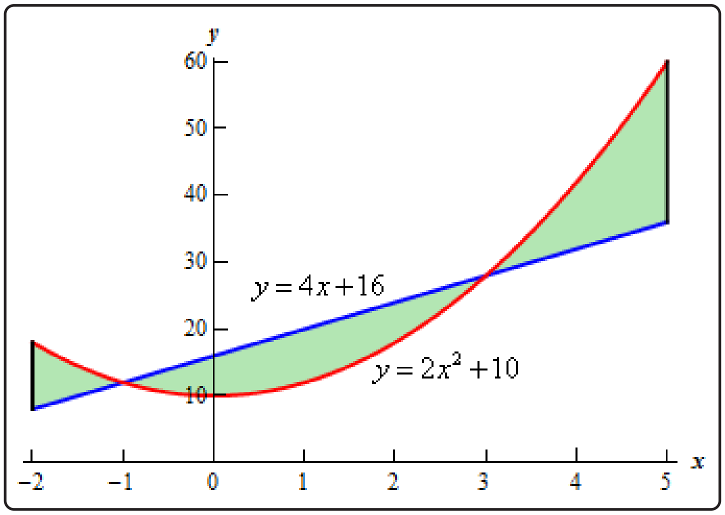
Determine the area of the region bounded by $y = 2x^2 + 10$, $y = 4x + 16$, $x = -2$ and $x = 5$.

Solution

So, the functions used in this problem are identical to the functions from the first problem. The difference is that we've extended the bounded region out from the intersection points.

Since these are the same functions we used in the previous example we won't bother finding the intersection points again.

Here is a graph of this region.



Okay, we have a small problem here. Our formula requires that one function always be the upper function and the other function always be the lower function and we clearly do not have that here. However, this actually isn't the problem that it might at first appear to be. There are three regions in which one function is always the upper function and the other is always the lower function. So, all that we need to do is find the area of each of the three regions, which we can do, and then add them all up.

Here is the area.

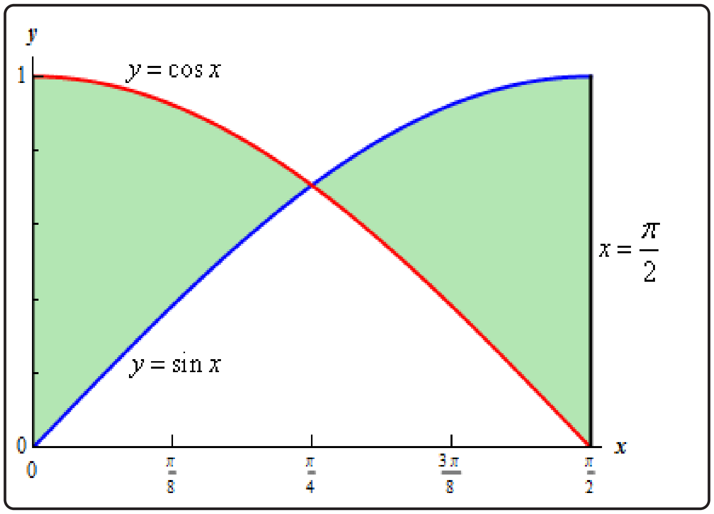
$$\begin{aligned}
A &= \int_{-2}^{-1} 2x^2 + 10 - (4x + 16)\, dx + \int_{-1}^{3} 4x + 16 - \left(2x^2 + 10\right)\, dx \\
&\qquad\qquad\qquad\qquad + \int_{3}^{5} 2x^2 + 10 - (4x + 16)\, dx \\
&= \int_{-2}^{-1} 2x^2 - 4x - 6\, dx + \int_{-1}^{3} -2x^2 + 4x + 6\, dx + \int_{3}^{5} 2x^2 - 4x - 6\, dx \\
&= \left(\frac{2}{3}x^3 - 2x^2 - 6x\right)\bigg|_{-2}^{-1} + \left(-\frac{2}{3}x^3 + 2x^2 + 6x\right)\bigg|_{-1}^{3} + \left(\frac{2}{3}x^3 - 2x^2 - 6x\right)\bigg|_{3}^{5} \\
&= \frac{14}{3} + \frac{64}{3} + \frac{64}{3} \\
&= \frac{142}{3}
\end{aligned}$$

Example 5

Determine the area of the region enclosed by $y = \sin(x)$, $y = \cos(x)$, $x = \frac{\pi}{2}$, and the y-axis.

Solution

First let's get a graph of the region.



So, we have another situation where we will need to do two integrals to get the area. The intersection point will be where

$$\sin(x) = \cos(x)$$

in the interval. We'll leave it to you to verify that this will be $x = \frac{\pi}{4}$. The area is then,

$$\begin{aligned} A &= \int_0^{\frac{\pi}{4}} \cos(x) - \sin(x)\,dx + \int_{\pi/4}^{\pi/2} \sin(x) - \cos(x)\,dx \\ &= \left(\sin(x) + \cos(x)\right)\Big|_0^{\frac{\pi}{4}} + \left(-\cos(x) - \sin(x)\right)\Big|_{\pi/4}^{\pi/2} \\ &= \sqrt{2} - 1 + \left(\sqrt{2} - 1\right) \\ &= 2\sqrt{2} - 2 = 0.828427 \end{aligned}$$

We will need to be careful with this next example.

Example 6

Determine the area of the region enclosed by $x = \frac{1}{2}y^2 - 3$ and $y = x - 1$.

Solution

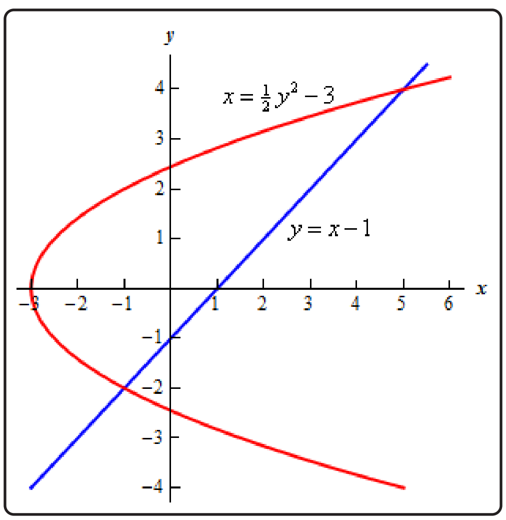
Don't let the first equation get you upset. We will have to deal with these kinds of equations occasionally so we'll need to get used to dealing with them.

As always, it will help if we have the intersection points for the two curves. In this case we'll get the intersection points by solving the second equation for x and then setting them equal. Here is that work,

$$\begin{aligned} y + 1 &= \frac{1}{2}y^2 - 3 \\ 2y + 2 &= y^2 - 6 \\ 0 &= y^2 - 2y - 8 \\ 0 &= (y - 4)(y + 2) \end{aligned}$$

So, it looks like the two curves will intersect at $y = -2$ and $y = 4$ or if we need the full coordinates they will be : $(-1, -2)$ and $(5, 4)$.

Here is a sketch of the two curves.



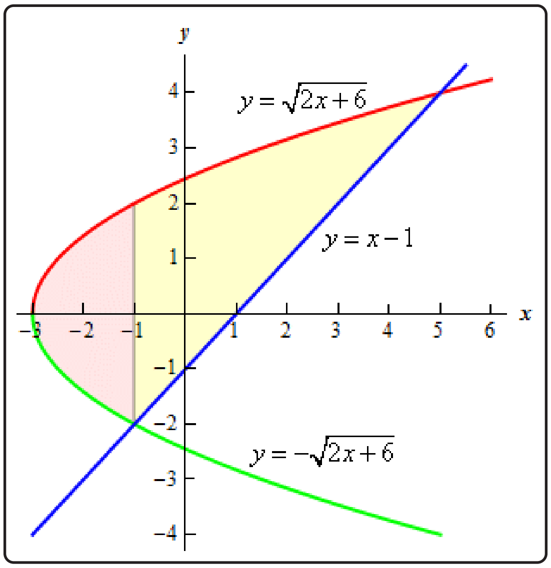
Now, we will have a serious problem at this point if we aren't careful. To this point we've been using an upper function and a lower function. To do that here notice that there are actually two portions of the region that will have different lower functions. In the range $[-3, -1]$ the parabola is actually both the upper and the lower function.

To use the formula that we've been using to this point we need to solve the parabola for y. This gives,

$$y = \pm\sqrt{2x+6}$$

where the "$+$" gives the upper portion of the parabola and the "-" gives the lower portion.

Here is a sketch of the complete area with each region shaded that we'd need if we were going to use the first formula.



The integrals for the area would then be,

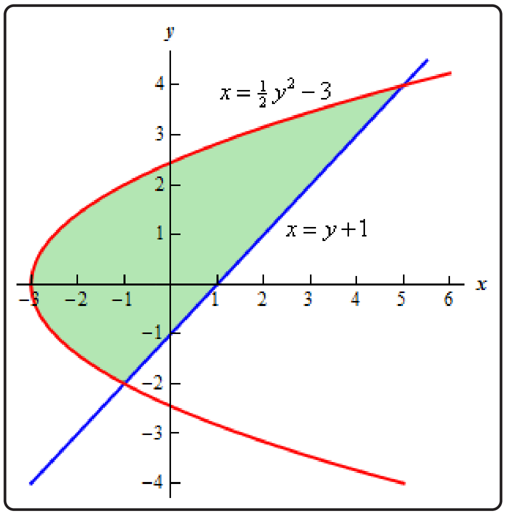
$$\begin{aligned} A &= \int_{-3}^{-1} \sqrt{2x+6} - \left(-\sqrt{2x+6}\right)\,dx + \int_{-1}^{5} \sqrt{2x+6} - (x-1)\,dx \\ &= \int_{-3}^{-1} 2\sqrt{2x+6}\,dx + \int_{-1}^{5} \sqrt{2x+6} - x + 1\,dx \\ &= \int_{-3}^{-1} 2\sqrt{2x+6}\,dx + \int_{-1}^{5} \sqrt{2x+6}\,dx + \int_{-1}^{5} -x + 1\,dx \\ &= \frac{2}{3}u^{\frac{3}{2}}\Big|_{0}^{4} + \frac{1}{3}u^{\frac{3}{2}}\Big|_{4}^{16} + \left(-\frac{1}{2}x^2 + x\right)\Big|_{-1}^{5} \\ &= 18 \end{aligned}$$

While these integrals aren't terribly difficult they are more difficult than they need to be.

Recall that there is another formula for determining the area. It is,

$$A = \int_{c}^{d} \begin{pmatrix}\text{right}\\\text{function}\end{pmatrix} - \begin{pmatrix}\text{left}\\\text{function}\end{pmatrix}\,dy, \qquad c \le y \le d$$

and in our case we do have one function that is always on the left and the other is always on the right. So, in this case this is definitely the way to go. Note that we will need to rewrite the equation of the line since it will need to be in the form $x = f(y)$ but that is easy enough to do. Here is the graph for using this formula.



The area is,

$$\begin{aligned} A &= \int_c^d \left(\begin{array}{c}\text{right}\\\text{function}\end{array}\right) - \left(\begin{array}{c}\text{left}\\\text{function}\end{array}\right) dy \\ &= \int_{-2}^{4} (y+1) - \left(\frac{1}{2}y^2 - 3\right) dy \\ &= \int_{-2}^{4} -\frac{1}{2}y^2 + y + 4\, dy \\ &= \left(-\frac{1}{6}y^3 + \frac{1}{2}y^2 + 4y\right)\Bigg|_{-2}^{4} \\ &= 18 \end{aligned}$$

This is the same that we got using the first formula and this was definitely easier than the first method.

So, in this last example we've seen a case where we could use either formula to find the area. However, the second was definitely easier.

Students often come into a calculus class with the idea that the only easy way to work with functions is to use them in the form $y = f(x)$. However, as we've seen in this previous example there are definitely times when it will be easier to work with functions in the form $x = f(y)$. In fact, there are going to be occasions when this will be the only way in which a problem can be worked so make sure that you can deal with functions in this form.

Let's take a look at one more example to make sure we can deal with functions in this form.

Example 7

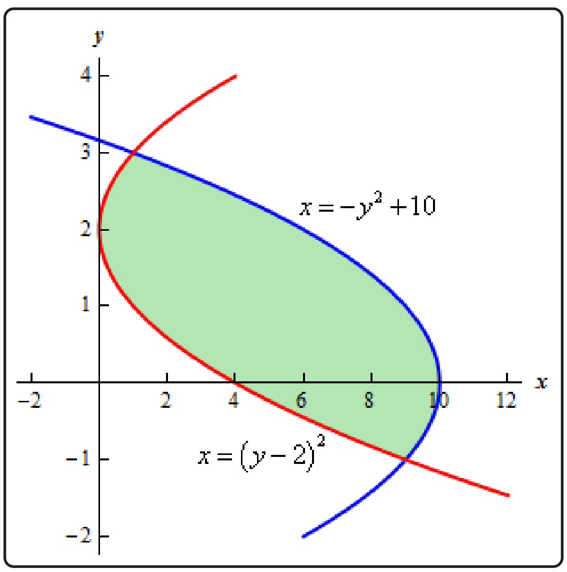
Determine the area of the region bounded by $x = -y^2 + 10$ and $x = (y-2)^2$.

Solution

First, we will need intersection points.

$$\begin{aligned} -y^2 + 10 &= (y-2)^2 \\ -y^2 + 10 &= y^2 - 4y + 4 \\ 0 &= 2y^2 - 4y - 6 \\ 0 &= 2(y+1)(y-3) \end{aligned}$$

The intersection points are $y = -1$ and $y = 3$. Here is a sketch of the region.



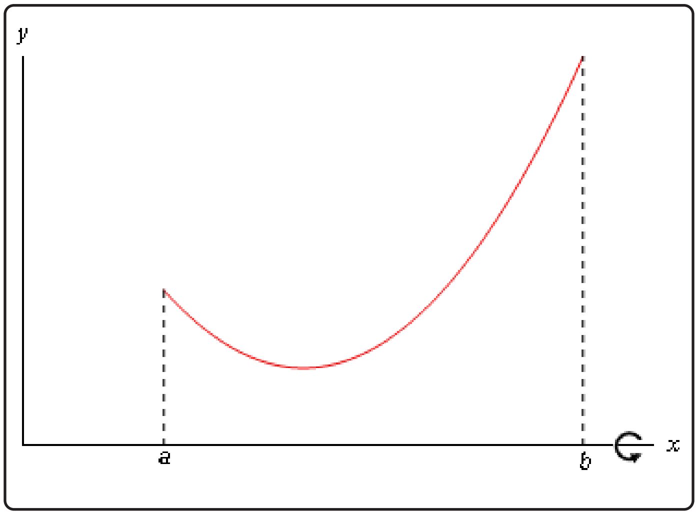
This is definitely a region where the second area formula will be easier. If we used the first formula there would be three different regions that we'd have to look at.

The area in this case is,

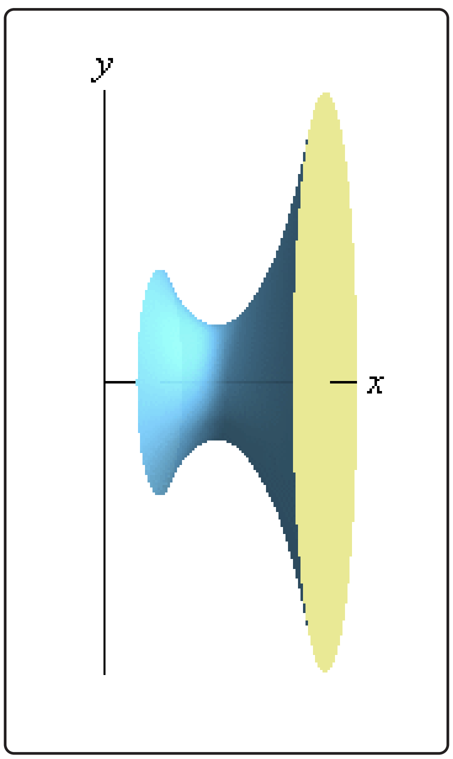
$$\begin{aligned} A &= \int_c^d \begin{pmatrix} \text{right} \\ \text{function} \end{pmatrix} - \begin{pmatrix} \text{left} \\ \text{function} \end{pmatrix} dy \\ &= \int_{-1}^{3} -y^2 + 10 - (y-2)^2 \, dy \\ &= \int_{-1}^{3} -2y^2 + 4y + 6 \, dy \\ &= \left(-\frac{2}{3}y^3 + 2y^2 + 6y \right) \Bigg|_{-1}^{3} = \frac{64}{3} \end{aligned}$$

6.3 Solids of Revolution / Method of Rings

In this section we will start looking at the volume of a solid of revolution. We should first define just what a solid of revolution is. To get a solid of revolution we start out with a function, $y = f(x)$, on an interval $[a, b]$.



We then rotate this curve about a given axis to get the surface of the solid of revolution. For purposes of this discussion let's rotate the curve about the x-axis, although it could be any vertical or horizontal axis. Doing this for the curve above gives the following three dimensional region.



What we want to do over the course of the next two sections is to determine the volume of this object.

In the Area and Volume Formulas section of the Extras appendix we derived the following formulas for the volume of this solid.

Volume Formulas

$$V = \int_a^b A(x)\,dx \qquad\qquad V = \int_c^d A(y)\,dy$$

where, $A(x)$ and $A(y)$ are the cross-sectional area functions of the solid. There are many ways to get the cross-sectional area and we'll see two (or three depending on how you look at it) over the next two sections. Whether we will use $A(x)$ or $A(y)$ will depend upon the method and the axis of rotation used for each problem.

One of the easier methods for getting the cross-sectional area is to cut the object perpendicular to the axis of rotation. Doing this the cross section will be either a solid disk if the object is solid (as our above example is) or a ring if we've hollowed out a portion of the solid (we will see this eventually).

In the case that we get a solid disk the area is,

$$A = \pi(\text{radius})^2$$

where the radius will depend upon the function and the axis of rotation.

In the case that we get a ring the area is,

Area of Ring

$$A = \pi\left(\left(\begin{array}{c}\text{outer}\\ \text{radius}\end{array}\right)^2 - \left(\begin{array}{c}\text{inner}\\ \text{radius}\end{array}\right)^2\right)$$

where again both of the radii will depend on the functions given and the axis of rotation. Note as well that in the case of a solid disk we can think of the inner radius as zero and we'll arrive at the correct formula for a solid disk and so this is a much more general formula to use.

Also, in both cases, whether the area is a function of x or a function of y will depend upon the axis of rotation as we will see.

This method is often called the **method of disks** or the **method of rings**.

Let's do an example.

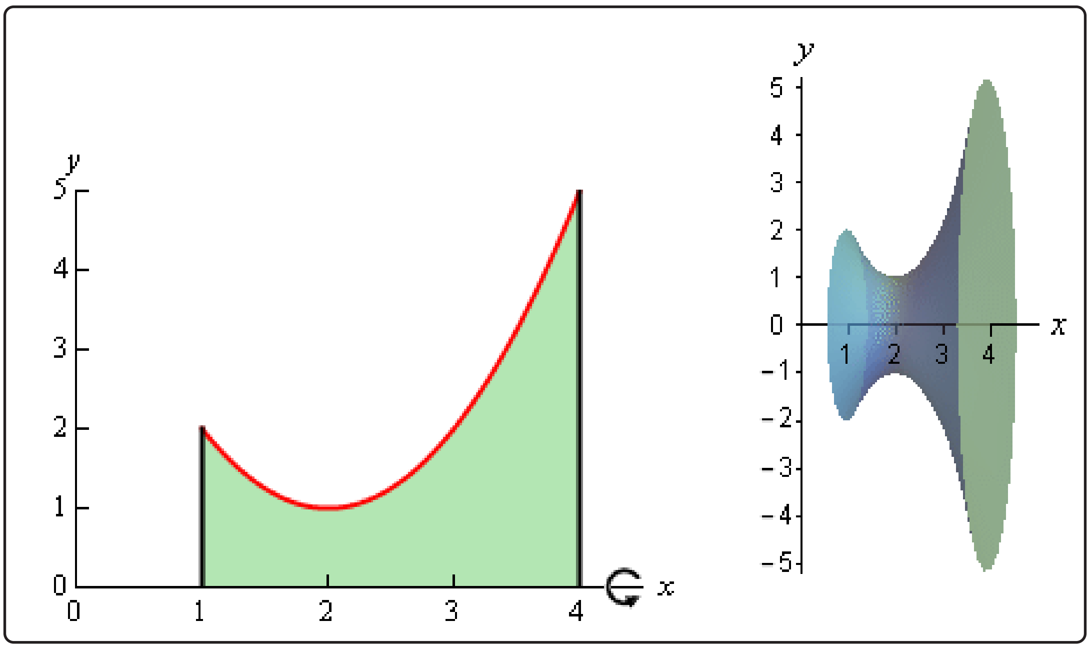
Example 1

Determine the volume of the solid obtained by rotating the region bounded by $y = x^2 - 4x + 5$, $x = 1$, $x = 4$, and the x-axis about the x-axis.

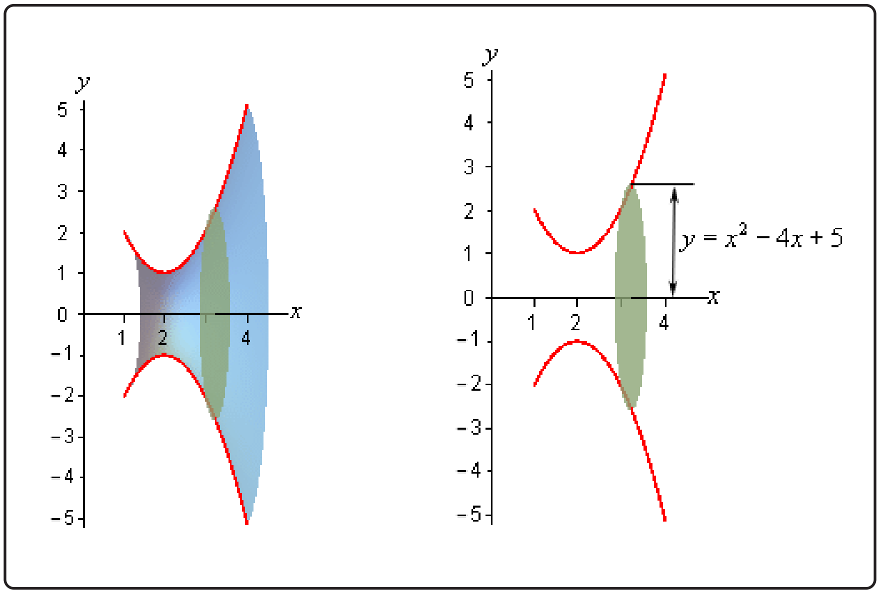
Solution

The first thing to do is get a sketch of the bounding region and the solid obtained by rotating the region about the x-axis. We don't need a picture perfect sketch of the curves we just need something that will allow us to get a feel for what the bounded region looks like so we can get a quick sketch of the solid. With that in mind we can note that the first equation is just a parabola with vertex $(2, 1)$ (you do remember how to get the vertex of a parabola right?) and opens upward and so we don't really need to put a lot of time into sketching it.

Here are both of these sketches.



Okay, to get a cross section we cut the solid at any x. Below are a couple of sketches showing a typical cross section. The sketch on the right shows a cut away of the object with a typical cross section without the caps. The sketch on the left shows just the curve we're rotating as well as its mirror image along the bottom of the solid.



In this case the radius is simply the distance from the x-axis to the curve and this is nothing more than the function value at that particular x as shown above. The cross-sectional area is then,

$$A(x) = \pi\left(x^2 - 4x + 5\right)^2 = \pi\left(x^4 - 8x^3 + 26x^2 - 40x + 25\right)$$

Next, we need to determine the limits of integration. Working from left to right the first cross section will occur at $x = 1$ and the last cross section will occur at $x = 4$. These are the limits of integration.

The volume of this solid is then,

$$\begin{aligned} V &= \int_a^b A(x)\, dx \\ &= \pi \int_1^4 x^4 - 8x^3 + 26x^2 - 40x + 25\, dx \\ &= \pi \left(\frac{1}{5}x^5 - 2x^4 + \frac{26}{3}x^3 - 20x^2 + 25x\right)\Bigg|_1^4 \\ &= \frac{78\pi}{5} \end{aligned}$$

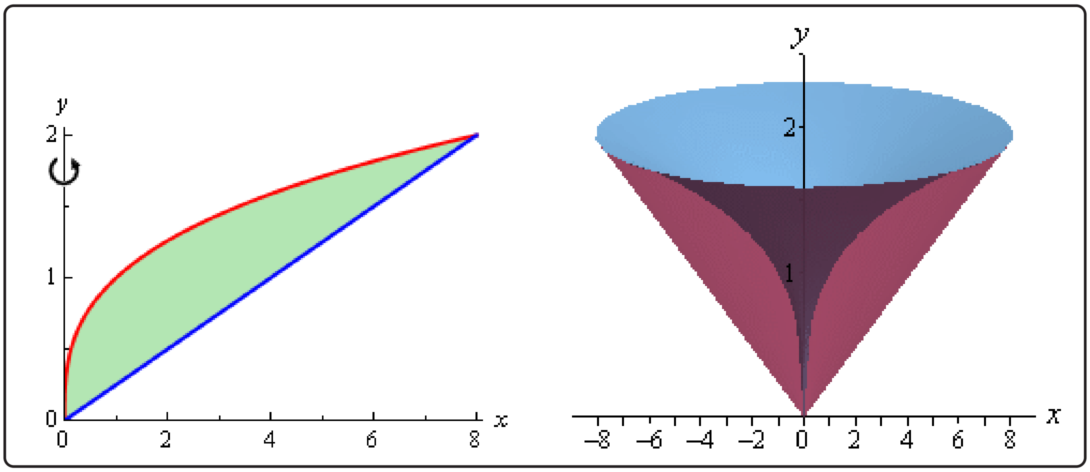
In the above example the object was a solid object, but the more interesting objects are those that are not solid so let's take a look at one of those.

Example 2

Determine the volume of the solid obtained by rotating the portion of the region bounded by $y = \sqrt[3]{x}$ and $y = \frac{x}{4}$ that lies in the first quadrant about the y-axis.

Solution

First, let's get a graph of the bounding region and a graph of the object. Remember that we only want the portion of the bounding region that lies in the first quadrant. There is a portion of the bounding region that is in the third quadrant as well, but we don't want that for this problem.

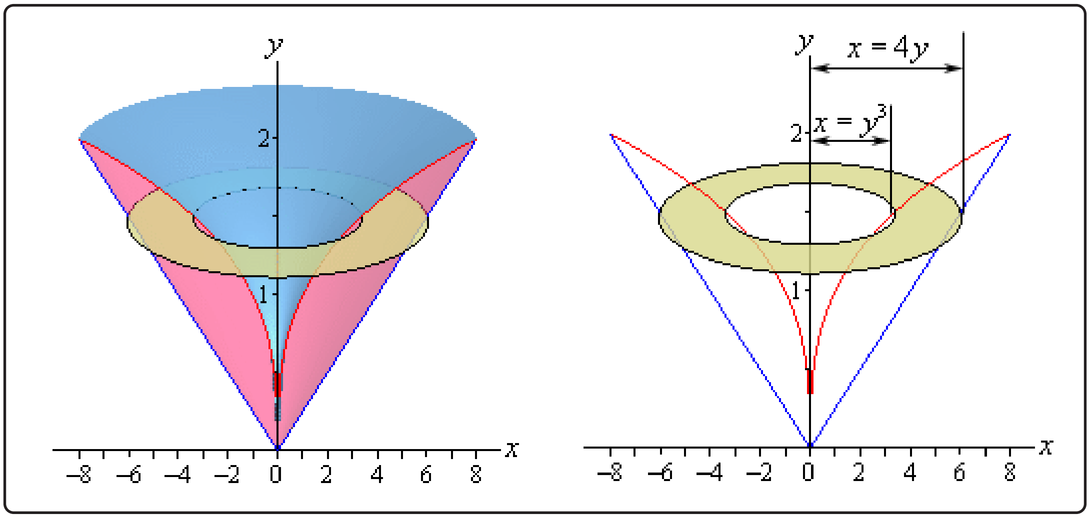


There are a couple of things to note with this problem. First, we are only looking for the volume of the "walls" of this solid, not the complete interior as we did in the last example.

Next, we will get our cross section by cutting the object perpendicular to the axis of rotation. The cross section will be a ring (remember we are only looking at the walls) for this example and it will be horizontal at some y. This means that the inner and outer radius for the ring will be x values and so we will need to rewrite our functions into the form $x = f(y)$. Here are the functions written in the correct form for this example.

$$y = \sqrt[3]{x} \qquad \Rightarrow \qquad x = y^3$$
$$y = \frac{x}{4} \qquad \Rightarrow \qquad x = 4y$$

Here are a couple of sketches of the boundaries of the walls of this object as well as a typical ring. The sketch on the left includes the back portion of the object to give a little context to the figure on the right.



The inner radius in this case is the distance from the y-axis to the inner curve while the outer radius is the distance from the y-axis to the outer curve. Both of these are then x distances and so are given by the equations of the curves as shown above.

The cross-sectional area is then,

$$A(y) = \pi\left((4y)^2 - \left(y^3\right)^2\right) = \pi\left(16y^2 - y^6\right)$$

Working from the bottom of the solid to the top we can see that the first cross-section will occur at $y = 0$ and the last cross-section will occur at $y = 2$. These will be the limits of integration. The volume is then,

$$\begin{aligned} V &= \int_c^d A(y)\, dy \\ &= \pi \int_0^2 16y^2 - y^6\, dy \\ &= \pi\left(\frac{16}{3}y^3 - \frac{1}{7}y^7\right)\Bigg|_0^2 \\ &= \frac{512\pi}{21} \end{aligned}$$

With these two examples out of the way we can now make a generalization about this method. If we rotate about a horizontal axis (the x-axis for example) then the cross-sectional area will be a function of x. Likewise, if we rotate about a vertical axis (the y-axis for example) then the cross-sectional area will be a function of y.

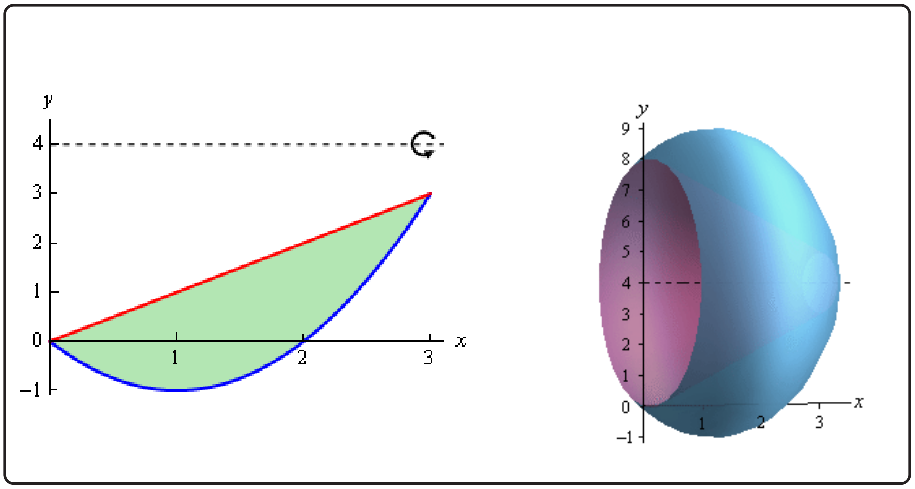
The remaining two examples in this section will make sure that we don't get too used to the idea of always rotating about the x or y-axis.

Example 3

Determine the volume of the solid obtained by rotating the region bounded by $y = x^2 - 2x$ and $y = x$ about the line $y = 4$.

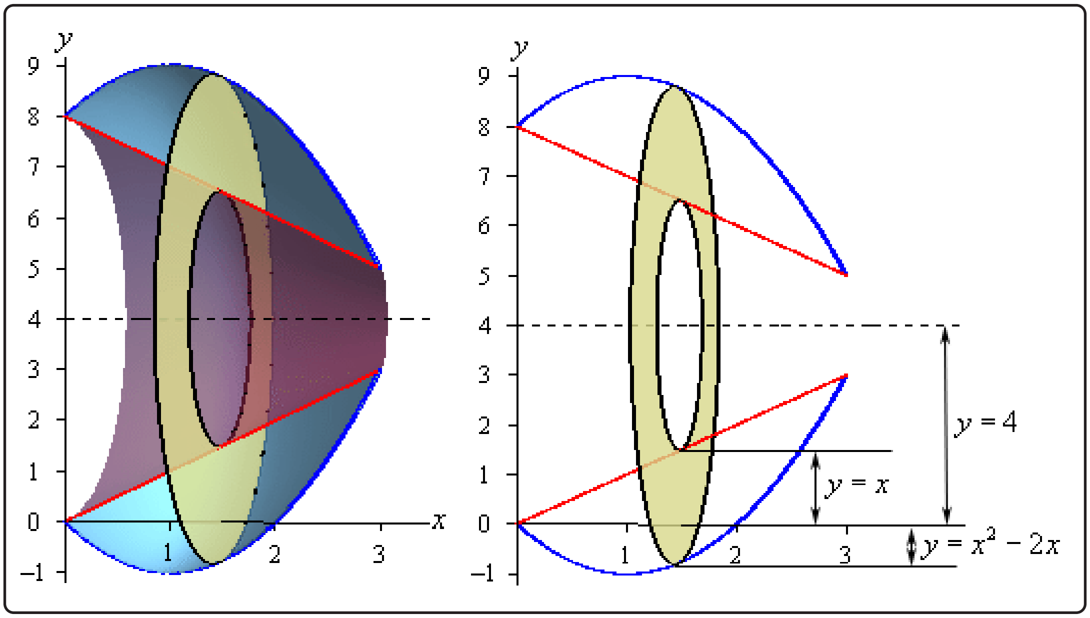
Solution

First let's get the bounding region and the solid graphed.



Again, we are going to be looking for the volume of the walls of this object. Also, since we are rotating about a horizontal axis we know that the cross-sectional area will be a function of x.

Here are a couple of sketches of the boundaries of the walls of this object as well as a typical ring. The sketch on the left includes the back portion of the object to give a little context to the figure on the right.



Now, we're going to have to be careful here in determining the inner and outer radius as they aren't going to be quite as simple they were in the previous two examples.

Let's start with the inner radius as this one is a little clearer. First, the inner radius is NOT x. The distance from the x-axis to the inner edge of the ring is x, but we want the radius and that is the distance from the axis of rotation to the inner edge of the ring. So, we know that the distance from the axis of rotation to the x-axis is 4 and the distance from the x-axis to the inner ring is x. The inner radius must then be the difference between these two. Or,

$$\text{inner radius} = 4 - x$$

The outer radius works the same way. The outer radius is,

$$\text{outer radius} = 4 - \left(x^2 - 2x\right) = -x^2 + 2x + 4$$

Note that given the location of the typical ring in the sketch above the formula for the outer radius may not look quite right but it is in fact correct. As sketched the outer edge of the ring is below the x-axis and at this point the value of the function will be negative and so when we do the subtraction in the formula for the outer radius we'll actually be subtracting off a negative number which has the net effect of adding this distance onto 4 and that gives the correct outer radius. Likewise, if the outer edge is above the x-axis, the function value will be positive and so we'll be doing an honest subtraction here and again we'll get the correct radius in this case.

The cross-sectional area for this case is,

$$A\left(x\right) = \pi\left(\left(-x^2 + 2x + 4\right)^2 - \left(4 - x\right)^2\right) = \pi\left(x^4 - 4x^3 - 5x^2 + 24x\right)$$

The first ring will occur at $x = 0$ and the last ring will occur at $x = 3$ and so these are our limits of integration. The volume is then,

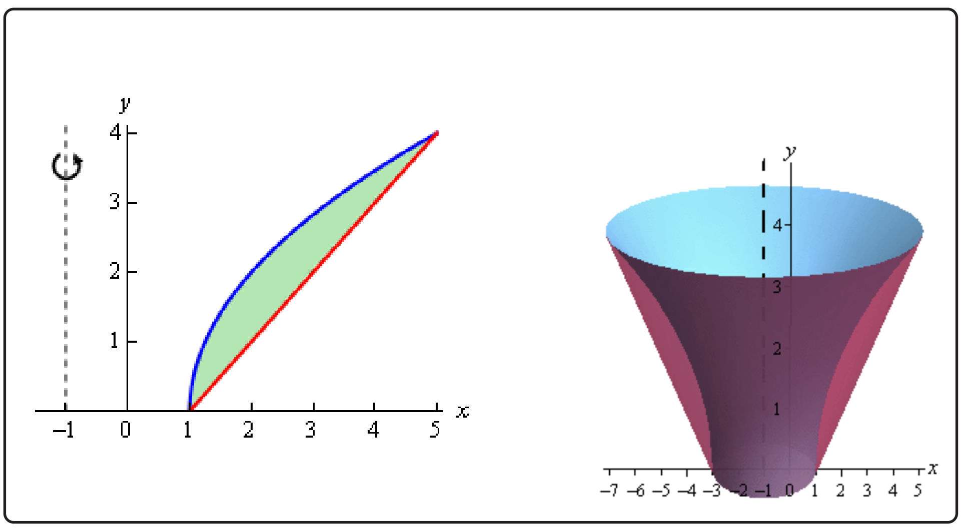
$$\begin{aligned} V &= \int_a^b A(x)\,dx \\ &= \pi \int_0^3 x^4 - 4x^3 - 5x^2 + 24x\,dx \\ &= \pi \left(\frac{1}{5}x^5 - x^4 - \frac{5}{3}x^3 + 12x^2 \right)\Bigg|_0^3 \\ &= \frac{153\pi}{5} \end{aligned}$$

Example 4

Determine the volume of the solid obtained by rotating the region bounded by $y = 2\sqrt{x-1}$ and $y = x - 1$ about the line $x = -1$.

Solution

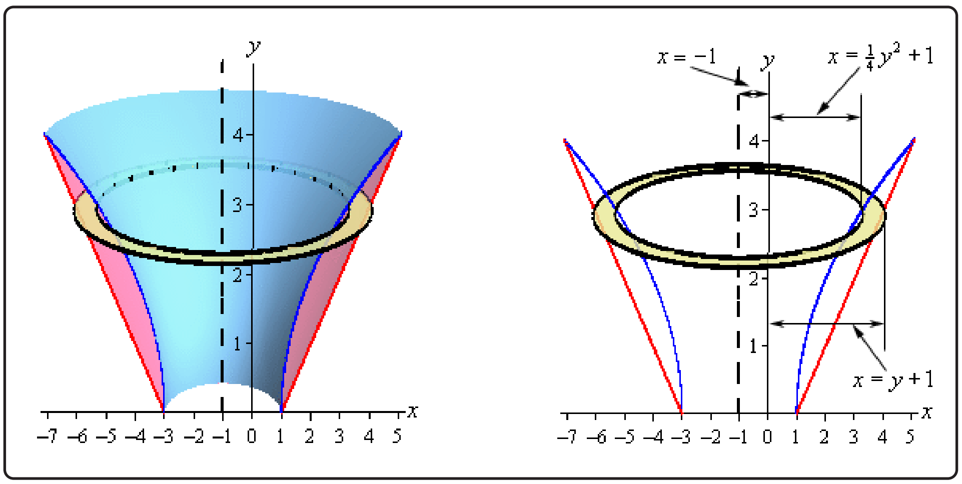
As with the previous examples, let's first graph the bounded region and the solid.



Now, let's notice that since we are rotating about a vertical axis and so the cross-sectional area will be a function of y. This also means that we are going to have to rewrite the functions to also get them in terms of y.

$$y = 2\sqrt{x-1} \qquad \Rightarrow \qquad x = \frac{y^2}{4} + 1$$
$$y = x - 1 \qquad \Rightarrow \qquad x = y + 1$$

Here are a couple of sketches of the boundaries of the walls of this object as well as a typical ring. The sketch on the left includes the back portion of the object to give a little context to the figure on the right.



The inner and outer radius for this case is both similar and different from the previous example. This example is similar in the sense that the radii are not just the functions. In this example the functions are the distances from the y-axis to the edges of the rings. The center of the ring however is a distance of 1 from the y-axis. This means that the distance from the center to the edges is a distance from the axis of rotation to the y-axis (a distance of 1) and then from the y-axis to the edge of the rings.

So, the radii are then the functions plus 1 and that is what makes this example different from the previous example. Here we had to add the distance to the function value whereas in the previous example we needed to subtract the function from this distance. Note that without sketches the radii on these problems can be difficult to get.

So, in summary, we've got the following for the inner and outer radius for this example.

$$\text{outer radius} = y + 1 + 1 = y + 2$$
$$\text{inner radius} = \frac{y^2}{4} + 1 + 1 = \frac{y^2}{4} + 2$$

The cross-sectional area is then,

$$A(y) = \pi\left((y+2)^2 - \left(\frac{y^2}{4}+2\right)^2\right) = \pi\left(4y - \frac{y^4}{16}\right)$$

The first ring will occur at $y = 0$ and the final ring will occur at $y = 4$ and so these will be our limits of integration.

The volume is,

$$\begin{aligned} V &= \int_c^d A(y)\, dy \\ &= \pi \int_0^4 4y - \frac{y^4}{16}\, dy \\ &= \pi\left(2y^2 - \frac{1}{80}y^5\right)\Big|_0^4 \\ &= \frac{96\pi}{5} \end{aligned}$$

6.4 Solids of Revolution / Method of Cylinders

In the previous section we started looking at finding volumes of solids of revolution. In that section we took cross sections that were rings or disks, found the cross-sectional area and then used the following formulas to find the volume of the solid.

$$V = \int_a^b A(x)\,dx \qquad\qquad V = \int_c^d A(y)\,dy$$

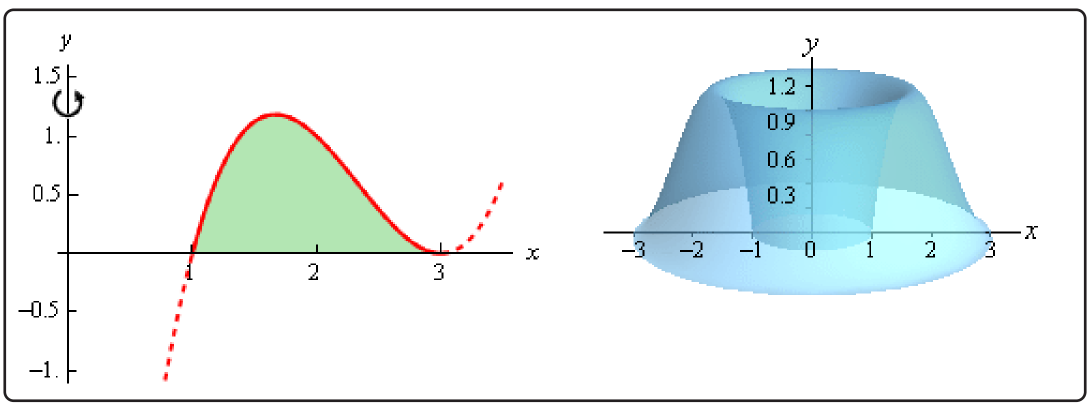
In the previous section we only used cross sections that were in the shape of a disk or a ring. This however does not always need to be the case. We can use any shape for the cross sections as long as it can be expanded or contracted to completely cover the solid we're looking at. This is a good thing because as our first example will show us we can't always use rings/disks.

Example 1

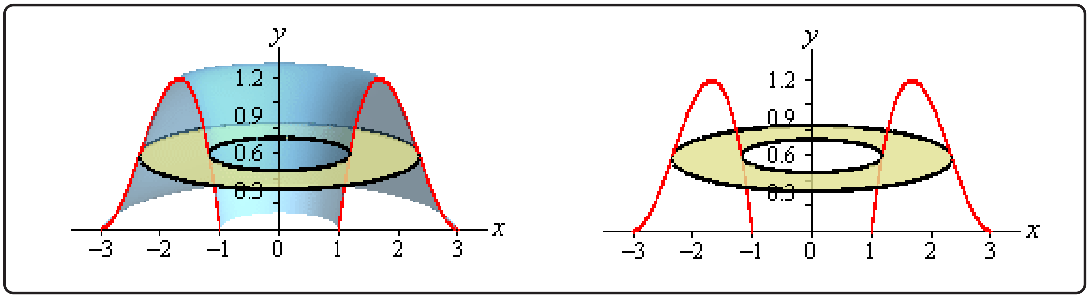
Determine the volume of the solid obtained by rotating the region bounded by $y = (x-1)(x-3)^2$ and the x-axis about the y-axis.

Solution

As we did in the previous section, let's first graph the bounded region and solid. Note that the bounded region here is the shaded portion shown. The curve is extended out a little past this for the purposes of illustrating what the curve looks like.



So, we've basically got something that's roughly doughnut shaped. If we were to use rings on this solid here is what a typical ring would look like.



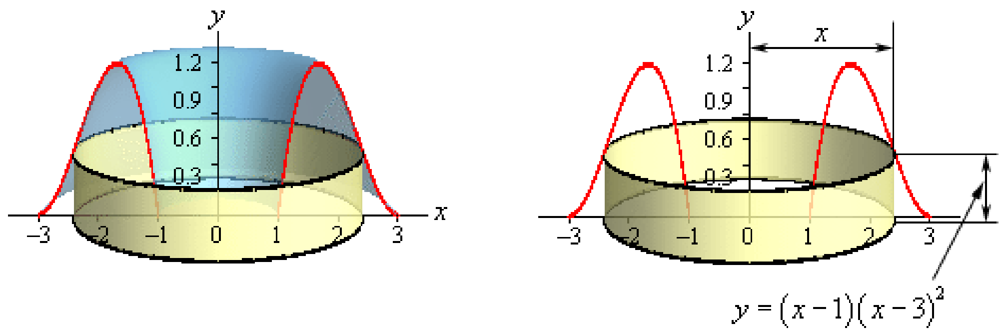
This leads to several problems. First, both the inner and outer radius are defined by the same function. This, in itself, can be dealt with on occasion as we saw in a example in the Area Between Curves section. However, this usually means more work than other methods so it's often not the best approach.

This leads to the second problem we got here. In order to use rings we would need to put this function into the form $x = f(y)$. That is NOT easy in general for a cubic polynomial and in other cases may not even be possible to do. Even when it is possible to do this the resulting equation is often significantly messier than the original which can also cause problems.

The last problem with rings in this case is not so much a problem as it's just added work. If we were to use rings the limit would be y limits and this means that we will need to know how high the graph goes. To this point the limits of integration have always been intersection points that were fairly easy to find. However, in this case the highest point is not an intersection point, but instead a relative maximum. We spent several sections in the Applications of Derivatives chapter talking about how to find maximum values of functions. However, finding them can, on occasion, take some work.

So, we've seen three problems with rings in this case that will either increase our work load or outright prevent us from using rings.

What we need to do is to find a different way to cut the solid that will give us a cross-sectional area that we can work with. One way to do this is to think of our solid as a lump of cookie dough and instead of cutting it perpendicular to the axis of rotation we could instead center a cylindrical cookie cutter on the axis of rotation and push this down into the solid. Doing this would give the following picture,



Doing this gives us a cylinder or shell in the object and we can easily find its surface area. The surface area of this cylinder is,

$$\begin{aligned} A(x) &= 2\pi\,(\text{radius})\,(\text{height}) \\ &= 2\pi\,(x)\left((x-1)(x-3)^2\right) \\ &= 2\pi\left(x^4 - 7x^3 + 15x^2 - 9x\right) \end{aligned}$$

Notice as well that as we increase the radius of the cylinder we will completely cover the solid and so we can use this in our formula to find the volume of this solid. All we need are limits of integration. The first cylinder will cut into the solid at $x = 1$ and as we increase x to $x = 3$ we will completely cover both sides of the solid since expanding the cylinder in one direction will automatically expand it in the other direction as well.

The volume of this solid is then,

$$\begin{aligned} V &= \int_a^b A(x)\,dx \\ &= 2\pi \int_1^3 x^4 - 7x^3 + 15x^2 - 9x\,dx \\ &= 2\pi\left(\frac{1}{5}x^5 - \frac{7}{4}x^4 + 5x^3 - \frac{9}{2}x^2\right)\Bigg|_1^3 \\ &= \frac{24\pi}{5} \end{aligned}$$

The method used in the last example is called the **method of cylinders** or **method of shells**. The formula for the area in all cases will be,

Area of Cylinder

$$A = 2\pi\,(\text{radius})\,(\text{height})$$

There are a couple of important differences between this method and the method of rings/disks that we should note before moving on. First, rotation about a vertical axis will give an area that is a function of x and rotation about a horizontal axis will give an area that is a function of y. This is exactly opposite of the method of rings/disks.

Second, we don't take the complete range of x or y for the limits of integration as we did in the previous section. Instead we take a range of x or y that will cover one side of the solid. As we noted in the first example if we expand out the radius to cover one side we will automatically expand in the other direction as well to cover the other side.

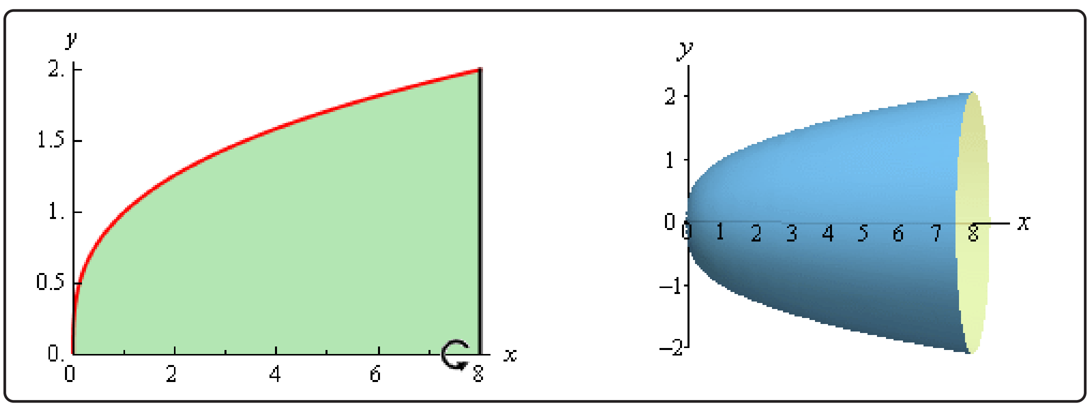
Let's take a look at another example.

Example 2

Determine the volume of the solid obtained by rotating the region bounded by $y = \sqrt[3]{x}$, $x = 8$ and the x-axis about the x-axis.

Solution

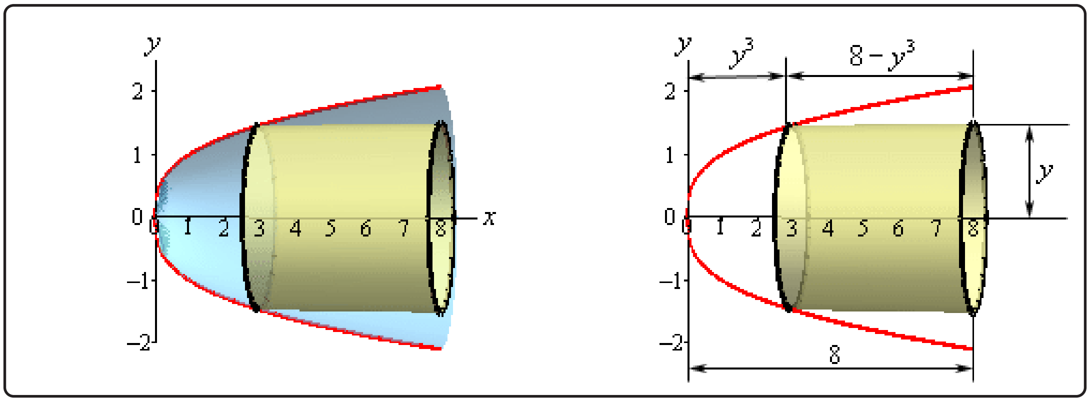
First let's get a graph of the bounded region and the solid.



Okay, we are rotating about a horizontal axis. This means that the area will be a function of y and so our equation will also need to be written in $x = f(y)$ form.

$$y = \sqrt[3]{x} \qquad \Rightarrow \qquad x = y^3$$

As we did in the ring/disk section let's take a couple of looks at a typical cylinder. The sketch on the left shows a typical cylinder with the back half of the object also in the sketch to give the right sketch some context. The sketch on the right contains a typical cylinder and only the curves that define the edge of the solid.



In this case the width of the cylinder is not the function value as it was in the previous example. In this case the function value is the distance between the edge of the cylinder and the y-axis. The distance from the edge out to the line is $x = 8$ and so the width is then $8 - y^3$. The cross-sectional area in this case is,

$$\begin{aligned} A(y) &= 2\pi\,(\text{radius})\,(\text{width}) \\ &= 2\pi\,(y)\left(8 - y^3\right) \\ &= 2\pi\left(8y - y^4\right) \end{aligned}$$

The first cylinder will cut into the solid at $y = 0$ and the final cylinder will cut in at $y = 2$ and so these are our limits of integration.

The volume of this solid is,

$$\begin{aligned} V &= \int_c^d A(y)\,dy \\ &= 2\pi \int_0^2 8y - y^4\,dy \\ &= 2\pi\left(4y^2 - \frac{1}{5}y^5\right)\Big|_0^2 \\ &= \frac{96\pi}{5} \end{aligned}$$

The remaining examples in this section will have axis of rotation about axis other than the x and

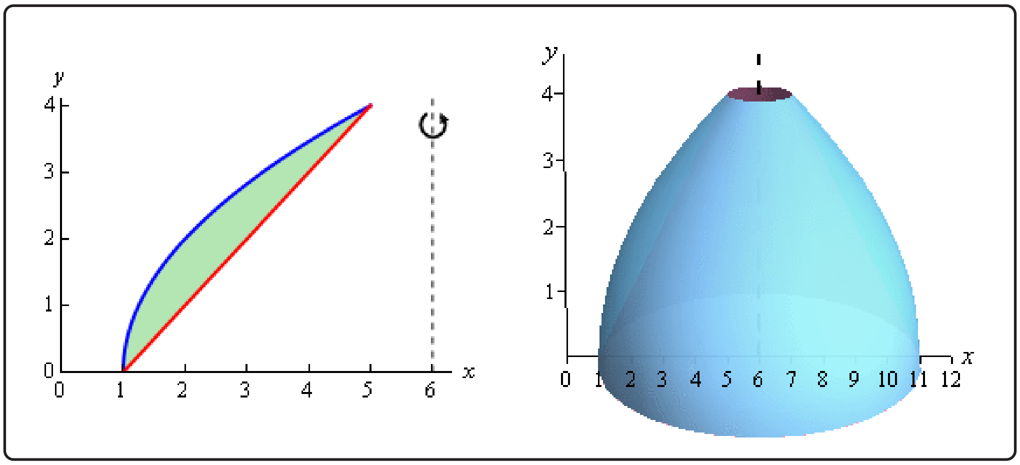
y-axis. As with the method of rings/disks we will need to be a little careful with these.

Example 3

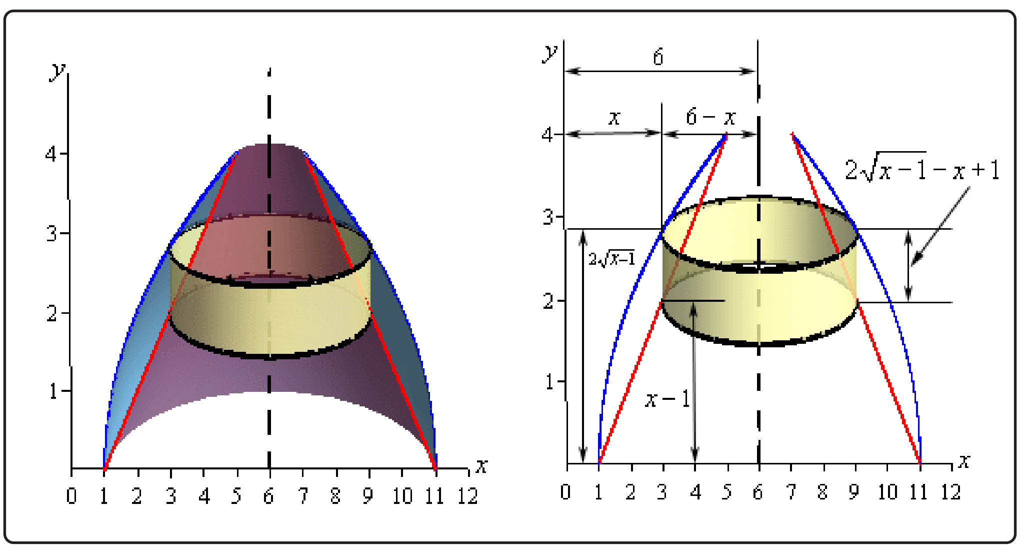
Determine the volume of the solid obtained by rotating the region bounded by $y = 2\sqrt{x-1}$ and $y = x - 1$ about the line $x = 6$.

Solution

Here's a graph of the bounded region and solid.



Here are our sketches of a typical cylinder. Again, the sketch on the left is here to provide some context for the sketch on the right.



Okay, there is a lot going on in the sketch to the left. First notice that the radius is not just an x or y as it was in the previous two cases. In this case x is the distance from the *y*-axis to the edge of the cylinder and we need the distance from the axis of rotation to the edge of the cylinder. That means that the radius of this cylinder is $6 - x$.

Secondly, the height of the cylinder is the difference of the two functions in this case.

The cross-sectional area is then,

$$\begin{aligned} A(x) &= 2\pi\,(\text{radius})\,(\text{height}) \\ &= 2\pi\,(6-x)\left(2\sqrt{x-1} - x + 1\right) \\ &= 2\pi\left(x^2 - 7x + 6 + 12\sqrt{x-1} - 2x\sqrt{x-1}\right) \end{aligned}$$

Now the first cylinder will cut into the solid at $x = 1$ and the final cylinder will cut into the solid at $x = 5$ so there are our limits.

Here is the volume.

$$\begin{aligned} V &= \int_a^b A(x)\,dx \\ &= 2\pi\int_1^5 x^2 - 7x + 6 + 12\sqrt{x-1} - 2x\sqrt{x-1}\,dx \\ &= 2\pi\left(\frac{1}{3}x^3 - \frac{7}{2}x^2 + 6x + 8(x-1)^{\frac{3}{2}} - \frac{4}{3}(x-1)^{\frac{3}{2}} - \frac{4}{5}(x-1)^{\frac{5}{2}}\right)\Bigg|_1^5 \\ &= 2\pi\left(\frac{136}{15}\right) \\ &= \frac{272\pi}{15} \end{aligned}$$

The integration of the last term is a little tricky so let's do that here. It will use the substitution,

$$u = x - 1 \qquad du = dx \qquad x = u + 1$$

$$\begin{aligned} \int 2x\sqrt{x-1}\,dx &= 2\int (u+1)\,u^{\frac{1}{2}}\,du \\ &= 2\int u^{\frac{3}{2}} + u^{\frac{1}{2}}\,du \\ &= 2\left(\frac{2}{5}u^{\frac{5}{2}} + \frac{2}{3}u^{\frac{3}{2}}\right) + c \\ &= \frac{4}{5}(x-1)^{\frac{5}{2}} + \frac{4}{3}(x-1)^{\frac{3}{2}} + c \end{aligned}$$

We saw one of these kinds of substitutions back in the substitution section.

Example 4

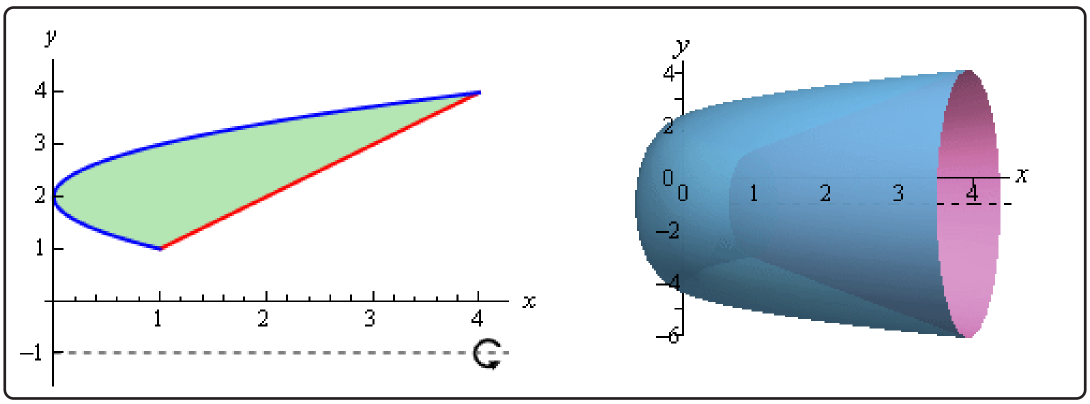
Determine the volume of the solid obtained by rotating the region bounded by $x = (y-2)^2$ and $y = x$ about the line $y = -1$.

Solution

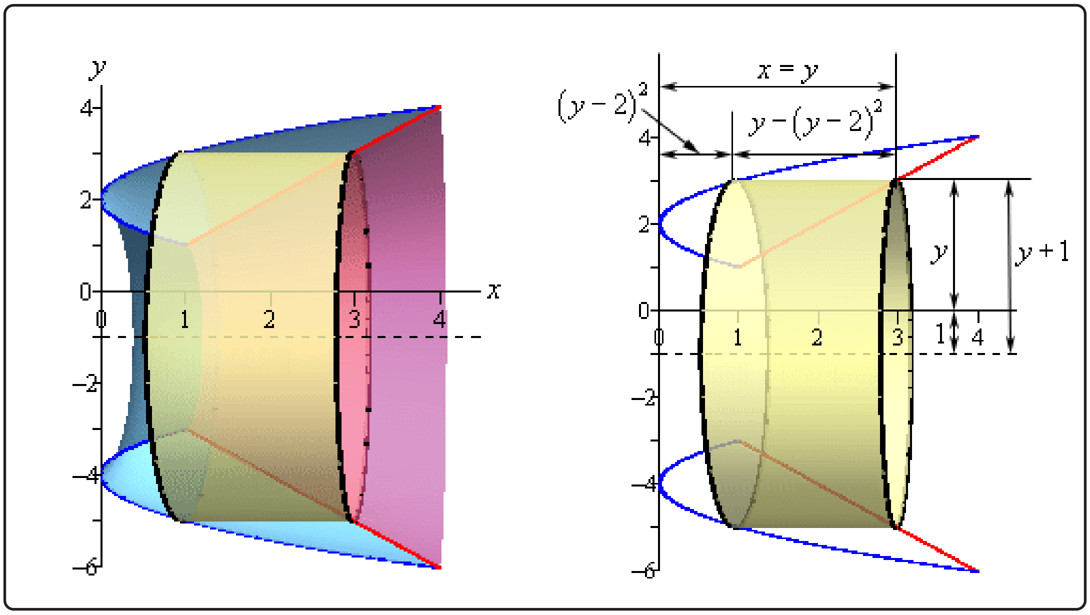
We should first get the intersection points there.

$$\begin{aligned} y &= (y-2)^2 \\ y &= y^2 - 4y + 4 \\ 0 &= y^2 - 5y + 4 \\ 0 &= (y-4)(y-1) \end{aligned}$$

So, the two curves will intersect at $y = 1$ and $y = 4$. Here is a sketch of the bounded region and the solid.



Here are our sketches of a typical cylinder. The sketch on the left is here to provide some context for the sketch on the right.



Here's the cross-sectional area for this cylinder.

$$\begin{aligned} A(y) &= 2\pi \,(\text{radius})\,(\text{width}) \\ &= 2\pi\,(y+1)\left(y-(y-2)^2\right) \\ &= 2\pi\left(-y^3+4y^2+y-4\right) \end{aligned}$$

The first cylinder will cut into the solid at $y = 1$ and the final cylinder will cut in at $y = 4$. The volume is then,

$$\begin{aligned} V &= \int_c^d A(y)\,dy \\ &= 2\pi \int_1^4 -y^3+4y^2+y-4\,dy \\ &= 2\pi\left(-\frac{1}{4}y^4+\frac{4}{3}y^3+\frac{1}{2}y^2-4y\right)\Bigg|_1^4 \\ &= \frac{63\pi}{2} \end{aligned}$$

6.5 More Volume Problems

In this section we're going to take a look at some more volume problems. However, the problems we'll be looking at here will not be solids of revolution as we looked at in the previous two sections. There are many solids out there that cannot be generated as solids of revolution, or at least not easily and so we need to take a look at how to do some of these problems.

Now, having said that these will not be solids of revolutions they will still be worked in pretty much the same manner. For each solid we'll need to determine the cross-sectional area, either $A(x)$ or $A(y)$, and they use the formulas we used in the previous two sections,

Volume Formulas

$$V = \int_a^b A(x)\, dx \qquad\qquad V = \int_c^d A(y)\, dy$$

The "hard" part of these problems will be determining what the cross-sectional area for each solid is. Each problem will be different and so each cross-sectional area will be determined by different means.

Also, before we proceed with any examples we need to acknowledge that the integrals in this section might look a little tricky at first. There are going to be very few numbers in these problems. All of the examples in this section are going to be more general derivation of volume formulas for certain solids. As such we'll be working with things like circles of radius r and we'll not be giving a specific value of r and we'll have heights of h instead of specific heights, *etc*.

All the letters in the integrals are going to make the integrals look a little tricky, but all you have to remember is that the r's and the h's are just letters being used to represent a fixed quantity for the problem, *i.e.* it is a constant. So, when we integrate we only need to worry about the letter in the differential as that is the variable we're actually integrating with respect to. All other letters in the integral should be thought of as constants. If you have trouble doing that, just think about what you'd do if the r was a 2 or the h was a 3 for example.

Let's start with a simple example that we don't actually need to do an integral that will illustrate how these problems work in general and will get us used to seeing multiple letters in integrals.

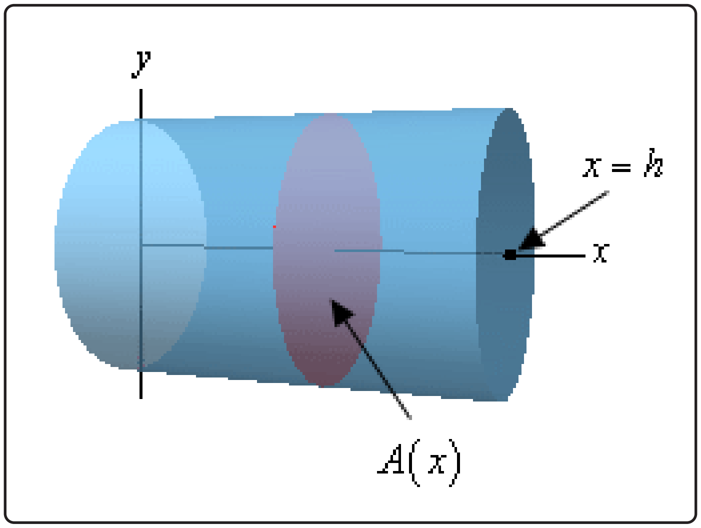
Example 1

Find the volume of a cylinder of radius r and height h.

Solution

Now, as we mentioned before starting this example we really don't need to use an integral to find this volume, but it is a good example to illustrate the method we'll need to use for these types of problems.

We'll start off with the sketch of the cylinder below.



We'll center the cylinder on the x-axis and the cylinder will start at $x = 0$ and end at $x = h$ as shown. Note that we're only choosing this particular set up to get an integral in terms of x and to make the limits nice to deal with. There are many other orientations that we could use.

What we need here is to get a formula for the cross-sectional area at any x. In this case the cross-sectional area is constant and will be a disk of radius r. Therefore, for any x we'll have the following cross-sectional area,

$$A(x) = \pi r^2$$

Next the limits for the integral will be $0 \le x \le h$ since that is the range of x in which the cylinder lives. Here is the integral for the volume,

$$V = \int_0^h \pi r^2 \, dx = \pi r^2 \int_0^h dx = \pi r^2 x \Big|_0^h = \pi r^2 h$$

So, we get the expected formula.

Also, recall we are using r to represent the radius of the cylinder. While r can clearly take different values it will never change once we start the problem. Cylinders do not change their radius in the middle of a problem and so as we move along the center of the cylinder (*i.e.* the x-axis) r is a fixed number and won't change. In other words, it is a constant that will not change as we change the x. Therefore, because we integrated with respect to x the r will be a constant as far as the integral is concerned. The r can then be pulled out of the integral as shown (although that's not required, we just did it to make the point). At this point we're just integrating dx and we know how to do that.

When we evaluate the integral remember that the limits are x values and so we plug into the x and NOT the r. Again, remember that r is just a letter that is being used to represent the radius of the cylinder and, once we start the integration, is assumed to be a fixed constant.

As noted before we started this example if you're having trouble with the r just think of what you'd do if there was a 2 there instead of an r. In this problem, because we're integrating with respect to x, both the 2 and the r will behave in the same manner. Note however that you should NEVER actually replace the r with a 2 as that WILL lead to a wrong answer. You should just think of what you would do IF the r was a 2.

So, to work these problems we'll first need to get a sketch of the solid with a set of x and y axes to help us see what's going on. At the very least we'll need the sketch to get the limits of the integral, but we will often need it to see just what the cross-sectional area is. Once we have the sketch we'll need to determine a formula for the cross-sectional area and then do the integral.

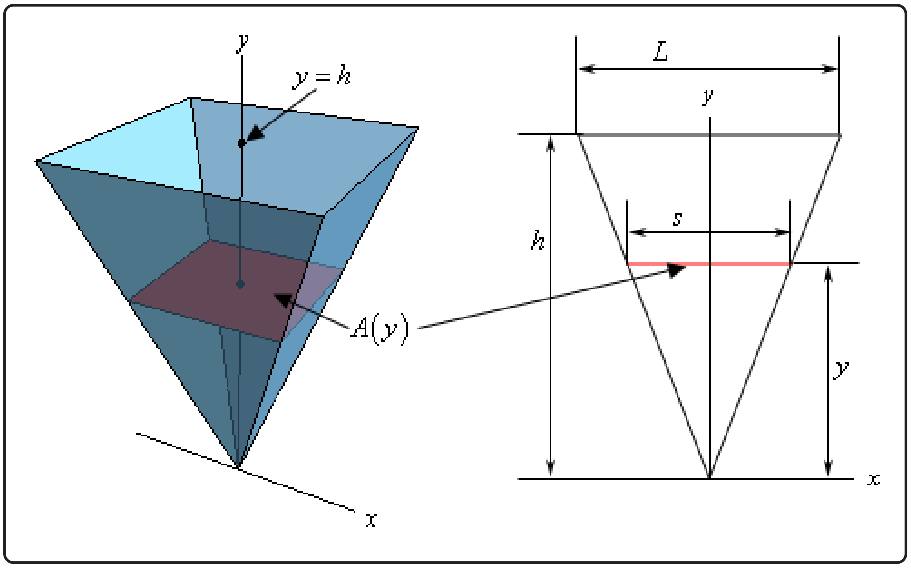
Let's work a couple of more complicated examples. In these examples the main issue is going to be determining what the cross-sectional areas are.

Example 2

Find the volume of a pyramid whose base is a square with sides of length L and whose height is h.

Solution

Let's start off with a sketch of the pyramid. In this case we'll center the pyramid on the y-axis and to make the equations easier we are going to position the point of the pyramid at the origin.



Now, as shown here the cross-sectional area will be a function of y and it will also be a square with sides of length s. The area of the square is easy, but we'll need to get the length of the side in terms of y. To determine this, consider the figure on the right above. If we look at the pyramid directly from the front we'll see that we have two similar triangles and we know that the ratio of any two sides of similar triangles must be equal. In other words, we know that,

$$\frac{s}{L} = \frac{y}{h} \qquad \Rightarrow \qquad s = \frac{y}{h}L = \frac{L}{h}y$$

So, the cross-sectional area is then,

$$A(y) = s^2 = \frac{L^2}{h^2}y^2$$

The limit for the integral will be $0 \le y \le h$ and the volume will be,

$$V = \int_0^h \frac{L^2}{h^2}y^2\,dy = \frac{L^2}{h^2}\int_0^h y^2\,dy = \frac{L^2}{h^2}\left(\frac{1}{3}y^3\right)\bigg|_0^h = \frac{1}{3}L^2 h$$

Again, do not get excited about the L and the h in the integral. Once we start the problem if we change y they will not change and so they are constants as far as the integral is concerned and can get pulled out of the integral. Also, remember that when we evaluate will only plug the limits into the variable we integrated with respect to, y in this case.

Before we proceed with some more complicated examples we should once again remind you to not get excited about the other letters in the integrals. If we're integrating with respect to x or y then all other letters in the formula that represent fixed quantities (*i.e.* radius, height, length of a side, *etc.*) are just constants and can be treated as such when doing the integral.

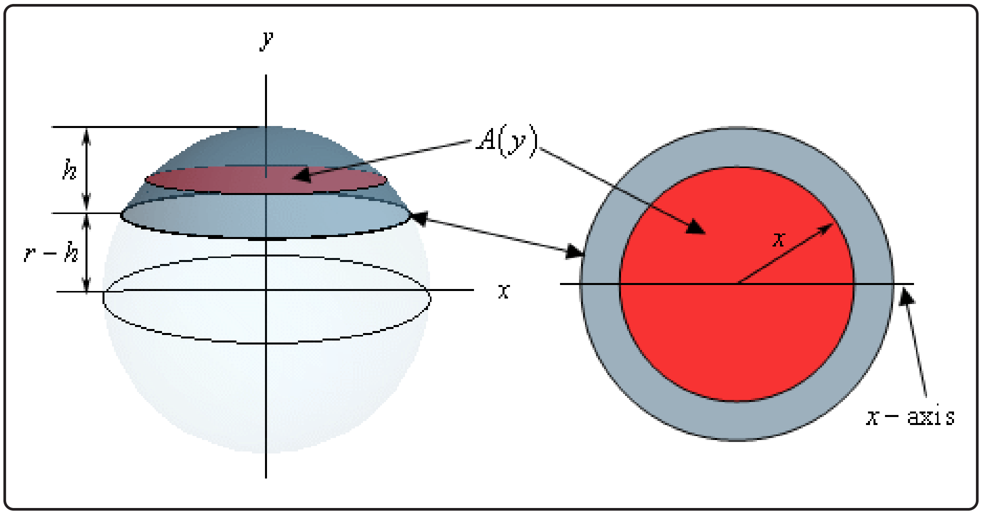
Now let's do some more examples.

Example 3

For a sphere of radius r find the volume of the cap of height h.

Solution

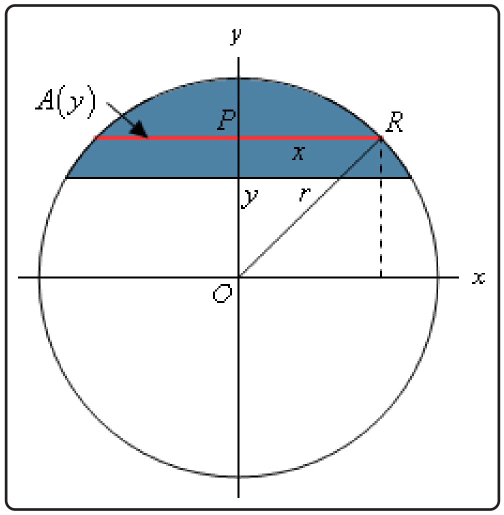
A sketch is probably best to illustrate what we're after here.



We are after the top portion of the sphere and the height of this is portion is h. In this case we'll use a cross-sectional area that starts at the bottom of the cap, which is at $y = r - h$, and moves up towards the top, which is at $y = r$. So, each cross-section will be a disk of radius x. It is a little easier to see that the radius will be x if we look at it from the top as shown in the sketch to the right above. The area of this disk is then,

$$\pi x^2$$

This is a problem however as we need the cross-sectional area in terms of y. So, what we really need to determine what x^2 will be for any given y at the cross-section. To get this let's look at the sphere from the front.



In particular look at the triangle POR. Because the point R lies on the sphere itself we can see that the length of the hypotenuse of this triangle (the line OR) is r, the radius of the sphere. The line PR has a length of x and the line OP has length y so by the Pythagorean Theorem we know,

$$x^2 + y^2 = r^2 \qquad \Rightarrow \qquad x^2 = r^2 - y^2$$

So, we now know what x^2 will be for any given y and so the cross-sectional area is,

$$A(y) = \pi\left(r^2 - y^2\right)$$

As we noted earlier the limits on y will be $r - h \le y \le r$ and so the volume is,

$$\begin{aligned} V &= \int_{r-h}^{r} \pi\left(r^2 - y^2\right)\,dy \\ &= \pi\left(r^2 y - \frac{1}{3}y^3\right)\Big|_{r-h}^{r} \\ &= \pi\left(h^2 r - \frac{1}{3}h^3\right) = \pi h^2\left(r - \frac{1}{3}h\right) \end{aligned}$$

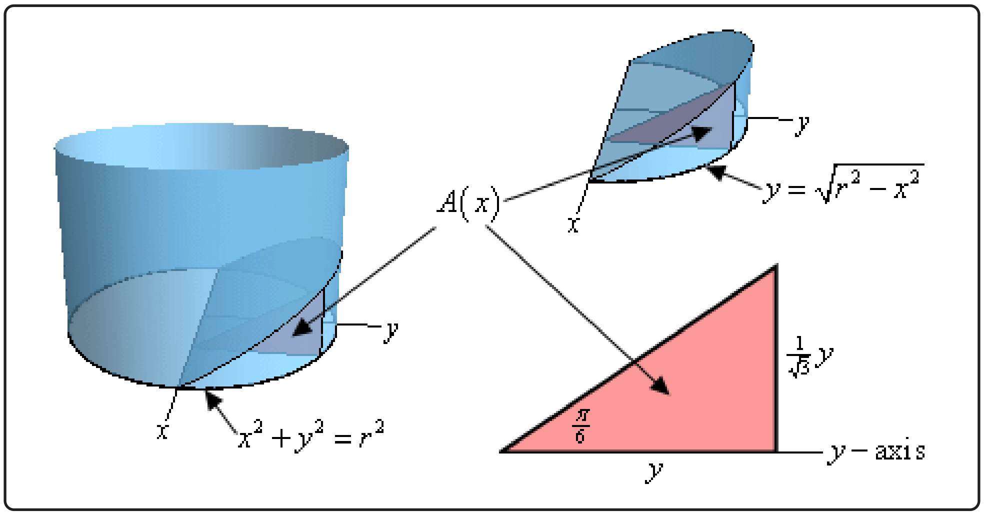
In the previous example we again saw an r in the integral. However, unlike the previous two examples it was not multiplied times the x or the y and so could not be pulled out of the integral. In this case it was like we were integrating $4 - y^2$ and we know how to integrate that. So, in this case we need to treat the r^2 like the 4 and so when we integrate that we'll pick up a y.

Example 4

Find the volume of a wedge cut out of a cylinder of radius r if the angle between the top and bottom of the wedge is $\frac{\pi}{6}$.

Solution

We should really start off with a sketch of just what we're looking for here.



On the left we see how the wedge is being cut out of the cylinder. The base of the cylinder is the circle give by $x^2 + y^2 = r^2$ and the angle between this circle and the top of the wedge is $\frac{\pi}{6}$. The sketch in the upper right position is the actual wedge itself. Given the orientation of the axes here we get the portion of the circle with positive y and so we can write the equation of the circle as $y = \sqrt{r^2 - x^2}$ since we only need the positive y values. Note as well that this is the reason for the way we oriented the axes here.

We get positive y's and we can write the equation of the circle as a function only of x's.

Now, as we can see in the two sketches of the wedge the cross-sectional area will be a right triangle and the area will be a function of x as we move from the back of the cylinder, at $x = -r$, to the front of the cylinder, at $x = r$.

The right angle of the triangle will be on the circle itself while the point on the x-axis will have an interior angle of $\frac{\pi}{6}$. The base of the triangle will have a length of y and using a little right triangle trig we see that the height of the rectangle is,

$$\text{height} = y \tan\left(\frac{\pi}{6}\right) = \frac{1}{\sqrt{3}}y$$

So, we now know the base and height of our triangle, in terms of y, and we have an equation for y in terms of x and so we can see that the area of the triangle, *i.e.* the cross-sectional area is,

$$A(x) = \frac{1}{2}(y)\left(\frac{1}{\sqrt{3}}y\right) = \frac{1}{2}\sqrt{r^2 - x^2}\left(\frac{1}{\sqrt{3}}\sqrt{r^2 - x^2}\right) = \frac{1}{2\sqrt{3}}\left(r^2 - x^2\right)$$

The limits on x are $-r \le x \le r$ and so the volume is then,

$$V = \int_{-r}^{r} \frac{1}{2\sqrt{3}}\left(r^2 - x^2\right) dx = \frac{1}{2\sqrt{3}}\left(r^2 x - \frac{1}{3}x^3\right)\Bigg|_{-r}^{r} = \frac{2r^3}{3\sqrt{3}}$$

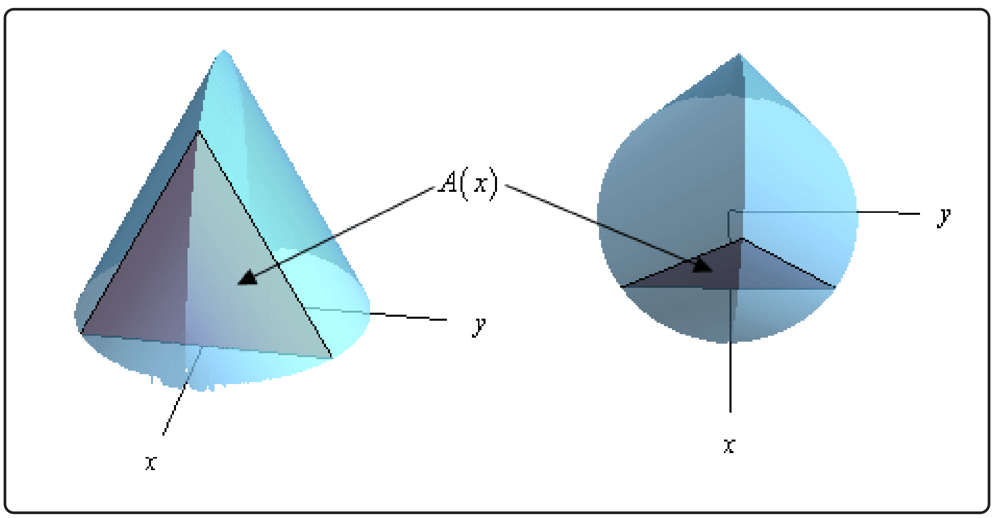
The next example is very similar to the previous one except it can be a little difficult to visualize the solid itself.

Example 5

Find the volume of the solid whose base is a disk of radius r and whose cross-sections are equilateral triangles.

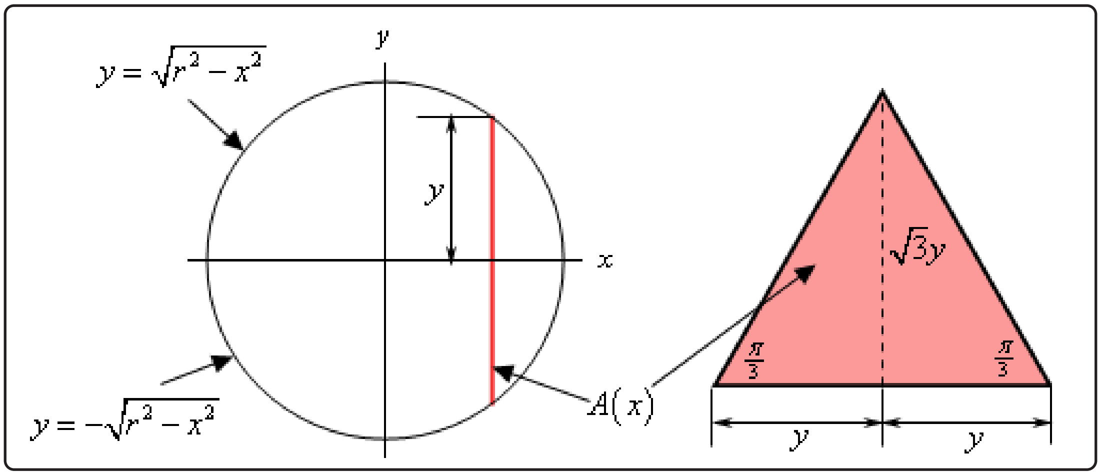
Solution

Let's start off with a couple of sketches of this solid. The sketch on the left is from the "front" of the solid and the sketch on the right is more from the top of the solid.



The base of this solid is the disk of radius r and we move from the back of the disk at $x = -r$ to the front of the disk at $x = r$ we form equilateral triangles to form the solid. A sample equilateral triangle, which is also the cross-sectional area, is shown above to hopefully make it a little clearer how the solid is formed.

Now, let's get a formula for the cross-sectional area. Let's start with the two sketches below.



In the left hand sketch we are looking at the solid from directly above and notice that we "reoriented" the sketch a little to put the x and y-axis in the "normal" orientation. The solid vertical line in this sketch is the cross-sectional area. From this we can see that the cross-section occurs at a given x and the top half will have a length of y where the value of y will be the y-coordinate of the point on the circle and so is,

$$y = \sqrt{r^2 - x^2}$$

Also, because the cross-section is an equilateral triangle that is centered on the x-axis the bottom half will also have a length of y. Thus, the base of the cross-section must have a length of $2y$.

The sketch to the right is of one of the cross-sections. As noted above the base of the triangle has a length of $2y$. Also note that because it is an equilateral triangle the angles are all $\frac{\pi}{3}$. If we divide the cross-section in two (as shown with the dashed line) we now have two right triangles and using right triangle trig we can see that the length of the dashed line is,

$$\text{dashed line} = y \tan\left(\frac{\pi}{3}\right) = \sqrt{3}\, y$$

Therefore, the height of the cross-section is $\sqrt{3}\, y$. Because the cross-section is a triangle we know that that it's area must then be,

$$A(x) = \frac{1}{2}(2y)\left(\sqrt{3}\, y\right) = \frac{1}{2}\left(2\sqrt{r^2 - x^2}\right)\left(\sqrt{3}\sqrt{r^2 - x^2}\right) = \sqrt{3}\left(r^2 - x^2\right)$$

Note that we used the cross-sectional area in terms of x because each of the cross-sections is perpendicular to the x-axis and this pretty much forces us to integrate with respect to x.

The volume of the solid is then,

$$V(x) = \int_{-r}^{r} \sqrt{3}\left(r^2 - x^2\right)\,dx = \sqrt{3}\left(r^2 x - \frac{1}{3}x^3\right)\Big|_{-r}^{r} = \frac{4}{\sqrt{3}}r^3$$

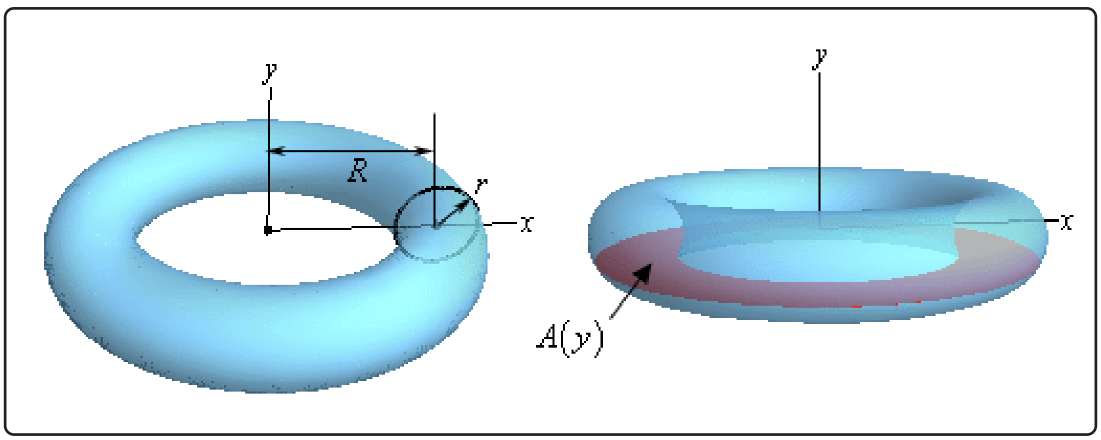
The final example we're going to work here is a little tricky both in seeing how to set it up and in doing the integral.

Example 6

Find the volume of a torus with radii r and R.

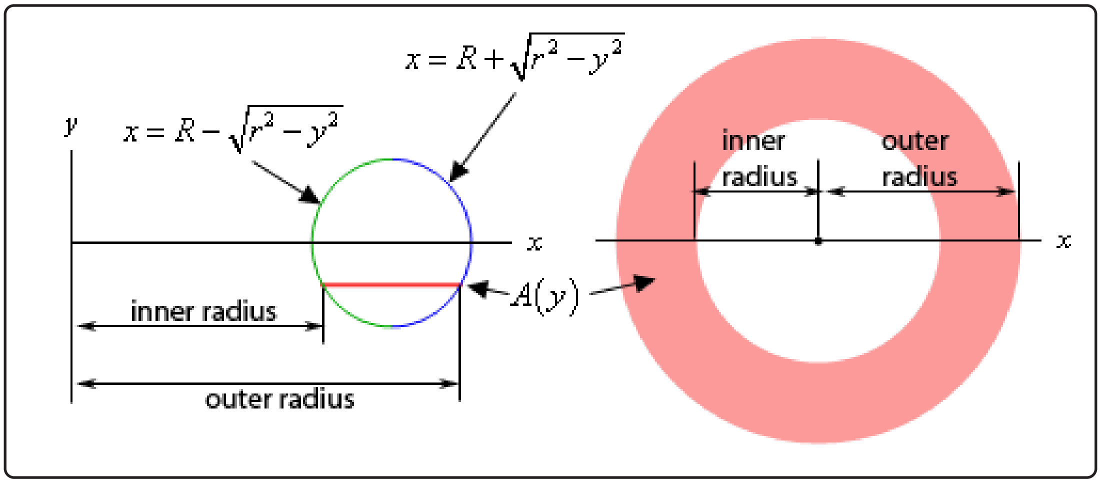
Solution

First, just what is a torus? A torus is a donut shaped solid that is generated by rotating the circle of radius r and centered at $(R, 0)$ about the y-axis. This is shown in the sketch to the left below.



One of the trickiest parts of this problem is seeing what the cross-sectional area needs to be. There is an obvious one. Most people would probably think of using the circle of radius r that we're rotating about the y-axis as the cross-section. It is definitely one of the more obvious choices, however setting up an integral using this is not so easy.

So, what we'll do is use a cross-section as shown in the sketch to the right above. If we cut the torus perpendicular to the y-axis we'll get a cross-section of a ring and finding the area of that shouldn't be too bad. To do that let's take a look at the two sketches below.



The sketch to the left is a sketch of the full cross-section. The sketch to the right is more important however. This is a sketch of the circle that we are rotating about the y-axis. Included is a line representing where the cross-sectional area would be in the torus.

Notice that the inner radius will always be the left portion of the circle and the outer radius will always be the right portion of the circle. Now, we know that the equation of this is,

$$(x - R)^2 + y^2 = r^2$$

and so if we solve for x we can get the equations for the left and right sides as shown above in the sketch. This however means that we also now have equations for the inner and outer radii.

inner radius : $x = R - \sqrt{r^2 - y^2}$ outer radius : $x = R + \sqrt{r^2 - y^2}$

The cross-sectional area is then,

$$\begin{aligned} A(y) &= \pi(\text{outer radius})^2 - \pi(\text{inner radius})^2 \\ &= \pi\left[\left(R + \sqrt{r^2 - y^2}\right)^2 - \left(R - \sqrt{r^2 - y^2}\right)^2\right] \\ &= \pi\left[R^2 + 2R\sqrt{r^2 - y^2} + r^2 - y^2 - \left(R^2 - 2R\sqrt{r^2 - y^2} + r^2 - y^2\right)\right] \\ &= 4\pi R\sqrt{r^2 - y^2} \end{aligned}$$

Next, the lowest cross-section will occur at $y = -r$ and the highest cross-section will occur at $y = r$ and so the limits for the integral will be $-r \le y \le r$.

The integral giving the volume is then,

$$V = \int_{-r}^{r} 4\pi R\sqrt{r^2 - y^2}\,dy = 2\int_0^r 4\pi R\sqrt{r^2 - y^2}\,dy = 8\pi R\int_0^r \sqrt{r^2 - y^2}\,dy$$

Note that we used the fact that because the integrand is an even function and we're integrating over $[-r, r]$ we could change the lower limit to zero and double the value of the integral. We saw this fact back in the Computing Definite Integrals section.

We've now reached the second really tricky part of this example. With the knowledge that we've currently got at this point this integral is not possible to do. It requires something called a trig substitution and that is a topic for Calculus II. Luckily enough for us, and this is the tricky part, in this case we can actually determine the integral's value using what we know about integrals.

Just for a second let's think about a different problem. Let's suppose we wanted to use an integral to determine the area under the portion of the circle of radius r and centered at the origin that is in the first quadrant. There are a couple of ways to do this, but to match what we're doing here let's do the following.

We know that the equation of the circle is $x^2 + y^2 = r^2$ and if we solve for x the equation of the circle in the first (and fourth for that matter) quadrant is,

$$x = \sqrt{r^2 - y^2}$$

If we want an integral for the area in the first quadrant we can think of this area as the region between the curve $x = \sqrt{r^2 - y^2}$ and the y-axis for $0 \le y \le r$ and this is,

$$A = \int_0^r \sqrt{r^2 - y^2}\, dy$$

In other words, this integral represents one quarter of the area of a circle of radius r and from basic geometric formulas we now know that this integral must have the value,

$$A = \int_0^r \sqrt{r^2 - y^2}\, dy = \frac{1}{4}\pi r^2$$

So, putting all this together the volume of the torus is then,

$$V = 8R\pi \int_0^r \sqrt{r^2 - y^2}\, dy = 8\pi R \left(\frac{1}{4}\pi r^2\right) = 2R\pi^2 r^2$$

6.6 Work

This is the final application of integral that we'll be looking at in this course. In this section we will be looking at the amount of work that is done by a force in moving an object.

In a first course in Physics you typically look at the work that a constant force, F, does when moving an object over a distance of d. In these cases the work is,

$$W = Fd$$

However, most forces are not constant and will depend upon where exactly the force is acting. So, let's suppose that the force at any x is given by $F(x)$. Then the work done by the force in moving an object from $x = a$ to $x = b$ is given by,

$$W = \int_a^b F(x)\, dx$$

To see a justification of this formula see the Proof of Various Integral Properties section of the Extras appendix.

Notice that if the force is constant we get the correct formula for a constant force.

$$\begin{aligned} W &= \int_a^b F\, dx \\ &= Fx\Big|_a^b \\ &= F(b-a) \end{aligned}$$

where $b - a$ is simply the distance moved, or d.

So, let's take a look at a couple of examples of non-constant forces.

Example 1

A spring has a natural length of 20 cm. A 40 N force is required to stretch (and hold the spring) to a length of 30 cm. How much work is done in stretching the spring from 35 cm to 38 cm?

Solution

This example will require Hooke's Law to determine the force. Hooke's Law tells us that the force required to stretch a spring a distance of x meters from its natural length is,

$$F(x) = k\,x$$

where $k > 0$ is called the spring constant. It is important to remember that the x in this formula is the distance the spring is stretched from its natural length and not the actual length of the spring.

So, the first thing that we need to do is determine the spring constant for this spring. We can do that using the initial information. A force of 40 N is required to stretch the spring

$$30 \text{ cm} - 20 \text{ cm} = 10 \text{ cm} = 0.1 \text{ m}$$

from its natural length. Using Hooke's Law we have,

$$40 = 0.10k \qquad \Rightarrow \qquad k = 400$$

So, according to Hooke's Law the force required to hold this spring x meters from its natural length is,

$$F(x) = 400x$$

We want to know the work required to stretch the spring from 35cm to 38cm. First, we need to convert these into distances from the natural length in meters. Doing that gives us x's of 0.15m and 0.18m.

The work is then,

$$\begin{aligned} W &= \int_{0.15}^{0.18} 400x \, dx \\ &= 200x^2 \Big|_{0.15}^{0.18} \\ &= 1.98 \text{ J} \end{aligned}$$

Example 2

We have a cable that weighs 2 lbs/ft attached to a bucket filled with coal that weighs 800 lbs. The bucket is initially at the bottom of a 500 ft mine shaft. Answer each of the following about this.

(a) Determine the amount of work required to lift the bucket to the midpoint of the shaft.

(b) Determine the amount of work required to lift the bucket from the midpoint of the shaft to the top of the shaft.

(c) Determine the amount of work required to lift the bucket all the way up the shaft.

Solution

Before answering either part we first need to determine the force. In this case the force will be the weight of the bucket and cable at any point in the shaft.

To determine a formula for this we will first need to set a convention for x. For this problem we will set x to be the amount of cable that has been pulled up. So at the bottom of the shaft $x = 0$, at the midpoint of the shaft $x = 250$ and at the top of the shaft $x = 500$. Also, at any point in the shaft there is $500 - x$ feet of cable still in the shaft.

The force then for any x is then nothing more than the weight of the cable and bucket at that point. This is,

$$\begin{aligned} F(x) &= \text{weight of cable} + \text{ weight of bucket/coal} \\ &= 2(500 - x) + 800 \\ &= 1800 - 2x \end{aligned}$$

We can now answer the questions.

(a) Determine the amount of work required to lift the bucket to the midpoint of the shaft.

In this case we want to know the work required to move the cable and bucket/coal from $x = 0$ to $x = 250$. The work required is,

$$\begin{aligned} W &= \int_0^{250} F(x)\,dx \\ &= \int_0^{250} 1800 - 2x\,dx \\ &= \left(1800x - x^2\right)\Big|_0^{250} \\ &= 387500 \text{ ft - lb} \end{aligned}$$

(b) Determine the amount of work required to lift the bucket from the midpoint of the shaft to the top of the shaft.

In this case we want to move the cable and bucket/coal from $x = 250$ to $x = 500$. The work required is,

$$\begin{aligned} W &= \int_{250}^{500} F(x)\,dx \\ &= \int_{250}^{500} 1800 - 2x\,dx \\ &= \left(1800x - x^2\right)\Big|_{250}^{500} \\ &= 262500 \text{ ft - lb} \end{aligned}$$

(c) Determine the amount of work required to lift the bucket all the way up the shaft.

In this case the work is,

$$\begin{aligned} W &= \int_0^{500} F\left(x\right)\,dx \\ &= \int_0^{500} 1800 - 2x\,dx \\ &= \left(1800x - x^2\right)\Big|_0^{500} \\ &= 650000 \text{ ft - lb} \end{aligned}$$

Note that we could have instead just added the results from the first two parts and we would have gotten the same answer to the third part.

Example 3

A 20 ft cable weighs 80 lbs and hangs from the ceiling of a building without touching the floor. Determine the work that must be done to lift the bottom end of the chain all the way up until it touches the ceiling.

Solution

First, we need to determine the weight per foot of the cable. This is easy enough to get,

$$\frac{80 \text{ lbs}}{20 \text{ ft}} = 4 \text{ lb/ft}$$

Next, let x be the distance from the ceiling to any point on the cable. Using this convention we can see that the portion of the cable in the range $10 < x \le 20$ will actually be lifted. The portion of the cable in the range $0 \le x \le 10$ will not be lifted at all since once the bottom of the cable has been lifted up to the ceiling the cable will be doubled up and each portion will have a length of 10 ft. So, the upper 10 foot portion of the cable will never be lifted while the lower 10 ft portion will be lifted.

Now, the very bottom of the cable, $x = 20$, will be lifted 10 feet to get to the midpoint and then a further 10 feet to get to the ceiling. A point 2 feet from the bottom of the cable, $x = 18$ will lift 8 feet to get to the midpoint and then a further 8 feet until it reaches its final position (if it is 2 feet from the bottom then its final position will be 2 feet from the ceiling). Continuing on in this fashion we can see that for any point on the lower half of the cable, *i.e.* $10 \le x \le 20$ it will be lifted a total of $2\left(x - 10\right)$.

As with the previous example the force required to lift any point of the cable in this range is simply the distance that point will be lifted times the weight/foot of the cable. So, the force is then,

$$\begin{aligned} F(x) &= (\text{distance lifted})\,(\text{weight per foot of cable}) \\ &= 2(x-10)(4) \\ &= 8(x-10) \end{aligned}$$

The work required is now,

$$\begin{aligned} W &= \int_{10}^{20} 8(x-10)\,dx \\ &= \left(4x^2 - 80x\right)\Big|_{10}^{20} \\ &= 400 \text{ ft - lb} \end{aligned}$$

Provided we can find the force, $F(x)$, for a given problem then using the above method for determining the work is (generally) pretty simple. However, there are some problems where this approach won't easily work. Let's take a look at one of those kinds of problems.

Example 4

A tank in the shape of an inverted cone has a height of 15 meters and a base radius of 4 meters and is filled with water to a depth of 12 meters. Determine the amount of work needed to pump all of the water to the top of the tank. Assume that the density of the water is 1000 kg/m^3.

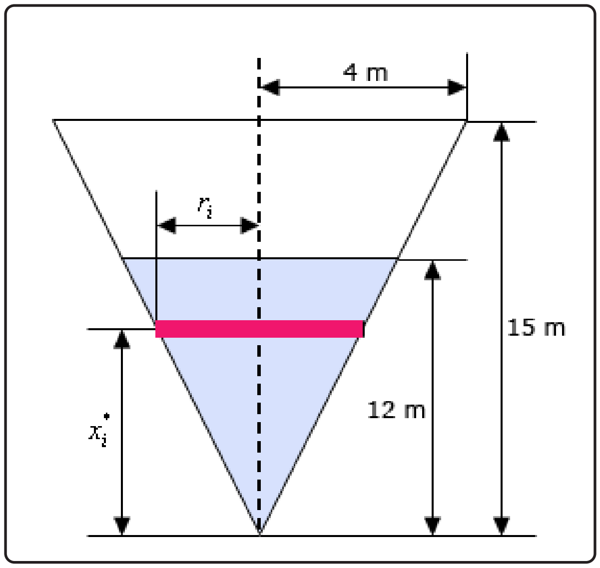
Solution

Okay, in this case we cannot just determine a force function, $F(x)$ that will work for us. So, we are going to need to approach this from a different standpoint.

Let's first set $x = 0$ to be the lower end of the tank/cone and $x = 15$ to be the top of the tank/cone. With this *definition* of our x's we can now see that the water in the tank will correspond to the interval $[0, 12]$.

So, let's start off by dividing $[0, 12]$ into n subintervals each of width Δx and let's also let x_i^* be any point from the i^{th} subinterval where $i = 1, 2, \ldots n$. Now, for each subinterval we will approximate the water in the tank corresponding to that interval as a cylinder of radius r_i and height Δx.

Here is a quick sketch of the tank. Note that the sketch really isn't to scale and we are looking at the tank from directly in front so we can see all the various quantities that we need to work with.



The red strip in the sketch represents the “cylinder” of water in the i^{th} subinterval. A quick application of similar triangles will allow us to relate r_i to x_i^* (which we'll need in a bit) as follows.

$$\frac{r_i}{x_i^*} = \frac{4}{15} \qquad \Rightarrow \qquad r_i = \frac{4}{15} x_i^*$$

Okay, the mass, m_i, of the volume of water, V_i, for the i^{th} subinterval is simply,

$$m_i = \text{density} \times V_i$$

We know the density of the water (it was given in the problem statement) and because we are approximating the water in the i^{th} subinterval as a cylinder we can easily approximate the volume as,

$$V_i \approx \pi(\text{radius})^2 (\text{height})$$

We can now approximate the mass of water in the i^{th} subinterval,

$$m_i \approx (1000) \left[\pi r_i^2 \Delta x\right] = 1000\pi \left(\frac{4}{15} x_i^*\right)^2 \Delta x = \frac{640}{9} \pi (x_i^*)^2 \Delta x$$

To raise this volume of water we need to overcome the force of gravity that is acting on the volume and that is, $F = m_i g$, where $g = 9.8\, m/s^2$ is the gravitational acceleration. The force to raise the volume of water in the i^{th} subinterval is then approximately,

$$F_i = m_i g \approx (9.8)\, \frac{640}{9} \pi (x_i^*)^2 \Delta x$$

Next, in order to reach to the top of the tank the water in the i^{th} subinterval will need to travel approximately $15 - x_i^*$ to reach the top of the tank. Because the volume of the water in the

i^{th} subinterval is constant the force needed to raise the water through any distance is also a constant force.

Therefore, the work to move the volume of water in the i^{th} subinterval to the top of the tank, *i.e.* raise it a distance of $15 - x_i^*$, is then approximately,

$$W_i \approx F_i \left(15 - x_i^*\right) = (9.8)\,\frac{640}{9}\pi (x_i^*)^2 \left(15 - x_i^*\right) \Delta x$$

The total amount of work required to raise all the water to the top of the tank is then approximately the sum of each of the W_i for $i = 1, 2, \ldots n$. Or,

$$W \approx \sum_{i=1}^{n} (9.8)\,\frac{640}{9}\pi (x_i^*)^2 \left(15 - x_i^*\right) \Delta x$$

To get the actual amount of work we simply need to take $n \to \infty$. *I.e.* compute the following limit,

$$W = \lim_{n \to \infty} \sum_{i=1}^{n} (9.8)\,\frac{640}{9}\pi (x_i^*)^2 \left(15 - x_i^*\right) \Delta x$$

This limit of a summation should look somewhat familiar to you. It's probably been some time, but recalling the definition of the definite integral we can see that this is nothing more than the following definite integral,

$$\begin{aligned} W &= \int_0^{12} (9.8)\,\frac{640}{9}\pi x^2 \left(15 - x\right)\,dx = (9.8)\,\frac{640}{9}\pi \int_0^{12} 15x^2 - x^3\,dx \\ &= (9.8)\,\frac{640}{9}\pi \left(5x^3 - \frac{1}{4}x^4\right)\Bigg|_0^{12} = 7,566,362.543\,\text{J} \end{aligned}$$

As we've seen in the previous example we sometimes need to compute "incremental" work and then use that to determine the actual integral we need to compute. This idea does arise on occasion and we shouldn't forget it!

7 Integration Techniques

By this point we've now looked at basic integration techniques. We've seen how to integrate most of the "basic" functions we're liable to run into : polynomials, roots, trig, exponential, logarithm and inverse trig functions to name a few. In addition, we've seen how to do basic u-substitutions allowing us to integrate some more complicated functions.

We've also taken a look at some basic applications of (definite) integrals. However, as was noted at the time, there are applications of (definite) integrals that will, on occasion, have integrals that need more than just a basic u-substitution. So, before we can take a look at those applications we'll need to first talk about some more involved integration techniques.

Before getting into the new techniques we first need to make it clear that in this chapter it is assumed at you are comfortable with basic integration, including u-substitutions. Many of the problems in this chapter will not have a lot, if any, discussion of the basic integration work under the assumption that you are comfortable enough with the basic work that discussion in simply not needed. In addition, we will usually, although not always, give the substitution that we're using for the u-substitution but we will generally not show the actual substitution work. Again, this is under the assumption that you are comfortable enough with basic u-substitutions that you can fill in the details if you need to.

The reason for skipping the discussion of the basic integration work and/or not showing the full substitution work is so we can concentrate our discussion on the particular method that we are covering in that particular section. This is not to "punish" you but simply to acknowledge that we only have so much time in which to discuss the material and just can't afford to spend a lot of time basically re-lecturing basic integration material. We realize that, for many of you, this is the start of your Calculus II course and so you may have had some time off and may well have some "rust" on your basic integration skills. This is a warning to start scraping that rust off. If you need do scrape some rust off you can check out the practice problems for some practice problems covering basic integration to refresh your memory on how basic integration works.

It is also very important for you to understand that most of the problems we'll be looking at in this chapter will involve u-substitutions in one way or another. In fact, many of the techniques in this chapter are really just substitutions. The only difference is that either they need a fair amount of work to get to the point where the substitutions can be used or they will involve substitutions used in ways that we've not seen to this point. So, again, if you have some rust on your u-substitution skills you'll need to get it scraped off so you can do the work in this chapter.

In addition, we will be doing indefinite integrals almost exclusively in most of the sections in this chapter. There are a few sections were we'll be doing some definite integrals but for the most part we'll keep the problems in most of the sections shorter by just doing indefinite integrals. It is assumed that if you were given a definite integral you could do the extra evaluation steps needed to finish the definite integral. Having said that, there are a few sections were definite integrals are done either because there are some subtleties that need to be dealt with for definite integrals or because the topic at hand, the last few sections in particular, involve only definite integrals.

So, with all that out of the way, here is a quick rundown of the new integration techniques we'll take a look at in this section.

Probably the most important technique, in this sense that it will be the most commonly seen technique out of this class, is integration by parts. This is the one new technique in this chapter that is not just u-substitutions done in new ways. Integration by Parts will involve u-substitutions at various steps the process on occasion but it will not be just a new way of doing a u-substitution.

As noted a lot of the techniques in this chapter are really just u-substitutions except they will need some manipulation of the integrand prior to actually doing the substitution. The techniques using this idea will include integrating some, but not all, products and quotients of trig functions, some integrands involving roots or quadratics that can't be done without manipulation of the integrand or "different" u-substitutions that we are used to. We'll also see how to use partial fractions to write some integrands involving rational expressions into a form that we can actually do the integral.

We'll also take a look at something called trig substitutions. This is probably the one technique that is usually considered the most difficult, or at the least, the longest method. As we'll see a trig substitution is really a substitution but it is not a traditional u-substitution. However, having said that, if you understand how basic u-substitutions work it will help greatly when it comes to working with trig substitutions as the basic concepts are the same.

Next we'll be taking a look at a new kind of integral, Improper Integrals. This topic will address how to deal with definite integrals for which one or both of the limits of integration will be an infinity. In addition, we'll see how we can, on occasion, deal with discontinuities in the integrand (we'll focus on division by zero in the integrand).

We'll close out the chapter with a quick section on approximating the value of definite integrals.

We will leave this introduction with a warning. It is with this chapter that you will find that you can't just memorize your way through the class anymore. We will acknowledge that up to this point it is possible, for the most part, to just memorize your way through the class. You may not get the highest grades through just memorization as there are some topics that require a fair amount of understanding of the topic, but you can survive up to this point if your really good at memorization.

Integration by Parts is a really good example of this warning. While you will need to memorize/know the basic integration by parts formula simply memorizing that will not help you to actually use integration by parts on the problem. You will need to actually understand how integration by parts works and how to "assign" various portions of the integrand to the various portions of the integration parts formula.

Also while there are some basic formulas we can, and do on occasion, give for some of the methods there are also situations that just don't fit into those formulas and so again you'll really need to understand how to do those methods in order to work problems for which basic formulas just won't work. Or, again, you can't just memorize your way out of most the methods taught in this chapter. Memorization may allow you to get through the basic problems but will not help all that much with more complicated problems.

Finally, we also need to warn you about seeing "patterns" and just assuming that all the problems will fall into those patterns. Integration by Parts is, again, a good example of this. There are some "patterns" that seem to show up because a lot of the problems we do in that section do fall into the patterns. The problem is that there are also some problems for which the "patterns" simply don't work and yet they still require integration by parts. If you get so locked into "patterns" you'll find it all but impossible to do some problems because they simply don't fall into those patterns.

This is not to say that recognizing that patterns in always a bad thing. Patterns do, on occasion, show up and they can be useful to understand/know as a possible solution method. However, you also need to always remember that there are problems that just don't fit easily into the patterns. This is also a warning that will be valid in other chapters in a typical Calculus II course as well. Again, patterns aren't bad per se, you just need to be careful to not always assume that every problem will fall into the patterns.

7.1 Integration by Parts

Let's start off with this section with a couple of integrals that we should already be able to do to get us started. First let's take a look at the following.

$$\int \mathbf{e}^x \, dx = \mathbf{e}^x + c$$

So, that was simple enough. Now, let's take a look at,

$$\int x\mathbf{e}^{x^2} \, dx$$

To do this integral we'll use the following substitution.

$$u = x^2 \qquad du = 2x\,dx \qquad \Rightarrow \qquad x\,dx = \frac{1}{2}\,du$$

$$\int x\mathbf{e}^{x^2} \, dx = \frac{1}{2}\int \mathbf{e}^u \, du = \frac{1}{2}\mathbf{e}^u + c = \frac{1}{2}\mathbf{e}^{x^2} + c$$

Again, simple enough to do provided you remember how to do substitutions. By the way make sure that you can do these kinds of substitutions quickly and easily. From this point on we are going to be doing these kinds of substitutions in our head. If you have to stop and write these out with every problem you will find that it will take you significantly longer to do these problems.

Now, let's look at the integral that we really want to do.

$$\int x\mathbf{e}^{6x} \, dx$$

If we just had an x by itself or $\mathbf{e}^{6x}$ by itself we could do the integral easily enough. Likewise, if the integrand was $x\mathbf{e}^{6x^2}$ we could do the integral with a substitution. Unfortunately, however, neither of these are options. So, at this point we don't have the knowledge to do this integral.

To do this integral we will need to use integration by parts so let's derive the integration by parts formula. We'll start with the product rule.

$$\Big(f(x)\,g(x)\Big)' = f'(x)\,g(x) + f(x)\,g'(x)$$

Now, integrate both sides of this.

$$\int \Big(f(x)\,g(x)\Big)' \, dx = \int f'(x)\,g(x) + f(x)\,g'(x)\,dx$$

The left side is easy enough to integrate (we know that integrating a derivative just "undoes" the derivative) and we'll split up the right side of the integral.

$$f(x)g(x) = \int f'(x)\,g(x)\,dx + \int f(x)\,g'(x)\,dx$$

Note that technically we should have had a constant of integration show up on the left side after doing the integration. We can drop it at this point since other constants of integration will be showing up down the road and they would just end up absorbing this one.

Finally, rewrite the formula as follows and we arrive at the integration by parts formula.

Integration by Parts (formal)

$$\int f(x)\,g'(x)\,dx = f(x)g(x) - \int f'(x)\,g(x)\,dx$$

This is not the easiest formula to use however. So, let's do a couple of substitutions.

$$u = f(x) \qquad\qquad v = g(x)$$
$$du = f'(x)\,dx \qquad\qquad dv = g(x)\,dx$$

Both of these are just the standard Calculus I substitutions that hopefully you are used to by now. Don't get excited by the fact that we are using two substitutions here. They will work the same way.

Using these substitutions gives us the formula that most people think of as the integration by parts formula.

Integration By Parts (simplified)

$$\int u\,dv = uv - \int v\,du$$

To use this formula, we will need to identify u and dv, compute du and v and then use the formula. Note as well that computing v is very easy. All we need to do is integrate dv.

$$v = \int dv$$

One of the more complicated things about using this formula is you need to be able to correctly identify both the u and the dv. It won't always be clear what the correct choices are and we will, on occasion, make the wrong choice. This is not something to worry about. If we make the wrong choice, we can always go back and try a different set of choices.

This does lead to the obvious question of how do we know if we made the correct choice for u and dv? The answer is actually pretty simple. We made the correct choices for u and dv if, after using the integration by parts formula the new integral (the one on the right of the formula) is one we can actually integrate.

So, let's take a look at the integral above that we mentioned we wanted to do.

Example 1

Evaluate the following integral.

$$\int x\mathbf{e}^{6x}\,dx$$

Solution

So, on some level, the problem here is the x that is in front of the exponential. If that wasn't there we could do the integral. Notice as well that in doing integration by parts anything that we choose for u will be differentiated. So, it seems that choosing $u = x$ will be a good choice since upon differentiating the x will drop out.

Now that we've chosen u we know that dv will be everything else that remains. So, here are the choices for u and dv as well as du and v.

$$u = x \qquad\qquad dv = \mathbf{e}^{6x}\,dx$$
$$du = dx \qquad\qquad v = \int \mathbf{e}^{6x}\,dx = \frac{1}{6}\mathbf{e}^{6x}$$

The integral is then,

$$\begin{aligned}\int x\mathbf{e}^{6x}\,dx &= \frac{x}{6}\mathbf{e}^{6x} - \int \frac{1}{6}\mathbf{e}^{6x}\,dx \\ &= \frac{x}{6}\mathbf{e}^{6x} - \frac{1}{36}\mathbf{e}^{6x} + c\end{aligned}$$

Once we have done the last integral in the problem we will add in the constant of integration to get our final answer.

Note as well that, as noted above, we know we made made a correct choice for u and dv when we got a new integral that we actually evaluate after applying the integration by parts formula.

Next, let's take a look at integration by parts for definite integrals. The integration by parts formula for definite integrals is,

Integration By Parts, Definite Integrals

$$\int_a^b u\,dv = uv\Big|_a^b - \int_a^b v\,du$$

Note that the $uv|_a^b$ in the first term is just the standard integral evaluation notation that you should be familiar with at this point. All we do is evaluate the term, *uv* in this case, at b then subtract off the evaluation of the term at a.

At some level we don't really need a formula here because we know that when doing definite integrals all we need to do is evaluate the indefinite integral and then do the evaluation. In fact, this is probably going to be slightly easier as we don't need to track evaluating each term this way.

Let's take a quick look at a definite integral using integration by parts.

Example 2

Evaluate the following integral.

$$\int_{-1}^{2} x\mathbf{e}^{6x}\,dx$$

Solution

This is the same integral that we looked at in the first example so we'll use the same u and dv to get,

$$\begin{aligned}\int_{-1}^{2} x\mathbf{e}^{6x}\,dx &= \frac{x}{6}\mathbf{e}^{6x}\Big|_{-1}^{2} - \frac{1}{6}\int_{-1}^{2}\mathbf{e}^{6x}\,dx \\ &= \frac{x}{6}\mathbf{e}^{6x}\Big|_{-1}^{2} - \frac{1}{36}\mathbf{e}^{6x}\Big|_{-1}^{2} \\ &= \frac{11}{36}\mathbf{e}^{12} + \frac{7}{36}\mathbf{e}^{-6}\end{aligned}$$

As noted above we could just as easily used the result from the first example to do the evaluation. We know, from the first example that,

$$\int x\mathbf{e}^{6x}\,dx = \frac{x}{6}\mathbf{e}^{6x} - \frac{1}{36}\mathbf{e}^{6x} + c$$

Using this we can quickly proceed to the evaluate of the definite integral as follows,

$$\begin{aligned}\int_{-1}^{2} x\mathbf{e}^{6x}\,dx &= \left(\frac{x}{6}\mathbf{e}^{6x} - \frac{1}{36}\mathbf{e}^{6x}\right)\Bigg|_{-1}^{2} \\ &= \left(\frac{1}{3}\mathbf{e}^{12} - \frac{1}{36}\mathbf{e}^{12}\right) - \left(-\frac{1}{6}\mathbf{e}^{-6} - \frac{1}{36}\mathbf{e}^{-6}\right) \\ &= \frac{11}{36}\mathbf{e}^{12} + \frac{7}{36}\mathbf{e}^{-6}\end{aligned}$$

Either method of evaluating definite integrals with integration by part are pretty simple so which one you choose to use is pretty much up to you.

Since we need to be able to do the indefinite integral in order to do the definite integral and doing the definite integral amounts to nothing more than evaluating the indefinite integral at a couple of points we will concentrate on doing indefinite integrals in the rest of this section. In fact, throughout most of this chapter this will be the case. We will be doing far more indefinite integrals than definite integrals.

Let's take a look at some more examples.

Example 3

Evaluate the following integral.

$$\int (3t+5)\cos\left(\frac{t}{4}\right)\,dt$$

Solution

There are two ways to proceed with this example. For many, the first thing that they try is multiplying the cosine through the parenthesis, splitting up the integral and then doing integration by parts on the first integral.

While that is a perfectly acceptable way of doing the problem it's more work than we really need to do. Instead of splitting the integral up let's instead use the following choices for u and dv.

$$u = 3t+5 \qquad dv = \cos\left(\frac{t}{4}\right)\,dt$$

$$du = 3\,dt \qquad v = 4\sin\left(\frac{t}{4}\right)$$

The integral is then,

$$\begin{aligned}\int (3t+5)\cos\left(\frac{t}{4}\right)\,dt &= 4\,(3t+5)\sin\left(\frac{t}{4}\right) - 12\int \sin\left(\frac{t}{4}\right)\,dt \\ &= 4\,(3t+5)\sin\left(\frac{t}{4}\right) + 48\cos\left(\frac{t}{4}\right) + c\end{aligned}$$

Notice that we pulled any constants out of the integral when we used the integration by parts formula. We will usually do this in order to simplify the integral a little.

Example 4

Evaluate the following integral.

$$\int w^2 \sin(10w)\, dw$$

Solution

For this example, we'll use the following choices for u and dv.

$$u = w^2 \qquad dv = \sin(10w)\, dw$$
$$du = 2w\, dw \qquad v = -\frac{1}{10}\cos(10w)$$

The integral is then,

$$\int w^2 \sin(10w)\, dw = -\frac{w^2}{10}\cos(10w) + \frac{1}{5}\int w\cos(10w)\, dw$$

In this example, unlike the previous examples, the new integral will also require integration by parts. For this second integral we will use the following choices.

$$u = w \qquad dv = \cos(10w)\, dw$$
$$du = dw \qquad v = \frac{1}{10}\sin(10w)$$

So, the integral becomes,

$$\begin{aligned}\int w^2 \sin(10w)\, dw &= -\frac{w^2}{10}\cos(10w) + \frac{1}{5}\left(\frac{w}{10}\sin(10w) - \frac{1}{10}\int \sin(10w)\, dw\right) \\ &= -\frac{w^2}{10}\cos(10w) + \frac{1}{5}\left(\frac{w}{10}\sin(10w) + \frac{1}{100}\cos(10w)\right) + c \\ &= -\frac{w^2}{10}\cos(10w) + \frac{w}{50}\sin(10w) + \frac{1}{500}\cos(10w) + c\end{aligned}$$

Be careful with the coefficient on the integral for the second application of integration by parts. Since the integral is multiplied by $\frac{1}{5}$ we need to make sure that the results of actually doing the integral are also multiplied by $\frac{1}{5}$. Forgetting to do this is one of the more common mistakes with integration by parts problems.

As this last example has shown us, we will sometimes need more than one application of integration by parts to completely evaluate an integral. This is something that will happen so don't get excited about it when it does.

In this next example we need to acknowledge an important point about integration techniques.

Some integrals can be done in using several different techniques. That is the case with the integral in the next example.

Example 5

Evaluate the following integral

$$\int x\sqrt{x+1}\,dx$$

(a) Using Integration by Parts.

(b) Using a standard Calculus I substitution.

Solution

(a) Using Integration by Parts.

First notice that there are no trig functions or exponentials in this integral. While a good many integration by parts integrals will involve trig functions and/or exponentials not all of them will so don't get too locked into the idea of expecting them to show up.

In this case we'll use the following choices for u and dv.

$$u = x \qquad dv = \sqrt{x+1}\,dx$$
$$du = dx \qquad v = \frac{2}{3}(x+1)^{\frac{3}{2}}$$

The integral is then,

$$\begin{aligned}\int x\sqrt{x+1}\,dx &= \frac{2}{3}x(x+1)^{\frac{3}{2}} - \frac{2}{3}\int (x+1)^{\frac{3}{2}}\,dx \\ &= \frac{2}{3}x(x+1)^{\frac{3}{2}} - \frac{4}{15}(x+1)^{\frac{5}{2}} + c\end{aligned}$$

(b) Using a standard Calculus I substitution.

Now let's do the integral with a substitution. We can use the following substitution.

$$u = x+1 \qquad x = u-1 \qquad du = dx$$

Notice that we'll actually use the substitution twice, once for the quantity under the square root and once for the x in front of the square root. The integral is then,

$$\begin{aligned}\int x\sqrt{x+1}\,dx &= \int (u-1)\sqrt{u}\,du \\ &= \int u^{\frac{3}{2}} - u^{\frac{1}{2}}\,du \\ &= \frac{2}{5}u^{\frac{5}{2}} - \frac{2}{3}u^{\frac{3}{2}} + c \\ &= \frac{2}{5}(x+1)^{\frac{5}{2}} - \frac{2}{3}(x+1)^{\frac{3}{2}} + c\end{aligned}$$

So, we used two different integration techniques in this example and we got two different answers. The obvious question then should be : Did we do something wrong?

It turns out that, we didn't do anything wrong. We need to remember the following fact from Calculus I.

$$\text{If } f'(x) = g'(x) \quad \text{then} \quad f(x) = g(x) + c$$

In other words, if two functions have the same derivative then they will differ by no more than a constant. So, how does this apply to the above problem? First define the following,

$$f'(x) = g'(x) = x\sqrt{x+1}$$

Then we can compute $f(x)$ and $g(x)$ by integrating as follows,

$$f(x) = \int f'(x)\,dx \qquad\qquad g(x) = \int g'(x)\,dx$$

We'll use integration by parts for the first integral and the substitution for the second integral. Then according to the fact $f(x)$ and $g(x)$ should differ by no more than a constant. Let's verify this and see if this is the case. We can verify that they differ by no more than a constant if we take a look at the difference of the two and do a little algebraic manipulation and simplification.

$$\begin{aligned}&\left(\tfrac{2}{3}x(x+1)^{\frac{3}{2}} - \tfrac{4}{15}(x+1)^{\frac{5}{2}}\right) - \left(\tfrac{2}{5}(x+1)^{\frac{5}{2}} - \tfrac{2}{3}(x+1)^{\frac{3}{2}}\right) \\ &\qquad = (x+1)^{\frac{3}{2}}\left(\tfrac{2}{3}x - \tfrac{4}{15}(x+1) - \tfrac{2}{5}(x+1) + \tfrac{2}{3}\right) \\ &\qquad = (x+1)^{\frac{3}{2}}(0) \\ &\qquad = 0\end{aligned}$$

So, in this case it turns out the two functions are exactly the same function since the difference is zero. Note that this won't always happen. Sometimes the difference will yield a nonzero constant. For an example of this check out the Constant of Integration section in the Calculus I notes.

So just what have we learned? First, there will, on occasion, be more than one method for evaluating an integral. Secondly, we saw that different methods will often lead to different answers. Last, even though the answers are different it can be shown, sometimes with a lot of work, that they differ by no more than a constant.

When we are faced with an integral the first thing that we'll need to decide is if there is more than one way to do the integral. If there is more than one way we'll then need to determine which method we should use. The general rule of thumb that I use in my classes is that you should use the method that *you* find easiest. This may not be the method that others find easiest, but that doesn't make it the wrong method.

One of the more common mistakes with integration by parts is for people to get too locked into perceived patterns. For instance, all of the previous examples used the basic pattern of taking u to be the polynomial that sat in front of another function and then letting dv be the other function. This will not always happen so we need to be careful and not get locked into any patterns that we think we see.

Let's take a look at some integrals that don't fit into the above pattern.

Example 6

Evaluate the following integral.

$$\int \ln(x)\,dx$$

Solution

So, unlike any of the other integral we've done to this point there is only a single function in the integral and no polynomial sitting in front of the logarithm.

The first choice of many people here is to try and fit this into the pattern from above and make the following choices for u and dv.

$$u = 1 \qquad dv = \ln(x)\,dx$$

This leads to a real problem however since that means v must be,

$$v = \int \ln(x)\,dx$$

In other words, we would need to know the answer ahead of time in order to actually do the problem. So, this choice simply won't work.

Therefore, if the logarithm doesn't belong in the dv it must belong instead in the u. So, let's use the following choices instead

$$\begin{aligned} u &= \ln(x) & dv &= dx \\ du &= \frac{1}{x}dx & v &= x \end{aligned}$$

The integral is then,

$$\begin{aligned}\int \ln(x)\,dx &= x\ln(x) - \int \frac{1}{x}\,x\,dx \\ &= x\ln(x) - \int dx \\ &= x\ln(x) - x + c\end{aligned}$$

Example 7

Evaluate the following integral.

$$\int x^5\sqrt{x^3+1}\,dx$$

Solution

So, if we again try to use the pattern from the first few examples for this integral our choices for u and dv would probably be the following.

$$u = x^5 \qquad\qquad dv = \sqrt{x^3+1}\,dx$$

However, as with the previous example this won't work since we can't easily compute v.

$$v = \int \sqrt{x^3+1}\,dx$$

This is not an easy integral to do. However, notice that if we had an x^2 in the integral along with the root we could very easily do the integral with a substitution. Also notice that we do have a lot of x's floating around in the original integral. So instead of putting all the x's (outside of the root) in the u let's split them up as follows.

$$\begin{aligned} u &= x^3 & dv &= x^2\sqrt{x^3+1}\,dx \\ du &= 3x^2dx & v &= \frac{2}{9}\left(x^3+1\right)^{\frac{3}{2}}\end{aligned}$$

We can now easily compute v and after using integration by parts we get,

$$\begin{aligned}\int x^5\sqrt{x^3+1}\,dx &= \frac{2}{9}x^3\left(x^3+1\right)^{\frac{3}{2}} - \frac{2}{3}\int x^2\left(x^3+1\right)^{\frac{3}{2}}\,dx \\ &= \frac{2}{9}x^3\left(x^3+1\right)^{\frac{3}{2}} - \frac{4}{45}\left(x^3+1\right)^{\frac{5}{2}} + c\end{aligned}$$

So, in the previous two examples we saw cases that didn't quite fit into any perceived pattern that we might have gotten from the first couple of examples. This is always something that we need to be on the lookout for with integration by parts.

Let's take a look at another example that also illustrates another integration technique that sometimes arises out of integration by parts problems.

Example 8

Evaluate the following integral.

$$\int \mathbf{e}^{\theta} \cos(\theta)\, d\theta$$

Solution

Okay, to this point we've always picked u in such a way that upon differentiating it would make that portion go away or at the very least put it the integral into a form that would make it easier to deal with. In this case no matter which part we make u it will never go away in the differentiation process.

It doesn't much matter which we choose to be u so we'll choose in the following way. Note however that we could choose the other way as well and we'll get the same result in the end.

$$u = \cos(\theta) \qquad dv = \mathbf{e}^{\theta}\, d\theta$$
$$du = -\sin(\theta)\, d\theta \qquad v = \mathbf{e}^{\theta}$$

The integral is then,

$$\int \mathbf{e}^{\theta} \cos(\theta)\, d\theta = \mathbf{e}^{\theta} \cos(\theta) + \int \mathbf{e}^{\theta} \sin(\theta)\, d\theta$$

So, it looks like we'll do integration by parts again. Here are our choices this time.

$$u = \sin(\theta) \qquad dv = \mathbf{e}^{\theta}\, d\theta$$
$$du = \cos(\theta)\, d\theta \qquad v = \mathbf{e}^{\theta}$$

The integral is now,

$$\int \mathbf{e}^{\theta} \cos(\theta)\, d\theta = \mathbf{e}^{\theta} \cos(\theta) + \mathbf{e}^{\theta} \sin(\theta) - \int \mathbf{e}^{\theta} \cos(\theta)\, d\theta$$

Now, at this point it looks like we're just running in circles. However, notice that we now have the same integral on both sides and on the right side it's got a minus sign in front of it. This means that we can add the integral to both sides to get,

$$2\int \mathbf{e}^{\theta} \cos(\theta)\, d\theta = \mathbf{e}^{\theta} \cos(\theta) + \mathbf{e}^{\theta} \sin(\theta)$$

All we need to do now is divide by 2 and we're done. The integral is,

$$\int \mathbf{e}^{\theta} \cos(\theta)\, d\theta = \frac{1}{2}\left(\mathbf{e}^{\theta}\cos(\theta) + \mathbf{e}^{\theta}\sin(\theta)\right) + c$$

Notice that after dividing by the two we add in the constant of integration at that point.

This idea of integrating until you get the same integral on both sides of the equal sign and then simply solving for the integral is kind of nice to remember. It doesn't show up all that often, but when it does it may be the only way to actually do the integral.

Note as well that this is really just Algebra, admittedly done in a way that you may not be used to seeing it, but it is really just Algebra.

At this stage of your mathematical career everyone can solve,

$$x = 3 - x \qquad \rightarrow \qquad x = \frac{3}{2}$$

We are still solving an "equation". The only difference is that instead of solving for an x in we are solving for an integral and instead of a nice constant, "3" in the above Algebra problem, we've got a "messier" function.

We've got one more example to do. As we will see some problems could require us to do integration by parts numerous times and there is a short hand method that will allow us to do multiple applications of integration by parts quickly and easily.

Example 9

Evaluate the following integral.

$$\int x^4 \mathbf{e}^{\frac{x}{2}}\, dx$$

Solution

We start off by choosing u and dv as we always would. However, instead of computing du and v we put these into the following table. We then differentiate down the column corresponding to u until we hit zero. In the column corresponding to dv we integrate once for each entry in the first column. There is also a third column which we will explain in a bit and it always starts with a ""+"" and then alternates signs as shown.

x^4	$\mathbf{e}^{\frac{x}{2}}$	$+$
$4x^3$	$2\mathbf{e}^{\frac{x}{2}}$	$-$
$12x^2$	$4\mathbf{e}^{\frac{x}{2}}$	$+$
$24x$	$8\mathbf{e}^{\frac{x}{2}}$	$-$
24	$16\mathbf{e}^{\frac{x}{2}}$	$+$
0	$32\mathbf{e}^{\frac{x}{2}}$	$-$

Now, multiply along the diagonals shown in the table. In front of each product put the sign in the third column that corresponds to the "u" term for that product. In this case this would give,

$$\begin{aligned}\int x^4\mathbf{e}^{\frac{x}{2}}\,dx &= \left(x^4\right)\left(2\mathbf{e}^{\frac{x}{2}}\right) - \left(4x^3\right)\left(4\mathbf{e}^{\frac{x}{2}}\right) + \left(12x^2\right)\left(8\mathbf{e}^{\frac{x}{2}}\right) - \left(24x\right)\left(16\mathbf{e}^{\frac{x}{2}}\right) + \left(24\right)\left(32\mathbf{e}^{\frac{x}{2}}\right)\\ &= 2x^4\mathbf{e}^{\frac{x}{2}} - 16x^3\mathbf{e}^{\frac{x}{2}} + 96x^2\mathbf{e}^{\frac{x}{2}} - 384x\mathbf{e}^{\frac{x}{2}} + 768\mathbf{e}^{\frac{x}{2}} + c\end{aligned}$$

We've got the integral. This is much easier than writing down all the various u's and dv's that we'd have to do otherwise.

So, in this section we've seen how to do integration by parts. In your later math classes this is liable to be one of the more frequent integration techniques that you'll encounter.

It is important to not get too locked into patterns that you may think you've seen. In most cases any pattern that you think you've seen can (and will be) violated at some point in time. Be careful!

7.2 Integrals Involving Trig Functions

In this section we are going to look at quite a few integrals involving trig functions and some of the techniques we can use to help us evaluate them. Let's start off with an integral that we should already be able to do.

$$\begin{aligned}\int \cos(x)\sin^5(x)\,dx &= \int u^5\,du \qquad \text{using the substitution } u = \sin(x)\\ &= \frac{1}{6}\sin^6(x) + c\end{aligned}$$

This integral is easy to do with a substitution because the presence of the cosine, however, what about the following integral.

Example 1

Evaluate the following integral.

$$\int \sin^5(x)\,dx$$

Solution

This integral no longer has the cosine in it that would allow us to use the substitution that we used above. Therefore, that substitution won't work and we are going to have to find another way of doing this integral.

Let's first notice that we could write the integral as follows,

$$\int \sin^5(x)\,dx = \int \sin^4(x)\,\sin(x)\,dx = \int \left(\sin^2(x)\right)^2\,\sin(x)\,dx$$

Now recall the trig identity,

$$\cos^2(x) + \sin^2(x) = 1 \qquad \Rightarrow \qquad \sin^2(x) = 1 - \cos^2(x)$$

With this identity the integral can be written as,

$$\int \sin^5(x)\,dx = \int \left(1 - \cos^2(x)\right)^2\,\sin(x)\,dx$$

and we can now use the substitution $u = \cos x$. Doing this gives us,

$$\begin{aligned}\int \sin^5(x)\,dx &= -\int \left(1 - u^2\right)^2\,du\\ &= -\int 1 - 2u^2 + u^4\,du\\ &= -\left(u - \frac{2}{3}u^3 + \frac{1}{5}u^5\right) + c\\ &= -\cos(x) + \frac{2}{3}\cos^3(x) - \frac{1}{5}\cos^5(x) + c\end{aligned}$$

So, with a little rewriting on the integrand we were able to reduce this to a fairly simple substitution.

Notice that we were able to do the rewrite that we did in the previous example because the exponent on the sine was odd. In these cases all that we need to do is strip out one of the sines. The exponent on the remaining sines will then be even and we can easily convert the remaining sines to cosines using the identity,

$$\cos^2(x) + \sin^2(x) = 1 \tag{7.1}$$

If the exponent on the sines had been even this would have been difficult to do. We could strip out a sine, but the remaining sines would then have an odd exponent and while we could convert them to cosines the resulting integral would often be even more difficult than the original integral in most cases.

Let's take a look at another example.

Example 2

Evaluate the following integral.

$$\int \sin^6(x)\cos^3(x)\,dx$$

Solution

So, in this case we've got both sines and cosines in the problem and in this case the exponent on the sine is even while the exponent on the cosine is odd. So, we can use a similar technique in this integral. This time we'll strip out a cosine and convert the rest to sines.

$$\begin{aligned}\int \sin^6(x)\cos^3(x)\,dx &= \int \sin^6(x)\cos^2(x)\,\cos(x)\,dx \\ &= \int \sin^6(x)\left(1-\sin^2(x)\right)\,\cos(x)\,dx \qquad\qquad u = \sin(x) \\ &= \int u^6\left(1-u^2\right)\,du \\ &= \int u^6 - u^8\,du \\ &= \frac{1}{7}\sin^7(x) - \frac{1}{9}\sin^9(x) + c\end{aligned}$$

At this point let's pause for a second to summarize what we've learned so far about integrating powers of sine and cosine.

$$\int \sin^n(x)\cos^m(x)\,dx \tag{7.2}$$

In this integral if the exponent on the sines (n) is odd we can strip out one sine, convert the rest to cosines using Equation 7.1 and then use the substitution $u = \cos(x)$. Likewise, if the exponent on the cosines (m) is odd we can strip out one cosine and convert the rest to sines and the use the substitution $u = \sin(x)$.

Of course, if both exponents are odd then we can use either method. However, in these cases it's usually easier to convert the term with the smaller exponent.

The one case we haven't looked at is what happens if both of the exponents are even? In this case the technique we used in the first couple of examples simply won't work and in fact there really isn't any one set method for doing these integrals. Each integral is different and in some cases there will be more than one way to do the integral.

With that being said most, if not all, of integrals involving products of sines and cosines in which both exponents are even can be done using one or more of the following formulas to rewrite the integrand.

$$\begin{aligned}\cos^2(x) &= \frac{1}{2}(1+\cos(2x))\\ \sin^2(x) &= \frac{1}{2}(1-\cos(2x))\\ \sin(x)\cos(x) &= \frac{1}{2}\sin(2x)\end{aligned}$$

The first two formulas are the standard half angle formula from a trig class written in a form that will be more convenient for us to use. The last is the standard double angle formula for sine, again with a small rewrite.

Let's take a look at an example.

Example 3

Evaluate the following integral.

$$\int \sin^2(x)\cos^2(x)\,dx$$

Solution

As noted above there are often more than one way to do integrals in which both of the exponents are even. This integral is an example of that. There are at least two solution techniques for this problem. We will do both solutions starting with what is probably the longer of the two, but it's also the one that many people see first.

Solution 1

In this solution we will use the two half angle formulas above and just substitute them into the integral.

$$\begin{aligned}\int \sin^2(x)\cos^2(x)\,dx &= \int \frac{1}{2}\left(1-\cos\left(2x\right)\right)\left(\frac{1}{2}\right)\left(1+\cos\left(2x\right)\right)dx \\ &= \frac{1}{4}\int 1-\cos^2\left(2x\right)\,dx\end{aligned}$$

So, we still have an integral that can't be completely done, however notice that we have managed to reduce the integral down to just one term causing problems (a cosine with an even power) rather than two terms causing problems.

In fact to eliminate the remaining problem term all that we need to do is reuse the first half angle formula given above.

$$\begin{aligned}\int \sin^2(x)\cos^2(x)\,dx &= \frac{1}{4}\int 1-\frac{1}{2}\left(1+\cos\left(4x\right)\right)dx \\ &= \frac{1}{4}\int \frac{1}{2}-\frac{1}{2}\cos\left(4x\right)\,dx \\ &= \frac{1}{4}\left(\frac{1}{2}x-\frac{1}{8}\sin\left(4x\right)\right)+c \\ &= \frac{1}{8}x-\frac{1}{32}\sin\left(4x\right)+c\end{aligned}$$

So, this solution required a total of three trig identities to complete.

Solution 2

In this solution we will use the double angle formula to help simplify the integral as follows.

$$\begin{aligned}\int \sin^2(x)\cos^2(x)\,dx &= \int \left(\sin(x)\cos(x)\right)^2 dx \\ &= \int \left(\frac{1}{2}\sin\left(2x\right)\right)^2 dx \\ &= \frac{1}{4}\int \sin^2\left(2x\right)\,dx\end{aligned}$$

Now, we use the half angle formula for sine to reduce to an integral that we can do.

$$\begin{aligned}\int \sin^2(x)\cos^2(x)\,dx &= \frac{1}{8}\int 1-\cos\left(4x\right)\,dx \\ &= \frac{1}{8}x-\frac{1}{32}\sin\left(4x\right)+c\end{aligned}$$

This method required only two trig identities to complete.

Notice that the difference between these two methods is more one of "messiness". The second method is not appreciably easier (other than needing one less trig identity) it is just not as messy and that will often translate into an "easier" process.

In the previous example we saw two different solution methods that gave the same answer. Note that this will not always happen. In fact, more often than not we will get different answers. However, as we discussed in the Integration by Parts section, the two answers will differ by no more than a constant.

In general, when we have products of sines and cosines in which both exponents are even we will need to use a series of half angle and/or double angle formulas to reduce the integral into a form that we can integrate. Also, the larger the exponents the more we'll need to use these formulas and hence the messier the problem.

Sometimes in the process of reducing integrals in which both exponents are even we will run across products of sine and cosine in which the arguments are different. These will require one of the following formulas to reduce the products to integrals that we can do.

$$\begin{aligned}\sin(\alpha)\cos(\beta) &= \frac{1}{2}\left[\sin(\alpha-\beta)+\sin(\alpha+\beta)\right]\\ \sin(\alpha)\sin(\beta) &= \frac{1}{2}\left[\cos(\alpha-\beta)-\cos(\alpha+\beta)\right]\\ \cos(\alpha)\cos(\beta) &= \frac{1}{2}\left[\cos(\alpha-\beta)+\cos(\alpha+\beta)\right]\end{aligned}$$

Let's take a look at an example of one of these kinds of integrals.

Example 4

Evaluate the following integral.

$$\int \cos(15x)\cos(4x)\,dx$$

Solution

This integral requires the last formula listed above.

$$\begin{aligned}\int \cos(15x)\cos(4x)\,dx &= \frac{1}{2}\int \cos(11x)+\cos(19x)\,dx\\ &= \frac{1}{2}\left(\frac{1}{11}\sin(11x)+\frac{1}{19}\sin(19x)\right)+c\end{aligned}$$

Okay, at this point we've covered pretty much all the possible cases involving products of sines and cosines. It's now time to look at integrals that involve products of secants and tangents.

This time, let's do a little analysis of the possibilities before we just jump into examples. The general integral will be,

$$\int \sec^n(x)\tan^m(x)\,dx \tag{7.3}$$

The first thing to notice is that we can easily convert even powers of secants to tangents and even powers of tangents to secants by using a formula similar to Equation 7.1. In fact, the formula can be derived from Equation 7.1 so let's do that.

$$\begin{aligned} \sin^2(x) + \cos^2(x) &= 1 \\ \frac{\sin^2(x)}{\cos^2(x)} + \frac{\cos^2(x)}{\cos^2(x)} &= \frac{1}{\cos^2(x)} \\ \tan^2(x) + 1 &= \sec^2(x) \end{aligned} \tag{7.4}$$

Now, we're going to want to deal with Equation 7.3 similarly to how we dealt with Equation 7.2. We'll want to eventually use one of the following substitutions.

$$\begin{aligned} u &= \tan(x) & du &= \sec^2(x)\,dx \\ u &= \sec(x) & du &= \sec(x)\tan(x)\,dx \end{aligned}$$

So, if we use the substitution $u = \tan(x)$ we will need two secants left for the substitution to work. This means that if the exponent on the secant (n) is even we can strip two out and then convert the remaining secants to tangents using Equation 7.4.

Next, if we want to use the substitution $u = \sec(x)$ we will need one secant and one tangent left over in order to use the substitution. This means that if the exponent on the tangent (m) is odd and we have at least one secant in the integrand we can strip out one of the tangents along with one of the secants of course. The tangent will then have an even exponent and so we can use Equation 7.4 to convert the rest of the tangents to secants. Note that this method does require that we have at least one secant in the integral as well. If there aren't any secants then we'll need to do something different.

If the exponent on the secant is even and the exponent on the tangent is odd then we can use either case. Again, it will be easier to convert the term with the smallest exponent.

Let's take a look at a couple of examples.

Example 5

Evaluate the following integral.

$$\int \sec^9(x)\tan^5(x)\,dx$$

Solution

First note that since the exponent on the secant isn't even we can't use the substitution $u = \tan(x)$. However, the exponent on the tangent is odd and we've got a secant in the integral and so we will be able to use the substitution $u = \sec(x)$. This means stripping out a single tangent (along with a secant) and converting the remaining tangents to secants using Equation 7.4.

Here's the work for this integral.

$$\begin{aligned}
\int \sec^9(x)\tan^5(x)\,dx &= \int \sec^8(x)\tan^4(x)\,\tan(x)\sec(x)\,dx \\
&= \int \sec^8(x)\left(\sec^2(x)-1\right)^2\,\tan(x)\sec(x)\,dx \qquad u=\sec(x) \\
&= \int u^8\left(u^2-1\right)^2\,du \\
&= \int u^{12}-2u^{10}+u^8\,du \\
&= \frac{1}{13}\sec^{13}(x)-\frac{2}{11}\sec^{11}(x)+\frac{1}{9}\sec^9(x)+c
\end{aligned}$$

Example 6

Evaluate the following integral.

$$\int \sec^4(x)\tan^6(x)\,dx$$

Solution

So, in this example the exponent on the tangent is even so the substitution $u = \sec(x)$ won't work. The exponent on the secant is even and so we can use the substitution $u = \tan(x)$ for this integral. That means that we need to strip out two secants and convert the rest to tangents. Here is the work for this integral.

$$\begin{aligned}
\int \sec^4(x)\tan^6(x)\,dx &= \int \sec^2(x)\tan^6(x)\sec^2(x)\,dx \\
&= \int \left(\tan^2(x)+1\right)\tan^6(x)\sec^2(x)\,dx \qquad u=\tan(x) \\
&= \int \left(u^2+1\right)u^6\,du \\
&= \int u^8+u^6\,du \\
&= \frac{1}{9}\tan^9(x)+\frac{1}{7}\tan^7(x)+c
\end{aligned}$$

Both of the previous examples fit very nicely into the patterns discussed above and so were not all that difficult to work. However, there are a couple of exceptions to the patterns above and in these cases there is no single method that will work for every problem. Each integral will be different and may require different solution methods in order to evaluate the integral.

Let's first take a look at a couple of integrals that have odd exponents on the tangents, but no secants. In these cases we can't use the substitution $u = \sec(x)$ since it requires there to be at least one secant in the integral.

Example 7

Evaluate the following integral.

$$\int \tan(x)\,dx$$

Solution

To do this integral all we need to do is recall the definition of tangent in terms of sine and cosine and then this integral is nothing more than a Calculus I substitution.

$$\begin{aligned} \int \tan(x)\,dx &= \int \frac{\sin(x)}{\cos(x)}\,dx & u = \cos(x) \\ &= -\int \frac{1}{u}\,du \\ &= -\ln\left|\cos(x)\right| + c & r\ln(x) = \ln\left(x^r\right) \\ &= \ln\left|\cos(x)\right|^{-1} + c \\ &= \ln\left|\sec(x)\right| + c \end{aligned}$$

Note that for many folks,

$$\int \tan(x)\,dx = -\ln\left|\cos(x)\right| + c$$

We went a step or two further with some simplification. The simplification was done solely to eliminate the minus sign that was in front of the logarithm. This does not have to be done in general, but it is always easy to lose minus signs and in this case it was easy to eliminate it without introducing any real complexity to the answer and so we did.

Example 8

Evaluate the following integral.

$$\int \tan^3(x)\,dx$$

Solution

The trick to this one is do the following manipulation of the integrand.

$$\begin{aligned}\int \tan^3(x)\,dx &= \int \tan(x)\tan^2(x)\,dx\\ &= \int \tan(x)\left(\sec^2(x)-1\right)\,dx\\ &= \int \tan(x)\sec^2(x)\,dx - \int \tan(x)\,dx\end{aligned}$$

We can now use the substitution $u = \tan x$ on the first integral and the results from the previous example on the second integral.

The integral is then,

$$\int \tan^3(x)\,dx = \frac{1}{2}\tan^2(x) - \ln\left|\sec(x)\right| + c$$

Note that all odd powers of tangent (with the exception of the first power) can be integrated using the same method we used in the previous example. For instance,

$$\int \tan^5(x)\,dx = \int \tan^3(x)\left(\sec^2(x)-1\right)\,dx = \int \tan^3(x)\sec^2(x)\,dx - \int \tan^3(x)\,dx$$

So, a quick substitution ($u = \tan(x)$) will give us the first integral and the second integral will always be the previous odd power.

Now let's take a look at a couple of examples in which the exponent on the secant is odd and the exponent on the tangent is even. In these cases the substitutions used above won't work.

It should also be noted that both of the following two integrals are integrals that we'll be seeing on occasion in later sections of this chapter and in later chapters. Because of this it wouldn't be a bad idea to make a note of these results so you'll have them ready when you need them later.

Example 9

Evaluate the following integral.

$$\int \sec(x)\,dx$$

Solution

This one isn't too bad once you see what you've got to do. By itself the integral can't be done. However, if we manipulate the integrand as follows we can do it.

$$\begin{aligned}\int \sec(x)\,dx &= \int \frac{\sec(x)\left(\sec(x)+\tan(x)\right)}{\sec(x)+\tan(x)}\,dx\\ &= \int \frac{\sec^2(x)+\tan(x)\sec(x)}{\sec(x)+\tan(x)}\,dx\end{aligned}$$

In this form we can do the integral using the substitution $u = \sec(x) + \tan(x)$. Doing this gives,

$$\int \sec(x)\,dx = \ln\left|\sec(x)+\tan(x)\right| + c$$

The idea used in the above example is a nice idea to keep in mind. Multiplying the numerator and denominator of a term by the same term above can, on occasion, put the integral into a form that can be integrated. Note that this method won't always work and even when it does it won't always be clear what you need to multiply the numerator and denominator by. However, when it does work and you can figure out what term you need it can greatly simplify the integral.

Here's the next example.

Example 10

Evaluate the following integral.

$$\int \sec^3(x)\,dx$$

Solution

This one is different from any of the other integrals that we've done in this section. The first step to doing this integral is to perform integration by parts using the following choices for u and dv.

$$\begin{aligned} u &= \sec(x) & dv &= \sec^2(x)\,dx\\ du &= \sec(x)\tan(x)\,dx & v &= \tan(x)\end{aligned}$$

Note that using integration by parts on this problem is not an obvious choice, but it does work very nicely here. After doing integration by parts we have,

$$\int \sec^3(x)\,dx = \sec(x)\tan(x) - \int \sec(x)\tan^2(x)\,dx$$

Now the new integral also has an odd exponent on the secant and an even exponent on the tangent and so the previous examples of products of secants and tangents still won't do us any good.

To do this integral we'll first write the tangents in the integral in terms of secants. Again, this is not necessarily an obvious choice but it's what we need to do in this case.

$$\begin{aligned}\int \sec^3(x)\,dx &= \sec(x)\tan(x) - \int \sec(x)\left(\sec^2(x)-1\right)dx \\ &= \sec(x)\tan(x) - \int \sec^3(x)\,dx + \int \sec(x)\,dx\end{aligned}$$

Now, we can use the results from the previous example to do the second integral and notice that the first integral is exactly the integral we're being asked to evaluate with a minus sign in front. So, add it to both sides to get,

$$2\int \sec^3(x)\,dx = \sec(x)\tan(x) + \ln\left|\sec(x)+\tan(x)\right|$$

Finally divide by two and we're done.

$$\int \sec^3(x)\,dx = \frac{1}{2}\Big(\sec(x)\tan(x) + \ln\left|\sec(x)+\tan(x)\right|\Big) + c$$

Again, note that we've again used the idea of integrating the right side until the original integral shows up and then moving this to the left side and dividing by its coefficient to complete the evaluation. We first saw this in the Integration by Parts section and noted at the time that this was a nice technique to remember. Here is another example of this technique.

Now that we've looked at products of secants and tangents let's also acknowledge that because we can relate cosecants and cotangents by

$$1 + \cot^2(x) = \csc^2(x)$$

all of the work that we did for products of secants and tangents will also work for products of cosecants and cotangents. We'll leave it to you to verify that.

There is one final topic to be discussed in this section before moving on.

To this point we've looked only at products of sines and cosines and products of secants and tangents. However, the methods used to do these integrals can also be used on some quotients involving sines and cosines and quotients involving secants and tangents (and hence quotients involving cosecants and cotangents).

Let's take a quick look at an example of this.

Example 11

Evaluate the following integral.

$$\int \frac{\sin^7(x)}{\cos^4(x)}\,dx$$

Solution

If this were a product of sines and cosines we would know what to do. We would strip out a sine (since the exponent on the sine is odd) and convert the rest of the sines to cosines. The same idea will work in this case. We'll strip out a sine from the numerator and convert the rest to cosines as follows,

$$\begin{aligned}\int \frac{\sin^7(x)}{\cos^4(x)}\,dx &= \int \frac{\sin^6(x)}{\cos^4(x)}\sin(x)\,dx\\ &= \int \frac{\left(\sin^2(x)\right)^3}{\cos^4(x)}\sin(x)\,dx\\ &= \int \frac{\left(1-\cos^2(x)\right)^3}{\cos^4(x)}\sin(x)\,dx\end{aligned}$$

At this point all we need to do is use the substitution $u = \cos(x)$ and we're done.

$$\begin{aligned}\int \frac{\sin^7(x)}{\cos^4(x)}\,dx &= -\int \frac{\left(1-u^2\right)^3}{u^4}\,du\\ &= -\int u^{-4} - 3u^{-2} + 3 - u^2\,du\\ &= -\left(-\frac{1}{3}\frac{1}{u^3} + 3\frac{1}{u} + 3u - \frac{1}{3}u^3\right) + c\\ &= \frac{1}{3\cos^3(x)} - \frac{3}{\cos(x)} - 3\cos(x) + \frac{1}{3}\cos^3(x) + c\end{aligned}$$

So, under the right circumstances, we can use the ideas developed to help us deal with products of trig functions to deal with quotients of trig functions. The natural question then, is just what are the right circumstances?

First notice that if the quotient had been reversed as in this integral,

$$\int \frac{\cos^4(x)}{\sin^7(x)}\,dx$$

we wouldn't have been able to strip out a sine.

$$\int \frac{\cos^4(x)}{\sin^7(x)}\,dx = \int \frac{\cos^4(x)}{\sin^6(x)}\,\frac{1}{\sin(x)}\,dx$$

In this case the "stripped out" sine remains in the denominator and it won't do us any good for the substitution $u = \cos(x)$ since this substitution requires a sine in the numerator of the quotient. Also note that, while we could convert the sines to cosines, the resulting integral would still be a fairly difficult integral.

So, we can use the methods we applied to products of trig functions to quotients of trig functions provided the term that needs parts stripped out in is the numerator of the quotient.

7.3 Trig Substitutions

As we have done in the last couple of sections, let's start off with a couple of integrals that we should already be able to do with a standard substitution.

$$\int x\sqrt{25x^2-4}\,dx = \frac{1}{75}\left(25x^2-4\right)^{\frac{3}{2}}+c \qquad \int \frac{x}{\sqrt{25x^2-4}}\,dx = \frac{1}{25}\sqrt{25x^2-4}+c$$

Both of these used the substitution $u = 25x^2 - 4$ and at this point should be pretty easy for you to do. However, let's take a look at the following integral.

Example 1

Evaluate the following integral.

$$\int \frac{\sqrt{25x^2-4}}{x}\,dx$$

Solution

In this case the substitution $u = 25x^2 - 4$ will not work (we don't have the $x\,dx$ in the numerator the substitution needs) and so we're going to have to do something different for this integral.

It would be nice if we could reduce the two terms in the root down to a single term somehow. The following substitution will do that for us.

$$x = \frac{2}{5}\sec(\theta)$$

Do not worry about where this came from at this point. As we work the problem you will see that it works and that if we have a similar type of square root in the problem we can always use a similar substitution.

Before we actually do the substitution however let's verify the claim that this will allow us to reduce the two terms in the root to a single term.

$$\sqrt{25x^2-4} = \sqrt{25\left(\frac{4}{25}\right)\sec^2(\theta)-4} = \sqrt{4\left(\sec^2(\theta)-1\right)} = 2\sqrt{\sec^2(\theta)-1}$$

Now reduce the two terms to a single term all we need to do is recall the relationship,

$$\tan^2(\theta)+1 = \sec^2(\theta) \qquad \Rightarrow \qquad \sec^2(\theta)-1 = \tan^2(\theta)$$

Using this fact the square root becomes,

$$\sqrt{25x^2-4} = 2\sqrt{\tan^2(\theta)} = 2\left|\tan(\theta)\right|$$

So, not only were we able to reduce the two terms to a single term in the process we were able to easily eliminate the root as well!

Note, however, the presence of the absolute value bars. These are important. Recall that

$$\sqrt{x^2} = |x|$$

There should always be absolute value bars at this stage. If we knew that $\tan(\theta)$ was always positive or always negative we could eliminate the absolute value bars using,

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

Without limits we won't be able to determine if $\tan(\theta)$ is positive or negative, however, we will need to eliminate them in order to do the integral. Therefore, since we are doing an indefinite integral we will assume that $\tan(\theta)$ will be positive and so we can drop the absolute value bars. This gives,

$$\sqrt{25x^2 - 4} = 2\tan(\theta)$$

So, we were able to reduce the two terms under the root to a single term with this substitution and in the process eliminate the root as well. Eliminating the root is a nice side effect of this substitution as the problem will now become somewhat easier to do.

Let's now do the substitution and see what we get. In doing the substitution don't forget that we'll also need to substitute for the dx. This is easy enough to get from the substitution.

$$x = \frac{2}{5}\sec(\theta) \qquad \Rightarrow \qquad dx = \frac{2}{5}\sec(\theta)\tan(\theta)\,d\theta$$

Using this substitution the integral becomes,

$$\begin{aligned} \int \frac{\sqrt{25x^2-4}}{x}\,dx &= \int \frac{2\tan(\theta)}{\frac{2}{5}\sec(\theta)}\left(\frac{2}{5}\sec(\theta)\tan(\theta)\right)d\theta \\ &= 2\int \tan^2(\theta)\,d\theta \end{aligned}$$

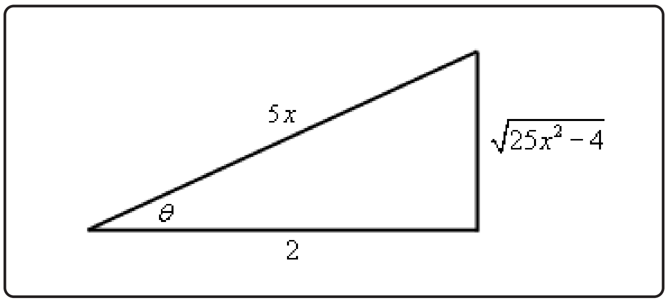
With this substitution we were able to reduce the given integral to an integral involving trig functions and we saw how to do these problems in the previous section. Let's finish the integral.

$$\begin{aligned} \int \frac{\sqrt{25x^2-4}}{x}\,dx &= 2\int \sec^2(\theta) - 1\,d\theta \\ &= 2\big(\tan(\theta) - \theta\big) + c \end{aligned}$$

So, we've got an answer for the integral. Unfortunately, the answer isn't given in x's as it should be. So, we need to write our answer in terms of x. We can do this with some right triangle trig. From our original substitution we have,

$$\sec(\theta) = \frac{5x}{2} = \frac{\text{hypotenuse}}{\text{adjacent}}$$

This gives the following right triangle.



From this we can see that,

$$\tan(\theta) = \frac{\sqrt{25x^2 - 4}}{2}$$

We can deal with the θ in one of any variety of ways. From our substitution we can see that,

$$\theta = \sec^{-1}\left(\frac{5x}{2}\right)$$

While this is a perfectly acceptable method of dealing with the θ we can use any of the possible six inverse trig functions and since sine and cosine are the two trig functions most people are familiar with we will usually use the inverse sine or inverse cosine. In this case we'll use the inverse cosine.

$$\theta = \cos^{-1}\left(\frac{2}{5x}\right)$$

So, with all of this the integral becomes,

$$\begin{aligned}\int \frac{\sqrt{25x^2 - 4}}{x}\,dx &= 2\left(\frac{\sqrt{25x^2 - 4}}{2} - \cos^{-1}\left(\frac{2}{5x}\right)\right) + c \\ &= \sqrt{25x^2 - 4} - 2\cos^{-1}\left(\frac{2}{5x}\right) + c\end{aligned}$$

We now have the answer back in terms of x.

Wow! That was a lot of work. Most of these won't take as long to work however. This first one needed lots of explanation since it was the first one. The remaining examples won't need quite as much explanation and so won't take as long to work.

However, before we move onto more problems let's first address the issue of definite integrals and how the process differs in these cases.

Example 2

Evaluate the following integral.

$$\int_{\frac{2}{5}}^{\frac{4}{5}} \frac{\sqrt{25x^2-4}}{x}\,dx$$

Solution

The limits here won't change the substitution so that will remain the same.

$$x = \frac{2}{5}\sec(\theta)$$

Using this substitution the square root still reduces down to,

$$\sqrt{25x^2-4} = 2\left|\tan(\theta)\right|$$

However, unlike the previous example we can't just drop the absolute value bars. In this case we've got limits on the integral and so we can use the limits as well as the substitution to determine the range of θ that we're in. Once we've got that we can determine how to drop the absolute value bars.

Here's the limits of θ and note that if you aren't good at solving trig equations in terms of secant you can always convert to cosine as we do below.

$$\begin{aligned} x=\frac{2}{5} &: \quad \frac{2}{5}=\frac{2}{5}\sec(\theta)=\frac{2}{5}\frac{1}{\cos(\theta)} \quad\rightarrow\quad \cos(\theta)=1 \quad\Rightarrow\quad \theta=\cos^{-1}(1)=0 \\ x=\frac{4}{5} &: \quad \frac{4}{5}=\frac{2}{5}\sec(\theta)=\frac{2}{5}\frac{1}{\cos(\theta)} \quad\rightarrow\quad \cos(\theta)=\frac{1}{2} \quad\Rightarrow\quad \theta=\cos^{-1}\left(\frac{1}{2}\right)=\frac{\pi}{3} \end{aligned}$$

Now, we know from solving trig equations, that there are in fact an infinite number of possible answers we could use. In fact, the more "correct" answer for the above work is,

$$\theta = 0+2\pi n = 2\pi n \qquad \& \qquad \theta=\frac{\pi}{3}+2\pi n \qquad n=0,\pm1,\pm2,\pm3,\ldots$$

So, which ones should we use? The answer is simple. When using a secant trig substitution and converting the limits we always assume that θ is in the range of inverse secant. Or,

$$\text{If } \theta=\sec^{-1}(x) \text{ then } 0\le\theta<\frac{\pi}{2} \text{ or } \frac{\pi}{2}<\theta\le\pi$$

Note that we have to avoid $\theta = \frac{\pi}{2}$ because secant will not exist at that point. Also note that the range of θ was given in terms of secant even though we actually used inverse cosine to get the answers. This will not be a problem because even though inverse cosine can give $\theta = \frac{\pi}{2}$ we'll never get it in our work above because that would require that we started with the secant being undefined and that will not happen when converting the limits as that would in turn require one of the limits to also be undefined!

So, in finding the new limits we didn't need all possible values of θ we just need the inverse cosine answers we got when we converted the limits. Therefore, if we are in the range $\frac{2}{5} \leq x \leq \frac{4}{5}$ then θ is in the range of $0 \leq \theta \leq \frac{\pi}{3}$ and in this range of θ's tangent is positive and so we can just drop the absolute value bars.

Let's do the substitution. Note that the work is identical to the previous example and so most of it is left out. We'll pick up at the final integral and then do the substitution.

$$\begin{aligned}\int_{\frac{2}{5}}^{\frac{4}{5}} \frac{\sqrt{25x^2-4}}{x}\,dx &= 2\int_0^{\frac{\pi}{3}} \sec^2(\theta) - 1\,d\theta \\ &= 2\big(\tan(\theta) - \theta\big)\Big|_0^{\pi/3} \\ &= 2\sqrt{3} - \frac{2\pi}{3}\end{aligned}$$

Note that because of the limits we didn't need to resort to a right triangle to complete the problem.

Let's take a look at a different set of limits for this integral.

Example 3

Evaluate the following integral.

$$\int_{-\frac{4}{5}}^{-\frac{2}{5}} \frac{\sqrt{25x^2-4}}{x}\,dx$$

Solution

Again, the substitution and square root are the same as the first two examples.

$$x = \frac{2}{5}\sec(\theta) \qquad \sqrt{25x^2-4} = 2\left|\tan(\theta)\right|$$

Let's next see the limits θ for this problem.

$$x = -\frac{2}{5} \; : \; -\frac{2}{5} = \frac{2}{5}\sec(\theta) = \frac{2}{5}\frac{1}{\cos(\theta)} \quad \rightarrow \quad \cos(\theta) = -1 \quad \Rightarrow \quad \theta = \cos^{-1}(-1) = \pi$$

$$x = -\frac{4}{5} \; : \; -\frac{4}{5} = \frac{2}{5}\sec(\theta) = \frac{2}{5}\frac{1}{\cos(\theta)} \quad \rightarrow \quad \cos(\theta) = -\frac{1}{2} \quad \Rightarrow \quad \theta = \cos^{-1}\left(-\frac{1}{2}\right) = \frac{2\pi}{3}$$

Remember that in converting the limits we use the results from the inverse secant/cosine. So, for this range of x's we have $\frac{2\pi}{3} \le \theta \le \pi$ and in this range of θ tangent is negative and so in this case we can drop the absolute value bars, but will need to add in a minus sign upon doing so. In other words,

$$\sqrt{25x^2 - 4} = -2\tan(\theta)$$

So, the only change this will make in the integration process is to put a minus sign in front of the integral. The integral is then,

$$\begin{aligned} \int_{-\frac{4}{5}}^{-\frac{2}{5}} \frac{\sqrt{25x^2-4}}{x}\,dx &= -2\int_{\frac{2\pi}{3}}^{\pi} \sec^2(\theta) - 1\,d\theta \\ &= -2\big(\tan(\theta) - \theta\big)\Big|_{2\pi/3}^{\pi} \\ &= \frac{2\pi}{3} - 2\sqrt{3} \end{aligned}$$

In the last two examples we saw that we have to be very careful with definite integrals. We need to make sure that we determine the limits on θ and whether or not this will mean that we can just drop the absolute value bars or if we need to add in a minus sign when we drop them.

Before moving on to the next example let's get the general form for the secant trig substitution that we used in the previous set of examples and the assumed limits on θ.

Let's work a new and different type of example.

Example 4

Evaluate the following integral.

$$\int \frac{1}{x^4\sqrt{9 - x^2}}\,dx$$

Solution

Now, the terms under the root in this problem looks to be (almost) the same as the previous ones so let's try the same type of substitution and see if it will work here as well.

$$x = 3\sec(\theta)$$

Using this substitution, the square root becomes,

$$\sqrt{9 - x^2} = \sqrt{9 - 9\sec^2(\theta)} = 3\sqrt{1 - \sec^2(\theta)} = 3\sqrt{-\tan^2(\theta)}$$

So, using this substitution we will end up with a negative quantity (the tangent squared is always positive of course) under the square root and this will be trouble. Using this substitution will give complex values and we don't want that. So, using secant for the substitution won't work.

However, the following substitution (and differential) will work.

$$x = 3\sin(\theta) \qquad\qquad dx = 3\cos(\theta)\,d\theta$$

With this substitution the square root is,

$$\sqrt{9 - x^2} = 3\sqrt{1 - \sin^2(\theta)} = 3\sqrt{\cos^2(\theta)} = 3\left|\cos(\theta)\right| = 3\cos(\theta)$$

We were able to drop the absolute value bars because we are doing an indefinite integral and so we'll assume that everything is positive.

The integral is now,

$$\begin{aligned}\int \frac{1}{x^4\sqrt{9 - x^2}}\,dx &= \int \frac{1}{81\sin^4(\theta)\left(3\cos(\theta)\right)}\,3\cos(\theta)\,d\theta\\ &= \frac{1}{81}\int \frac{1}{\sin^4(\theta)}\,d\theta\\ &= \frac{1}{81}\int \csc^4(\theta)\,d\theta\end{aligned}$$

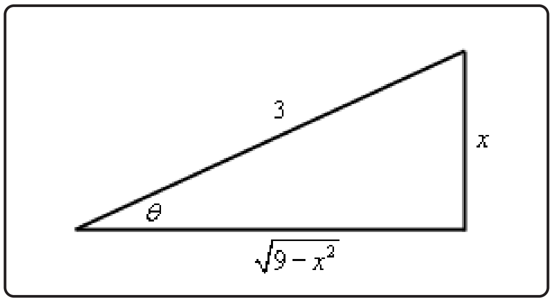
In the previous section we saw how to deal with integrals in which the exponent on the secant was even and since cosecants behave an awful lot like secants we should be able to do something similar with this.

Here is the integral.

$$\begin{aligned}\int \frac{1}{x^4\sqrt{9 - x^2}}\,dx &= \frac{1}{81}\int \csc^2(\theta)\,\csc^2(\theta)\,d\theta\\ &= \frac{1}{81}\int \left(\cot^2(\theta) + 1\right)\csc^2(\theta)\,d\theta \qquad\qquad u = \cot(\theta)\\ &= -\frac{1}{81}\int u^2 + 1\,du\\ &= -\frac{1}{81}\left(\frac{1}{3}\cot^3(\theta) + \cot(\theta)\right) + c\end{aligned}$$

Now we need to go back to x's using a right triangle. Here is the right triangle for this problem and trig functions for this problem.

$$\sin(\theta) = \frac{x}{3} \qquad \cot(\theta) = \frac{\sqrt{9-x^2}}{x}$$



The integral is then,

$$\begin{aligned}\int \frac{1}{x^4\sqrt{9-x^2}}\,dx &= -\frac{1}{81}\left(\frac{1}{3}\left(\frac{\sqrt{9-x^2}}{x}\right)^3 + \frac{\sqrt{9-x^2}}{x}\right) + c \\ &= -\frac{\left(9-x^2\right)^{\frac{3}{2}}}{243x^3} - \frac{\sqrt{9-x^2}}{81x} + c\end{aligned}$$

We aren't going to be doing a definite integral example with a sine trig substitution. However, if we had we would need to convert the limits and that would mean eventually needing to evaluate an inverse sine. So, much like with the secant trig substitution, the values of θ that we'll use will be those from the inverse sine or,

$$\text{If } \theta = \sin^{-1}(x) \text{ then } -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$$

Here is a summary for the sine trig substitution.

Fact

$$\sqrt{a^2 - b^2x^2} \quad \Rightarrow \quad x = \frac{a}{b}\sin(\theta), \quad -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$$

There is one final case that we need to look at. The next integral will also contain something that we need to make sure we can deal with.

Example 5

Evaluate the following integral.

$$\int_0^{\frac{1}{6}} \frac{x^5}{\left(36x^2+1\right)^{\frac{3}{2}}}\,dx$$

Solution

First, notice that there really is a square root in this problem even though it isn't explicitly written out. To see the root let's rewrite things a little.

$$\left(36x^2+1\right)^{\frac{3}{2}} = \left(\left(36x^2+1\right)^{\frac{1}{2}}\right)^3 = \left(\sqrt{36x^2+1}\right)^3$$

This terms under the root are not in the form we saw in the previous examples. Here we will use the substitution for this root.

$$x = \frac{1}{6}\tan(\theta) \qquad\qquad dx = \frac{1}{6}\sec^2(\theta)\,d\theta$$

With this substitution the denominator becomes,

$$\left(\sqrt{36x^2+1}\right)^3 = \left(\sqrt{\tan^2(\theta)+1}\right)^3 = \left(\sqrt{\sec^2(\theta)}\right)^3 = \left|\sec(\theta)\right|^3$$

Now, because we have limits we'll need to convert them to θ so we can determine how to drop the absolute value bars.

$$\begin{aligned} x = 0 &\;:\; 0 = \frac{1}{6}\tan(\theta) & \Rightarrow \qquad & \theta = \tan^{-1}(0) = 0 \\ x = \frac{1}{6} &\;:\; \frac{1}{6} = \frac{1}{6}\tan(\theta) & \Rightarrow \qquad & \theta = \tan^{-1}(1) = \frac{\pi}{4} \end{aligned}$$

As with the previous two cases when converting limits here we will use the results of the inverse tangent or,

$$\text{If } \theta = \tan^{-1}(x) \text{ then } -\frac{\pi}{2} < \theta < \frac{\pi}{2}$$

So, in this range of θ secant is positive and so we can drop the absolute value bars.

Here is the integral,

$$\begin{aligned} \int_0^{\frac{1}{6}} \frac{x^5}{\left(36x^2+1\right)^{\frac{3}{2}}}\,dx &= \int_0^{\frac{\pi}{4}} \frac{\frac{1}{7776}\tan^5(\theta)}{\sec^3(\theta)}\left(\frac{1}{6}\sec^2(\theta)\right)d\theta \\ &= \frac{1}{46656}\int_0^{\frac{\pi}{4}} \frac{\tan^5(\theta)}{\sec(\theta)}\,d\theta \end{aligned}$$

There are several ways to proceed from this point. Normally with an odd exponent on the tangent we would strip one of them out and convert to secants. However, that would require that we also have a secant in the numerator which we don't have. Therefore, it seems like the best way to do this one would be to convert the integrand to sines and cosines.

$$\begin{aligned}\int_0^{\frac{1}{6}} \frac{x^5}{(36x^2+1)^{\frac{3}{2}}}\,dx &= \frac{1}{46656}\int_0^{\frac{\pi}{4}} \frac{\sin^5(\theta)}{\cos^4(\theta)}\,d\theta \\ &= \frac{1}{46656}\int_0^{\frac{\pi}{4}} \frac{\left(1-\cos^2(\theta)\right)^2}{\cos^4(\theta)}\sin(\theta)\,d\theta\end{aligned}$$

We can now use the substitution $u = \cos(\theta)$ and we might as well convert the limits as well.

$$\begin{aligned}\theta &= 0 & u &= \cos(0) = 1 \\ \theta &= \frac{\pi}{4} & u &= \cos\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2}\end{aligned}$$

The integral is then,

$$\begin{aligned}\int_0^{\frac{1}{6}} \frac{x^5}{(36x^2+1)^{\frac{3}{2}}}\,dx &= -\frac{1}{46656}\int_1^{\frac{\sqrt{2}}{2}} u^{-4} - 2u^{-2} + 1\,du \\ &= -\frac{1}{46656}\left(-\frac{1}{3u^3} + \frac{2}{u} + u\right)\Bigg|_1^{\frac{\sqrt{2}}{2}} \\ &= \frac{1}{17496} - \frac{11\sqrt{2}}{279936}\end{aligned}$$

Here is a summary for this final type of trig substitution.

Fact

$$\sqrt{a^2+b^2x^2} \quad \Rightarrow \quad x = \frac{a}{b}\tan(\theta), \qquad -\frac{\pi}{2} < \theta < \frac{\pi}{2}$$

Before proceeding with some more examples let's discuss just how we knew to use the substitutions that we did in the previous examples.

The main idea was to determine a substitution that would allow us to reduce the two terms under the root that was always in the problem (more on this in a bit) into a single term and in doing so we were also able to easily eliminate the root. To do this we made use of the following formulas.

$$\begin{aligned}25x^2 - 4 \quad &\Rightarrow \quad \sec^2(\theta) - 1 = \tan^2(\theta) \\ 9 - x^2 \quad &\Rightarrow \quad 1 - \sin^2(\theta) = \cos^2(\theta) \\ 36x^2 + 1 \quad &\Rightarrow \quad \tan^2(\theta) + 1 = \sec^2(\theta)\end{aligned}$$

If we step back a bit we can notice that the terms we reduced look like the trig identities we used to reduce them in a vague way.

For instance, $25x^2 - 4$ is something squared (*i.e.* the $25x^2$) minus a number (*i.e.* the 4) and the left side of formula we used, $\sec^2(\theta) - 1$, also follows this basic form. So, because the two look alike in a very vague way that suggests using a secant substitution for that problem. We can notice similar vague similarities in the other two cases as well.

If we keep this idea in mind we don't need the "formulas" listed after each example to tell us which trig substitution to use and since we have to know the trig identities anyway to do the problems keeping this idea in mind doesn't really add anything to what we need to know for the problems.

Once we've identified the trig function to use in the substitution the coefficient, the $\frac{a}{b}$ in the formulas, is also easy to get. Just remember that in order to use the trig identities the coefficient of the trig function and the number in the identity must be the same, *i.e.* both 4 or 9, so that the trig identity can be used after we factor the common number out. What this means is that we need to "turn" the coefficient of the squared term into the constant number through our substitution.

So, in the first example we needed to "turn" the 25 into a 4 through our substitution. Remembering that we are eventually going to square the substitution that means we need to divide out by a 5 so the 25 will cancel out, upon squaring. Likewise, we'll need to add a 2 to the substitution so the coefficient will "turn" into a 4 upon squaring. In other words, we would need to use the substitution that we did in the problem.

The same idea holds for the other two trig substitutions.

Notice as well that we could have used cosecant in the first case, cosine in the second case and cotangent in the third case. So, why didn't we? Simply because of the differential work. Had we used these trig functions instead we would have picked up a minus sign in the differential that we'd need to keep track of. So, while these could be used they generally aren't to avoid extra minus signs that we need to keep track of.

Next, let's quickly address the fact that a root was in all of these problems. Note that the root is not required in order to use a trig substitution. Instead, the trig substitution gave us a really nice way of eliminating the root from the problem. In this section we will always be having roots in the problems, and in fact our summaries above all assumed roots, roots are not actually required in order use a trig substitution. We will be seeing an example or two of trig substitutions in integrals that do not have roots in the Integrals Involving Quadratics section.

Finally, let's summarize up all the ideas with the trig substitutions we've discussed and again we will be using roots in the summary simply because all the integrals in this section will have roots and those tend to be the most likely places for using trig substitutions but again, are not required in order to use a trig substitution.

Fact

Form	Looks Like	Substitution	Limit Assumptions
$\sqrt{b^2x^2 - a^2}$	$\sec^2(\theta) - 1 = \tan^2\theta$	$x = \frac{a}{b}\sec(\theta)$	$0 \leq \theta < \frac{\pi}{2}, \frac{\pi}{2} < \theta \leq \pi$
$\sqrt{a^2 - b^2x^2}$	$1 - \sin^2(\theta) = \cos^2(\theta)$	$x = \frac{a}{b}\sin(\theta)$	$-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$
$\sqrt{a^2 + b^2x^2}$	$\tan^2(\theta) + 1 = \sec^2(\theta)$	$x = \frac{a}{b}\tan(\theta)$	$-\frac{\pi}{2} < \theta < \frac{\pi}{2}$

Now, we have a couple of final examples to work in this section. Not all trig substitutions will just jump right out at us. Sometimes we need to do a little work on the integrand first to get it into the correct form and that is the point of the remaining examples.

Example 6

Evaluate the following integral.

$$\int \frac{x}{\sqrt{2x^2 - 4x - 7}}\,dx$$

Solution

In this case the quantity under the root doesn't obviously fit into any of the cases we looked at above and in fact isn't in the any of the forms we saw in the previous examples. Note however that if we complete the square on the quadratic we can make it look somewhat like the above integrals.

Remember that completing the square requires a coefficient of one in front of the x^2. Once we have that we take half the coefficient of the x, square it, and then add and subtract it to the quantity. Here is the completing the square for this problem.

$$2\left(x^2 - 2x - \frac{7}{2}\right) = 2\left(x^2 - 2x + 1 - 1 - \frac{7}{2}\right) = 2\left((x-1)^2 - \frac{9}{2}\right) = 2(x-1)^2 - 9$$

So, the root becomes,

$$\sqrt{2x^2 - 4x - 7} = \sqrt{2(x-1)^2 - 9}$$

Now, this looks (very) vaguely like $\sec^2(\theta) - 1$ (*i.e.* something squared minus a number) except we've got something more complicated in the squared term. That is okay we'll still be able to do a secant substitution and it will work in pretty much the same way.

$$x - 1 = \frac{3}{\sqrt{2}}\sec(\theta) \qquad x = 1 + \frac{3}{\sqrt{2}}\sec(\theta) \qquad dx = \frac{3}{\sqrt{2}}\sec(\theta)\tan(\theta)\,d\theta$$

Using this substitution the root reduces to,

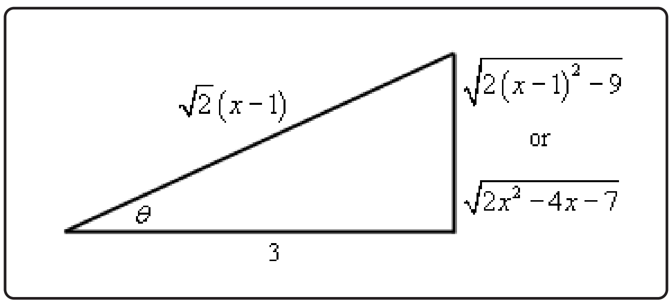
$$\sqrt{2x^2-4x-7}=\sqrt{2(x-1)^2-9}=\sqrt{9\sec^2(\theta)-9}=3\sqrt{\tan^2(\theta)}=3\left|\tan(\theta)\right|=3\tan(\theta)$$

Note we could drop the absolute value bars since we are doing an indefinite integral. Here is the integral.

$$\begin{aligned}\int\frac{x}{\sqrt{2x^2-4x-7}}\,dx&=\int\frac{1+\frac{3}{\sqrt{2}}\sec(\theta)}{3\tan(\theta)}\left(\frac{3}{\sqrt{2}}\sec(\theta)\tan(\theta)\right)\,d\theta\\&=\int\frac{1}{\sqrt{2}}\sec(\theta)+\frac{3}{2}\sec^2(\theta)\,d\theta\\&=\frac{1}{\sqrt{2}}\ln\left|\sec(\theta)+\tan(\theta)\right|+\frac{3}{2}\tan(\theta)+c\end{aligned}$$

And here is the right triangle for this problem.

$$\sec(\theta)=\frac{\sqrt{2}\,(x-1)}{3}\qquad\qquad\tan(\theta)=\frac{\sqrt{2x^2-4x-7}}{3}$$



The integral is then,

$$\int\frac{x}{\sqrt{2x^2-4x-7}}\,dx=\frac{1}{\sqrt{2}}\ln\left|\frac{\sqrt{2}\,(x-1)}{3}+\frac{\sqrt{2x^2-4x-7}}{3}\right|+\frac{\sqrt{2x^2-4x-7}}{2}+c$$

Example 7

Evaluate the following integral.

$$\int \mathbf{e}^{4x}\sqrt{1+\mathbf{e}^{2x}}\,dx$$

Solution

This doesn't look to be anything like the other problems in this section. However it is. To see this we first need to notice that,

$$\mathbf{e}^{2x} = (\mathbf{e}^{x})^2$$

Upon noticing this we can use the following standard Calculus I substitution.

$$u = \mathbf{e}^{x} \qquad\qquad du = \mathbf{e}^{x}\,dx$$

We do need to be a little careful with the differential work however. We don't have just an $\mathbf{e}^{x}$ out in front of the root. Instead we have an $\mathbf{e}^{4x}$. So, we'll need to strip one of those out for the differential and then use the substitution on the rest. Here is the substitution work.

$$\begin{aligned}\int \mathbf{e}^{4x}\sqrt{1+\mathbf{e}^{2x}}\,dx &= \int \mathbf{e}^{3x}\mathbf{e}^{x}\sqrt{1+\mathbf{e}^{2x}}\,dx\\ &= \int (\mathbf{e}^{x})^3\sqrt{1+(\mathbf{e}^{x})^2}\,\mathbf{e}^{x}dx = \int u^3\sqrt{1+u^2}\,du\end{aligned}$$

This is now a fairly obvious trig substitution (hopefully). The quantity under the root looks almost exactly like $1+\tan^2(\theta)$ and so we can use a tangent substitution. Here is that work.

$$u = \tan(\theta) \qquad du = \sec^2(\theta)\,d\theta \qquad\qquad \sqrt{1+u^2} = \sqrt{1+\tan^2\theta} = \sqrt{\sec^2(\theta)} = |\sec(\theta)|$$

Because we are doing an indefinite integral we can assume secant is positive and drop the absolute value bars. Applying this substitution to the integral gives,

$$\int \mathbf{e}^{4x}\sqrt{1+\mathbf{e}^{2x}}\,dx = \int \tan^3(\theta)\,(\sec(\theta))\,\left(\sec^2(\theta)\right)d\theta = \int \tan^3(\theta)\sec^3(\theta)d\theta$$

We'll finish this integral off in a bit. Before we get to that there is a "quicker" (although not super obvious) way of doing the substitutions above. Let's cover that first then we'll come back and finish working the integral.

We can notice that the u in the Calculus I substitution and the trig substitution are the same u and so we can combine them into the following substitution.

$$\mathbf{e}^{x} = \tan(\theta)$$

We can then compute the differential. Just remember that all we do is differentiate both sides and then tack on dx or $d\theta$ onto the appropriate side. Doing this gives,

$$\mathbf{e}^x\,dx = \sec^2(\theta)\,d\theta$$

With this substitution the square root becomes,

$$\sqrt{1+\mathbf{e}^{2x}} = \sqrt{1+(\mathbf{e}^x)^2} = \sqrt{1+\tan^2(\theta)} = \sqrt{\sec^2(\theta)} = |\sec(\theta)| = \sec(\theta)$$

Again, we can drop the absolute value bars because we are doing an indefinite integral. The integral then becomes,

$$\begin{aligned}\int \mathbf{e}^{4x}\sqrt{1+\mathbf{e}^{2x}}\,dx &= \int \mathbf{e}^{3x}\mathbf{e}^x\sqrt{1+\mathbf{e}^{2x}}\,dx \\ &= \int (\mathbf{e}^x)^3\sqrt{1+\mathbf{e}^{2x}}\,(\mathbf{e}^x)\,dx \\ &= \int \tan^3(\theta)\,(\sec(\theta))\left(\sec^2(\theta)\right)\,d\theta = \int \tan^3(\theta)\sec^3(\theta)d\theta\end{aligned}$$

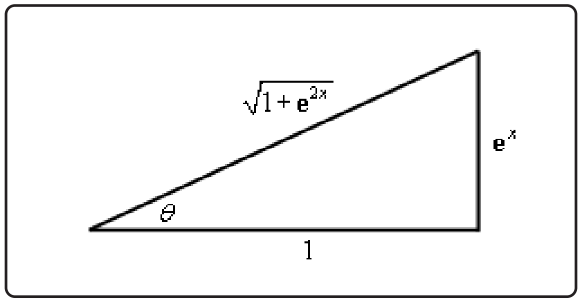
So, the same integral with less work. However, it does require that you be able to combine the two substitutions in to a single substitution. How you do this type of problem is up to you but if you don't feel comfortable with the single substitution (and there's nothing wrong if you don't!) then just do the two individual substitutions. The single substitution method was given only to show you that it can be done so that those that are really comfortable with both kinds of substitutions can do the work a little quicker.

Now, let's finish the integral work.

$$\begin{aligned}\int \mathbf{e}^{4x}\sqrt{1+\mathbf{e}^{2x}}\,dx &= \int \tan^3(\theta)\sec^3(\theta)d\theta \\ &= \int \left(\sec^2(\theta)-1\right)\sec^2(\theta)\,\sec(\theta)\tan(\theta)\,d\theta \qquad v=\sec(\theta) \\ &= \int v^4 - v^2\,dv \\ &= \frac{1}{5}\sec^5(\theta) - \frac{1}{3}\sec^3(\theta) + c\end{aligned}$$

Here is the right triangle for this integral.

$$\tan(\theta) = \frac{\mathbf{e}^x}{1} \qquad \sec(\theta) = \frac{\sqrt{1+\mathbf{e}^{2x}}}{1} = \sqrt{1+\mathbf{e}^{2x}}$$



The integral is then,

$$\int \mathbf{e}^{4x}\sqrt{1+\mathbf{e}^{2x}}\,dx = \frac{1}{5}\left(1+\mathbf{e}^{2x}\right)^{\frac{5}{2}} - \frac{1}{3}\left(1+\mathbf{e}^{2x}\right)^{\frac{3}{2}} + c$$

This was a messy problem, but we will be seeing some of this type of integral in later sections on occasion so we needed to make sure you'd seen at least one like it.

So, as we've seen in the final two examples in this section some integrals that look nothing like the first few examples can in fact be turned into a trig substitution problem with a little work.

7.4 Partial Fractions

In this section we are going to take a look at integrals of rational expressions of polynomials and once again let's start this section out with an integral that we can already do so we can contrast it with the integrals that we'll be doing in this section.

$$\begin{aligned} \int \frac{2x-1}{x^2-x-6}\,dx &= \int \frac{1}{u}\,du \qquad \text{using } u = x^2-x-6 \text{ and } du = (2x-1)\,dx \\ &= \ln\left|x^2-x-6\right| + c \end{aligned}$$

So, if the numerator is the derivative of the denominator (or a constant multiple of the derivative of the denominator) doing this kind of integral is fairly simple. However, often the numerator isn't the derivative of the denominator (or a constant multiple). For example, consider the following integral.

$$\int \frac{3x+11}{x^2-x-6}\,dx$$

In this case the numerator is definitely not the derivative of the denominator nor is it a constant multiple of the derivative of the denominator. Therefore, the simple substitution that we used above won't work. However, if we notice that the integrand can be broken up as follows,

$$\frac{3x+11}{x^2-x-6} = \frac{4}{x-3} - \frac{1}{x+2}$$

then the integral is actually quite simple.

$$\begin{aligned} \int \frac{3x+11}{x^2-x-6}\,dx &= \int \frac{4}{x-3} - \frac{1}{x+2}dx \\ &= 4\ln|x-3| - \ln|x+2| + c \end{aligned}$$

This process of taking a rational expression and decomposing it into simpler rational expressions that we can add or subtract to get the original rational expression is called **partial fraction decomposition**. Many integrals involving rational expressions can be done if we first do partial fractions on the integrand.

So, let's do a quick review of partial fractions. We'll start with a rational expression in the form,

$$f(x) = \frac{P(x)}{Q(x)}$$

where both $P(x)$ and $Q(x)$ are polynomials and the degree of $P(x)$ is smaller than the degree of $Q(x)$. Recall that the degree of a polynomial is the largest exponent in the polynomial. Partial fractions can only be done if the degree of the numerator is strictly less than the degree of the denominator. That is important to remember.

So, once we've determined that partial fractions can be done we factor the denominator as completely as possible. Then for each factor in the denominator we can use the following table to determine the term(s) we pick up in the partial fraction decomposition.

Factor in denominator	Term in partial fraction decomposition
$ax+b$	$\dfrac{A}{ax+b}$
$(ax+b)^k$	$\dfrac{A_1}{ax+b}+\dfrac{A_2}{(ax+b)^2}+\cdots+\dfrac{A_k}{(ax+b)^k},\ k=1,2,3,\ldots$
ax^2+bx+c	$\dfrac{Ax+B}{ax^2+bx+c}$
$(ax^2+bx+c)^k$	$\dfrac{A_1x+B_1}{ax^2+bx+c}+\dfrac{A_2x+B_2}{(ax^2+bx+c)^2}+\cdots+\dfrac{A_kx+B_k}{(ax^2+bx+c)^k},\ k=1,2,3,\ldots$

Notice that the first and third cases are really special cases of the second and fourth cases respectively.

There are several methods for determining the coefficients for each term and we will go over each of those in the following examples.

Let's start the examples by doing the integral above.

Example 1

Evaluate the following integral.

$$\int \frac{3x+11}{x^2-x-6}\,dx$$

Solution

The first step is to factor the denominator as much as possible and get the form of the partial fraction decomposition. Doing this gives,

$$\frac{3x+11}{(x-3)(x+2)}=\frac{A}{x-3}+\frac{B}{x+2}$$

The next step is to actually add the right side back up.

$$\frac{3x+11}{(x-3)(x+2)}=\frac{A(x+2)+B(x-3)}{(x-3)(x+2)}$$

Now, we need to choose A and B so that the numerators of these two are equal for every x. To do this we'll need to set the numerators equal.

$$3x+11=A(x+2)+B(x-3)$$

Note that in most problems we will go straight from the general form of the decomposition to this step and not bother with actually adding the terms back up. The only point to adding the terms is to get the numerator and we can get that without actually writing down the results of the addition.

At this point we have one of two ways to proceed. One way will always work but is often more work. The other, while it won't always work, is often quicker when it does work. In this case both will work and so we'll use the quicker way for this example. We'll take a look at the other method in a later example.

What we're going to do here is to notice that the numerators must be equal for *any x* that we would choose to use. In particular the numerators must be equal for $x = -2$ and $x = 3$. So, let's plug these in and see what we get.

$$\begin{array}{llll} x = -2 : & 5 = A(0) + B(-5) & \Rightarrow & B = -1 \\ x = 3 \quad : & 20 = A(5) + B(0) & \Rightarrow & A = 4 \end{array}$$

So, by carefully picking the x's we got the unknown constants to quickly drop out. Note that these are the values we claimed they would be above.

At this point there really isn't a whole lot to do other than the integral.

$$\begin{aligned} \int \frac{3x+11}{x^2-x-6}\,dx &= \int \frac{4}{x-3} - \frac{1}{x+2}dx \\ &= \int \frac{4}{x-3}\,dx - \int \frac{1}{x+2}dx \\ &= 4\ln|x-3| - \ln|x+2| + c \end{aligned}$$

Recall that to do this integral we first split it up into two integrals and then used the substitutions,

$$u = x - 3 \qquad\qquad v = x + 2$$

on the integrals to get the final answer.

Before moving onto the next example a couple of quick notes are in order here. First, many of the integrals in partial fractions problems come down to the type of integral seen above. Make sure that you can do those integrals.

There is also another integral that often shows up in these kinds of problems so we may as well give the formula for it here since we are already on the subject.

$$\int \frac{1}{x^2+a^2}\,dx = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right) + c$$

It will be an example or two before we use this so don't forget about it.

Now, let's work some more examples.

Example 2

Evaluate the following integral.

$$\int \frac{x^2+4}{3x^3+4x^2-4x}\,dx$$

Solution

We won't be putting as much detail into this solution as we did in the previous example. The first thing is to factor the denominator and get the form of the partial fraction decomposition.

$$\frac{x^2+4}{x\,(x+2)\,(3x-2)} = \frac{A}{x} + \frac{B}{x+2} + \frac{C}{3x-2}$$

The next step is to set numerators equal. If you need to actually add the right side together to get the numerator for that side then you should do so, however, it will definitely make the problem quicker if you can do the addition in your head to get,

$$x^2+4 = A\,(x+2)\,(3x-2) + Bx\,(3x-2) + Cx\,(x+2)$$

As with the previous example it looks like we can just pick a few values of x and find the constants so let's do that.

$$\begin{array}{llll} x=0 \quad : & 4 = A\,(2)\,(-2) & \Rightarrow & A=-1 \\ x=-2 : & 8 = B\,(-2)\,(-8) & \Rightarrow & B = \dfrac{1}{2} \\ x=\dfrac{2}{3} \; : & \dfrac{40}{9} = C\left(\dfrac{2}{3}\right)\left(\dfrac{8}{3}\right) & \Rightarrow & C = \dfrac{40}{16} = \dfrac{5}{2} \end{array}$$

Note that unlike the first example most of the coefficients here are fractions. That is not unusual so don't get excited about it when it happens.

Now, let's do the integral.

$$\begin{aligned} \int \frac{x^2+4}{3x^3+4x^2-4x}\,dx &= \int -\frac{1}{x} + \frac{\frac{1}{2}}{x+2} + \frac{\frac{5}{2}}{3x-2}\,dx \\ &= -\ln|x| + \frac{1}{2}\ln|x+2| + \frac{5}{6}\ln|3x-2| + c \end{aligned}$$

Again, as noted above, integrals that generate natural logarithms are very common in these problems so make sure you can do them. Also, you were able to correctly do the last integral right? The coefficient of $\frac{5}{6}$ is correct. Make sure that you do the substitution required for the term properly.

Example 3

Evaluate the following integral.

$$\int \frac{x^2 - 29x + 5}{(x-4)^2 (x^2+3)} \, dx$$

Solution

This time the denominator is already factored so let's just jump right to the partial fraction decomposition.

$$\frac{x^2 - 29x + 5}{(x-4)^2 (x^2+3)} = \frac{A}{x-4} + \frac{B}{(x-4)^2} + \frac{Cx+D}{x^2+3}$$

Setting numerators gives,

$$x^2 - 29x + 5 = A(x-4)\left(x^2+3\right) + B\left(x^2+3\right) + (Cx+D)(x-4)^2$$

In this case we aren't going to be able to just pick values of x that will give us all the constants. Therefore, we will need to work this the second (and often longer) way. The first step is to multiply out the right side and collect all the like terms together. Doing this gives,

$$x^2 - 29x + 5 = (A+C)\,x^3 + (-4A+B-8C+D)\,x^2 + (3A+16C-8D)\,x - 12A + 3B + 16D$$

Now we need to choose A, B, C, and D so that these two are equal. In other words, we will need to set the coefficients of like powers of x equal. This will give a system of equations that can be solved.

$$\left.\begin{array}{lr} x^3: & A + C = 0 \\ x^2: & -4A + B - 8C + D = 1 \\ x^1: & 3A + 16C - 8D = -29 \\ x^0: & -12A + 3B + 16D = 5 \end{array}\right\} \quad \Rightarrow \quad A = 1,\ B = -5,\ C = -1,\ D = 2$$

Note that we used x^0 to represent the constants. Also note that these systems can often be quite large and have a fair amount of work involved in solving them. The best way to deal with these is to use some form of computer aided solving techniques.

Now, let's take a look at the integral.

$$\begin{aligned} \int \frac{x^2 - 29x + 5}{(x-4)^2 (x^2+3)} \, dx &= \int \frac{1}{x-4} - \frac{5}{(x-4)^2} + \frac{-x+2}{x^2+3} \, dx \\ &= \int \frac{1}{x-4} - \frac{5}{(x-4)^2} - \frac{x}{x^2+3} + \frac{2}{x^2+3} \, dx \\ &= \ln|x-4| + \frac{5}{x-4} - \frac{1}{2}\ln\left|x^2+3\right| + \frac{2}{\sqrt{3}}\tan^{-1}\left(\frac{x}{\sqrt{3}}\right) + c \end{aligned}$$

In order to take care of the third term we needed to split it up into two separate terms. Once we've done this we can do all the integrals in the problem. The first two use the substitution $u = x - 4$, the third uses the substitution $v = x^2 + 3$ and the fourth term uses the formula given above for inverse tangents.

Example 4

Evaluate the following integral.

$$\int \frac{x^3 + 10x^2 + 3x + 36}{(x-1)\left(x^2+4\right)^2}\,dx$$

Solution

Let's first get the general form of the partial fraction decomposition.

$$\frac{x^3 + 10x^2 + 3x + 36}{(x-1)\left(x^2+4\right)^2} = \frac{A}{x-1} + \frac{Bx+C}{x^2+4} + \frac{Dx+E}{\left(x^2+4\right)^2}$$

Now, set numerators equal, expand the right side and collect like terms.

$$\begin{aligned} x^3 + 10x^2 + 3x + 36 &= A\left(x^2+4\right)^2 + (Bx+C)(x-1)\left(x^2+4\right) + (Dx+E)(x-1) \\ &= (A+B)\,x^4 + (C-B)\,x^3 + (8A+4B-C+D)\,x^2 + \\ &\qquad (-4B+4C-D+E)\,x + 16A - 4C - E \end{aligned}$$

Setting coefficient equal gives the following system.

$$\left.\begin{array}{ll} x^4: & A+B=0 \\ x^3: & C-B=1 \\ x^2: & 8A+4B-C+D=10 \\ x^1: & -4B+4C-D+E=3 \\ x^0: & 16A-4C-E=36 \end{array}\right\} \quad \Rightarrow \quad A=2,\ B=-2,\ C=-1,\ D=1,\ E=0$$

Don't get excited if some of the coefficients end up being zero. It happens on occasion.

Here's the integral.

$$\begin{aligned} \int \frac{x^3 + 10x^2 + 3x + 36}{(x-1)\left(x^2+4\right)^2}\,dx &= \int \frac{2}{x-1} + \frac{-2x-1}{x^2+4} + \frac{x}{\left(x^2+4\right)^2}\,dx \\ &= \int \frac{2}{x-1} - \frac{2x}{x^2+4} - \frac{1}{x^2+4} + \frac{x}{\left(x^2+4\right)^2}\,dx \\ &= 2\ln|x-1| - \ln\left|x^2+4\right| - \frac{1}{2}\tan^{-1}\left(\frac{x}{2}\right) - \frac{1}{2}\frac{1}{x^2+4} + c \end{aligned}$$

To this point we've only looked at rational expressions where the degree of the numerator was strictly less that the degree of the denominator. Of course, not all rational expressions will fit into this form and so we need to take a look at a couple of examples where this isn't the case.

Example 5

Evaluate the following integral.

$$\int \frac{x^4 - 5x^3 + 6x^2 - 18}{x^3 - 3x^2}\,dx$$

Solution

So, in this case the degree of the numerator is 4 and the degree of the denominator is 3. Therefore, partial fractions can't be done on this rational expression.

To fix this up we'll need to do long division on this to get it into a form that we can deal with. Here is the work for that.

$$\begin{array}{r} x - 2 \\ x^3 - 3x^2 \overline{\big) x^4 - 5x^3 + 6x^2 - 18} \\ -\left(x^4 - 3x^3\right) \\ \hline -2x^3 + 6x^2 - 18 \\ -\left(-2x^3 + 6x^2\right) \\ \hline -18 \end{array}$$

So, from the long division we see that,

$$\frac{x^4 - 5x^3 + 6x^2 - 18}{x^3 - 3x^2} = x - 2 - \frac{18}{x^3 - 3x^2}$$

and the integral becomes,

$$\begin{aligned} \int \frac{x^4 - 5x^3 + 6x^2 - 18}{x^3 - 3x^2}\,dx &= \int x - 2 - \frac{18}{x^3 - 3x^2}\,dx \\ &= \int x - 2\,dx - \int \frac{18}{x^3 - 3x^2}\,dx \end{aligned}$$

The first integral we can do easily enough and the second integral is now in a form that allows us to do partial fractions. So, let's get the general form of the partial fractions for the second integrand.

$$\frac{18}{x^2\left(x-3\right)} = \frac{A}{x} + \frac{B}{x^2} + \frac{C}{x-3}$$

Setting numerators equal gives us,

$$18 = Ax\left(x-3\right) + B\left(x-3\right) + Cx^2$$

Now, there is a variation of the method we used in the first couple of examples that will work here. There are a couple of values of x that will allow us to quickly get two of the three constants, but there is no value of x that will just hand us the third.

What we'll do in this example is pick x's to get the two constants that we can easily get and then we'll just pick another value of x that will be easy to work with (*i.e.* it won't give large/messy numbers anywhere) and then we'll use the fact that we also know the other two constants to find the third.

$$\begin{aligned} &x = 0: && 18 = B(-3) && \Rightarrow && B = -6 \\ &x = 3: && 18 = C(9) && \Rightarrow && C = 2 \\ &x = 1: && 18 = A(-2) + B(-2) + C = -2A + 14 && \Rightarrow && A = -2 \end{aligned}$$

The integral is then,

$$\begin{aligned} \int \frac{x^4 - 5x^3 + 6x^2 - 18}{x^3 - 3x^2}\,dx &= \int x - 2\,dx - \int -\frac{2}{x} - \frac{6}{x^2} + \frac{2}{x-3}\,dx \\ &= \frac{1}{2}x^2 - 2x + 2\ln|x| - \frac{6}{x} - 2\ln|x-3| + c \end{aligned}$$

In the previous example there were actually two different ways of dealing with the x^2 in the denominator. One is to treat it as a quadratic which would give the following term in the decomposition

$$\frac{Ax + B}{x^2}$$

and the other is to treat it as a linear term in the following way,

$$x^2 = (x - 0)^2$$

which gives the following two terms in the decomposition,

$$\frac{A}{x} + \frac{B}{x^2}$$

We used the second way of thinking about it in our example. Notice however that the two will give identical partial fraction decompositions. So, why talk about this? Simple. This will work for x^2, but what about x^3 or x^4? In these cases, we really will need to use the second way of thinking about these kinds of terms.

$$x^3 \quad \Rightarrow \quad \frac{A}{x} + \frac{B}{x^2} + \frac{C}{x^3} \qquad\qquad x^4 \quad \Rightarrow \quad \frac{A}{x} + \frac{B}{x^2} + \frac{C}{x^3} + \frac{D}{x^4}$$

Let's take a look at one more example.

Example 6

Evaluate the following integral.

$$\int \frac{x^2}{x^2 - 1}\,dx$$

Solution

In this case the numerator and denominator have the same degree. As with the last example we'll need to do long division to get this into the correct form. We'll leave the details of that to you to check.

$$\int \frac{x^2}{x^2 - 1}\,dx = \int 1 + \frac{1}{x^2 - 1}\,dx = \int dx + \int \frac{1}{x^2 - 1}\,dx$$

So, we'll need to partial fraction the second integral. Here's the decomposition.

$$\frac{1}{(x-1)(x+1)} = \frac{A}{x-1} + \frac{B}{x+1}$$

Setting numerator equal gives,

$$1 = A(x+1) + B(x-1)$$

Picking value of x gives us the following coefficients.

$$\begin{array}{llll} x = -1 : & 1 = B(-2) & \Rightarrow & B = -\frac{1}{2} \\ x = 1 \;\;\; : & 1 = A(2) & \Rightarrow & A = \frac{1}{2} \end{array}$$

The integral is then,

$$\begin{aligned} \int \frac{x^2}{x^2 - 1}\,dx &= \int dx + \int \frac{\frac{1}{2}}{x-1} - \frac{\frac{1}{2}}{x+1}\,dx \\ &= x + \frac{1}{2}\ln|x-1| - \frac{1}{2}\ln|x+1| + c \end{aligned}$$

7.5 Integrals Involving Roots

In this section we're going to look at an integration technique that can be useful for *some* integrals with roots in them. We've already seen some integrals with roots in them. Some can be done quickly with a simple Calculus I substitution and some can be done with trig substitutions.

However, not all integrals with roots will allow us to use one of these methods. Let's look at a couple of examples to see another technique that can be used on occasion to help with these integrals.

Example 1

Evaluate the following integral.

$$\int \frac{x+2}{\sqrt[3]{x-3}}\,dx$$

Solution

Sometimes when faced with an integral that contains a root we can use the following substitution to simplify the integral into a form that can be easily worked with.

$$u = \sqrt[3]{x-3}$$

So, instead of letting u be the stuff under the radical as we often did in Calculus I we let u be the whole radical. Now, there will be a little more work here since we will also need to know what x is so we can substitute in for that in the numerator and so we can compute the differential, dx. This is easy enough to get however. Just solve the substitution for x as follows,

$$x = u^3 + 3 \qquad\qquad dx = 3u^2\,du$$

Using this substitution the integral is now,

$$\begin{aligned}\int \frac{(u^3+3)+2}{u}\,3u^2du &= \int 3u^4 + 15u\,du\\ &= \frac{3}{5}u^5 + \frac{15}{2}u^2 + c\\ &= \frac{3}{5}(x-3)^{\frac{5}{3}} + \frac{15}{2}(x-3)^{\frac{2}{3}} + c\end{aligned}$$

So, sometimes, when an integral contains the root $\sqrt[n]{g(x)}$ the substitution,

$$u = \sqrt[n]{g(x)}$$

can be used to simplify the integral into a form that we can deal with.

Let's take a look at another example real quick.

Example 2

Evaluate the following integral.

$$\int \frac{2}{x - 3\sqrt{x+10}}\,dx$$

Solution

We'll do the same thing we did in the previous example. Here's the substitution and the extra work we'll need to do to get x in terms of u.

$$u = \sqrt{x+10} \qquad x = u^2 - 10 \qquad dx = 2u\,du$$

With this substitution the integral is,

$$\int \frac{2}{x - 3\sqrt{x+10}}\,dx = \int \frac{2}{u^2 - 10 - 3u}\left(2u\right)\,du = \int \frac{4u}{u^2 - 3u - 10}\,du$$

This integral can now be done with partial fractions.

$$\frac{4u}{\left(u-5\right)\left(u+2\right)} = \frac{A}{u-5} + \frac{B}{u+2}$$

Setting numerators equal gives,

$$4u = A\left(u+2\right) + B\left(u-5\right)$$

Picking value of u gives the coefficients.

$$\begin{array}{llll} u = -2 & -8 = B\left(-7\right) & \Rightarrow & B = \frac{8}{7} \\ u = 5 & 20 = A\left(7\right) & \Rightarrow & A = \frac{20}{7} \end{array}$$

The integral is then,

$$\begin{aligned} \int \frac{2}{x - 3\sqrt{x+10}}\,dx &= \int \frac{\frac{20}{7}}{u-5} + \frac{\frac{8}{7}}{u+2}\,du \\ &= \frac{20}{7}\ln\left|u-5\right| + \frac{8}{7}\ln\left|u+2\right| + c \\ &= \frac{20}{7}\ln\left|\sqrt{x+10}-5\right| + \frac{8}{7}\ln\left|\sqrt{x+10}+2\right| + c \end{aligned}$$

So, we've seen a nice method to eliminate roots from the integral and put it into a form that we can deal with. Note however, that this won't always work and sometimes the new integral will be just as difficult to do.

7.6 Integrals Involving Quadratics

To this point we've seen quite a few integrals that involve quadratics. A couple of examples are,

$$\int \frac{x}{x^2 \pm a}\,dx = \frac{1}{2}\ln\left|x^2 \pm a\right| + c \qquad\qquad \int \frac{1}{x^2 + a^2}\,dx = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right) + c$$

We also saw that integrals involving $\sqrt{b^2x^2 - a^2}$, $\sqrt{a^2 - b^2x^2}$ and $\sqrt{a^2 + b^2x^2}$ could be done with a trig substitution.

Notice however that all of these integrals were missing an x term. They all consist of only a quadratic term and a constant.

Some integrals involving general quadratics are easy enough to do. For instance, the following integral can be done with a quick substitution.

$$\begin{aligned}\int \frac{2x+3}{4x^2+12x-1}\,dx &= \frac{1}{4}\int \frac{1}{u}\,du \qquad\qquad u = 4x^2+12x-1 \qquad du = 4\left(2x+3\right)dx \\ &= \frac{1}{4}\ln\left|4x^2+12x-1\right| + c\end{aligned}$$

Some integrals with quadratics can be done with partial fractions. For instance,

$$\int \frac{10x-6}{3x^2+16x+5}\,dx = \int \frac{4}{x+5} - \frac{2}{3x+1}\,dx = 4\ln|x+5| - \frac{2}{3}\ln|3x+1| + c$$

Unfortunately, these methods won't work on a lot of integrals. A simple substitution will only work if the numerator is a constant multiple of the derivative of the denominator and partial fractions will only work if the denominator can be factored.

The topic of this section is how to deal with integrals involving quadratics when the techniques that we've looked at to this point simply won't work.

Back in the Trig Substitution section we saw how to deal with square roots that had a general quadratic in them. Let's take a quick look at another one like that since the idea involved in doing that kind of integral is exactly what we are going to need for the other integrals in this section.

Example 1

Evaluate the following integral.

$$\int \sqrt{x^2+4x+5}\,dx$$

Solution

Recall from the Trig Substitution section that in order to do a trig substitution here we first needed to complete the square on the quadratic. This gives,

$$x^2 + 4x + 5 = x^2 + 4x + 4 - 4 + 5 = (x+2)^2 + 1$$

After completing the square the integral becomes,

$$\int \sqrt{x^2 + 4x + 5}\, dx = \int \sqrt{(x+2)^2 + 1}\, dx$$

Upon doing this we can identify the trig substitution that we need. Here it is,

$$x + 2 = \tan(\theta) \qquad x = \tan(\theta) - 2 \qquad dx = \sec^2(\theta)\, d\theta$$

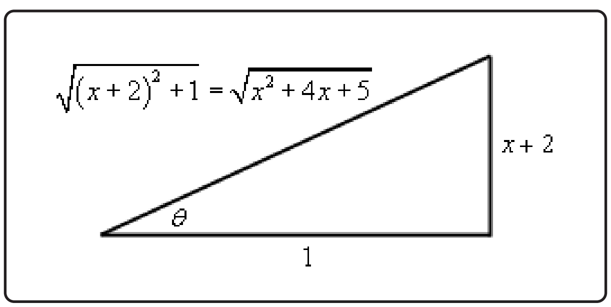
$$\sqrt{(x+2)^2 + 1} = \sqrt{\tan^2(\theta) + 1} = \sqrt{\sec^2(\theta)} = |\sec(\theta)| = \sec(\theta)$$

Recall that since we are doing an indefinite integral we can drop the absolute value bars. Using this substitution the integral becomes,

$$\begin{aligned} \int \sqrt{x^2 + 4x + 5}\, dx &= \int \sec^3(\theta)\, d\theta \\ &= \frac{1}{2}\Big(\sec(\theta)\tan(\theta) + \ln|\sec(\theta) + \tan(\theta)|\Big) + c \end{aligned}$$

We can finish the integral out with the following right triangle.

$$\tan(\theta) = \frac{x+2}{1} \qquad \sec(\theta) = \frac{\sqrt{x^2 + 4x + 5}}{1} = \sqrt{x^2 + 4x + 5}$$



$$\int \sqrt{x^2 + 4x + 5}\, dx = \frac{1}{2}\left((x+2)\sqrt{x^2 + 4x + 5} + \ln\left|x + 2 + \sqrt{x^2 + 4x + 5}\right|\right) + c$$

So, by completing the square we were able to take an integral that had a general quadratic in it and convert it into a form that allowed us to use a known integration technique.

Let's do a quick review of completing the square before proceeding. Here is the general completing the square formula that we'll use.

$$x^2 + bx + c = x^2 + bx + \left(\frac{b}{2}\right)^2 - \left(\frac{b}{2}\right)^2 + c = \left(x + \frac{b}{2}\right)^2 + c - \frac{b^2}{4}$$

This will always take a general quadratic and write it in terms of a squared term and a constant term.

Recall as well that in order to do this we must have a coefficient of one in front of the x^2. If not, we'll need to factor out the coefficient before completing the square. In other words,

$$ax^2 + bx + c = a\underbrace{\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right)}_{\text{complete the square on this!}}$$

Now, let's see how completing the square can be used to do integrals that we aren't able to do at this point.

Example 2

Evaluate the following integral.

$$\int \frac{1}{2x^2 - 3x + 2}\,dx$$

Solution

Okay, this doesn't factor so partial fractions just won't work on this. Likewise, since the numerator is just "1" we can't use the substitution $u = 2x^2 - 3x + 2$. So, let's see what happens if we complete the square on the denominator.

$$\begin{aligned} 2x^2 - 3x + 2 &= 2\left(x^2 - \frac{3}{2}x + 1\right) \\ &= 2\left(x^2 - \frac{3}{2}x + \frac{9}{16} - \frac{9}{16} + 1\right) \\ &= 2\left(\left(x - \frac{3}{4}\right)^2 + \frac{7}{16}\right) \end{aligned}$$

With this the integral is,

$$\int \frac{1}{2x^2 - 3x + 2}\,dx = \frac{1}{2}\int \frac{1}{\left(x - \frac{3}{4}\right)^2 + \frac{7}{16}}\,dx$$

Now this may not seem like all that great of a change. However, notice that we can now use the following substitution.

$$u = x - \frac{3}{4} \qquad du = dx$$

and the integral is now,

$$\int \frac{1}{2x^2 - 3x + 2}\,dx = \frac{1}{2}\int \frac{1}{u^2 + \frac{7}{16}}\,du$$

We can now see that this is an inverse tangent! So, using the formula from the start of the section we get,

$$\begin{aligned}\int \frac{1}{2x^2 - 3x + 2}\,dx &= \frac{1}{2}\left(\frac{4}{\sqrt{7}}\right)\tan^{-1}\left(\frac{4u}{\sqrt{7}}\right) + c\\ &= \frac{2}{\sqrt{7}}\tan^{-1}\left(\frac{4x-3}{\sqrt{7}}\right) + c\end{aligned}$$

Example 3

Evaluate the following integral.

$$\int \frac{3x - 1}{x^2 + 10x + 28}\,dx$$

Solution

This example is a little different from the previous one. In this case we do have an x in the numerator however the numerator still isn't a multiple of the derivative of the denominator and so a simple Calculus I substitution won't work.

So, let's again complete the square on the denominator and see what we get,

$$x^2 + 10x + 28 = x^2 + 10x + 25 - 25 + 28 = (x+5)^2 + 3$$

Upon completing the square the integral becomes,

$$\int \frac{3x - 1}{x^2 + 10x + 28}\,dx = \int \frac{3x - 1}{(x+5)^2 + 3}\,dx$$

At this point we can use the same type of substitution that we did in the previous example. The only real difference is that we'll need to make sure that we plug the substitution back into the numerator as well.

$$u = x + 5 \qquad x = u - 5 \qquad dx = du$$

$$\begin{aligned}
\int \frac{3x-1}{x^2+10x+28}\,dx &= \int \frac{3\,(u-5)-1}{u^2+3}\,du \\
&= \int \frac{3u}{u^2+3} - \frac{16}{u^2+3}\,du \\
&= \frac{3}{2}\ln\left|u^2+3\right| - \frac{16}{\sqrt{3}}\tan^{-1}\left(\frac{u}{\sqrt{3}}\right)+c \\
&= \frac{3}{2}\ln\left|(x+5)^2+3\right| - \frac{16}{\sqrt{3}}\tan^{-1}\left(\frac{x+5}{\sqrt{3}}\right)+c
\end{aligned}$$

So, in general when dealing with an integral in the form,

$$\int \frac{Ax+B}{ax^2+bx+c}\,dx \tag{7.5}$$

Here we are going to assume that the denominator doesn't factor and the numerator isn't a constant multiple of the derivative of the denominator. In these cases, we complete the square on the denominator and then do a substitution that will yield an inverse tangent and/or a logarithm depending on the exact form of the numerator.

Let's now take a look at a couple of integrals that are in the same general form as Equation 7.5 except the denominator will also be raised to a power. In other words, let's look at integrals in the form,

$$\int \frac{Ax+B}{(ax^2+bx+c)^n}\,dx \tag{7.6}$$

Example 4

Evaluate the following integral.

$$\int \frac{x}{(x^2-6x+11)^3}\,dx$$

Solution

For the most part this integral will work the same as the previous two with one exception that will occur down the road. So, let's start by completing the square on the quadratic in the denominator.

$$x^2-6x+11 = x^2-6x+9-9+11 = (x-3)^2+2$$

The integral is then,

$$\int \frac{x}{\left(x^2-6x+11\right)^3}\,dx = \int \frac{x}{\left[(x-3)^2+2\right]^3}\,dx$$

Now, we will use the same substitution that we've used to this point in the previous two examples.

$$u = x-3 \qquad x = u+3 \qquad dx = du$$

$$\begin{aligned}\int \frac{x}{\left(x^2-6x+11\right)^3}\,dx &= \int \frac{u+3}{\left(u^2+2\right)^3}\,du \\ &= \int \frac{u}{\left(u^2+2\right)^3}\,du + \int \frac{3}{\left(u^2+2\right)^3}\,du\end{aligned}$$

Now, here is where the differences start cropping up. The first integral can be done with the substitution $v = u^2+2$ and isn't too difficult. The second integral however, can't be done with the substitution used on the first integral and it isn't an inverse tangent.

It turns out that a trig substitution will work nicely on the second integral and it will be the same as we did when we had square roots in the problem.

$$u = \sqrt{2}\tan(\theta) \qquad du = \sqrt{2}\sec^2(\theta)\,d\theta$$

With these two substitutions the integrals become,

$$\begin{aligned}\int \frac{x}{\left(x^2-6x+11\right)^3}\,dx &= \frac{1}{2}\int \frac{1}{v^3}\,dv + \int \frac{3}{\left(2\tan^2(\theta)+2\right)^3}\left(\sqrt{2}\sec^2(\theta)\right)d\theta \\ &= -\frac{1}{4}\frac{1}{v^2} + \int \frac{3\sqrt{2}\sec^2(\theta)}{8\left(\tan^2(\theta)+1\right)^3}\,d\theta \\ &= -\frac{1}{4}\frac{1}{\left(u^2+2\right)^2} + \frac{3\sqrt{2}}{8}\int \frac{\sec^2(\theta)}{\left(\sec^2(\theta)\right)^3}\,d\theta \\ &= -\frac{1}{4}\frac{1}{\left((x-3)^2+2\right)^2} + \frac{3\sqrt{2}}{8}\int \frac{1}{\sec^4(\theta)}\,d\theta \\ &= -\frac{1}{4}\frac{1}{\left((x-3)^2+2\right)^2} + \frac{3\sqrt{2}}{8}\int \cos^4(\theta)\,d\theta\end{aligned}$$

Okay, at this point we've got two options for the remaining integral. We can either use the ideas we learned in the section about integrals involving trig integrals or we could use the following formula.

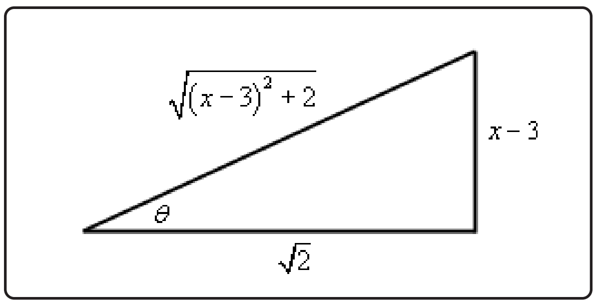
$$\int \cos^m(\theta)\,d\theta = \frac{1}{m}\sin(\theta)\cos^{m-1}(\theta) + \frac{m-1}{m}\int \cos^{m-2}(\theta)\,d\theta$$

Let's use this formula to do the integral.

$$\begin{aligned}\int \cos^4(\theta)\,d\theta &= \frac{1}{4}\sin(\theta)\cos^3(\theta) + \frac{3}{4}\int \cos^2(\theta)\,d\theta \\ &= \frac{1}{4}\sin(\theta)\cos^3(\theta) + \frac{3}{4}\left(\frac{1}{2}\sin(\theta)\cos(\theta) + \frac{1}{2}\int \cos^0(\theta)\,d\theta\right) \qquad \cos^0(\theta) = 1! \\ &= \frac{1}{4}\sin(\theta)\cos^3(\theta) + \frac{3}{8}\sin(\theta)\cos(\theta) + \frac{3}{8}\theta\end{aligned}$$

Next, let's use the following right triangle to get this back to x's.

$$\tan(\theta) = \frac{u}{\sqrt{2}} = \frac{x-3}{\sqrt{2}} \qquad \sin(\theta) = \frac{x-3}{\sqrt{(x-3)^2+2}} \qquad \cos(\theta) = \frac{\sqrt{2}}{\sqrt{(x-3)^2+2}}$$



The cosine integral is then,

$$\begin{aligned}\int \cos^4(\theta)\,d\theta &= \frac{1}{4}\frac{2\sqrt{2}\,(x-3)}{\left((x-3)^2+2\right)^2} + \frac{3}{8}\frac{\sqrt{2}\,(x-3)}{(x-3)^2+2} + \frac{3}{8}\tan^{-1}\left(\frac{x-3}{\sqrt{2}}\right) \\ &= \frac{\sqrt{2}}{2}\frac{x-3}{\left((x-3)^2+2\right)^2} + \frac{3\sqrt{2}}{8}\frac{x-3}{(x-3)^2+2} + \frac{3}{8}\tan^{-1}\left(\frac{x-3}{\sqrt{2}}\right)\end{aligned}$$

All told then the original integral is,

$$\begin{aligned}\int \frac{x}{(x^2-6x+11)^3}\,dx = -\frac{1}{4}\frac{1}{\left((x-3)^2+2\right)^2} + \\ \frac{3\sqrt{2}}{8}\left(\frac{\sqrt{2}}{2}\frac{x-3}{\left((x-3)^2+2\right)^2} + \frac{3\sqrt{2}}{8}\frac{x-3}{(x-3)^2+2} + \frac{3}{8}\tan^{-1}\left(\frac{x-3}{\sqrt{2}}\right)\right) \\ = \frac{1}{8}\frac{3x-11}{\left((x-3)^2+2\right)^2} + \frac{9}{32}\frac{x-3}{(x-3)^2+2} + \frac{9\sqrt{2}}{64}\tan^{-1}\left(\frac{x-3}{\sqrt{2}}\right) + c\end{aligned}$$

It's a long and messy answer, but there it is.

Example 5

Evaluate the following integral.

$$\int \frac{x-3}{\left(4-2x-x^2\right)^2}\,dx$$

Solution

As with the other problems we'll first complete the square on the denominator.

$$4-2x-x^2 = -\left(x^2+2x-4\right) = -\left(x^2+2x+1-1-4\right) = -\left((x+1)^2-5\right) = 5-(x+1)^2$$

The integral is,

$$\int \frac{x-3}{\left(4-2x-x^2\right)^2}\,dx = \int \frac{x-3}{\left[5-(x+1)^2\right]^2}\,dx$$

Now, let's do the substitution.

$$u = x+1 \qquad x = u-1 \qquad dx = du$$

and the integral is now,

$$\begin{aligned}\int \frac{x-3}{\left(4-2x-x^2\right)^2}\,dx &= \int \frac{u-4}{\left(5-u^2\right)^2}\,du \\ &= \int \frac{u}{\left(5-u^2\right)^2}\,du - \int \frac{4}{\left(5-u^2\right)^2}\,du\end{aligned}$$

In the first integral we'll use the substitution

$$v = 5-u^2$$

and in the second integral we'll use the following trig substitution

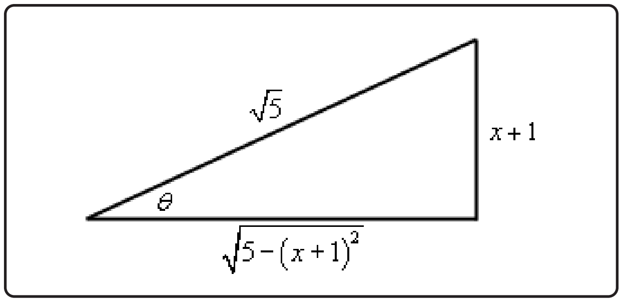
$$u = \sqrt{5}\sin(\theta) \qquad du = \sqrt{5}\cos(\theta)\,d\theta$$

Using these substitutions the integral becomes,

$$\begin{aligned}\int \frac{x-3}{\left(4-2x-x^2\right)^2}\,dx &= -\frac{1}{2}\int \frac{1}{v^2}\,dv - \int \frac{4}{\left(5-5\sin^2(\theta)\right)^2}\left(\sqrt{5}\cos(\theta)\right)\,d\theta \\ &= \frac{1}{2}\frac{1}{v} - \frac{4\sqrt{5}}{25}\int \frac{\cos(\theta)}{\left(1-\sin^2(\theta)\right)^2}\,d\theta \\ &= \frac{1}{2}\frac{1}{v} - \frac{4\sqrt{5}}{25}\int \frac{\cos(\theta)}{\cos^4(\theta)}\,d\theta \\ &= \frac{1}{2}\frac{1}{v} - \frac{4\sqrt{5}}{25}\int \sec^3(\theta)\,d\theta \\ &= \frac{1}{2}\frac{1}{v} - \frac{2\sqrt{5}}{25}\left(\sec(\theta)\tan(\theta) + \ln\left|\sec(\theta)+\tan(\theta)\right|\right) + c\end{aligned}$$

We'll need the following right triangle to finish this integral out.

$$\sin(\theta) = \frac{u}{\sqrt{5}} = \frac{x+1}{\sqrt{5}} \qquad \sec(\theta) = \frac{\sqrt{5}}{\sqrt{5-(x+1)^2}} \qquad \tan(\theta) = \frac{x+1}{\sqrt{5-(x+1)^2}}$$



So, going back to x's the integral becomes,

$$\begin{aligned}\int \frac{x-3}{\left(4-2x-x^2\right)^2}\,dx &= \frac{1}{2}\frac{1}{5-u^2} - \\ &\qquad \frac{2\sqrt{5}}{25}\left(\frac{\sqrt{5}\,(x+1)}{5-(x+1)^2} + \ln\left|\frac{\sqrt{5}}{\sqrt{5-(x+1)^2}} + \frac{x+1}{\sqrt{5-(x+1)^2}}\right|\right) + c \\ &= \frac{1}{10}\frac{1-4x}{5-(x+1)^2} - \frac{2\sqrt{5}}{25}\ln\left|\frac{x+1+\sqrt{5}}{\sqrt{5-(x+1)^2}}\right| + c\end{aligned}$$

Often the following formula is needed when using the trig substitution that we used in the previous example.

$$\int \sec^m(\theta)\,d\theta = \frac{1}{m-1}\tan(\theta)\sec^{m-2}(\theta) + \frac{m-2}{m-1}\int \sec^{m-2}(\theta)\,d\theta$$

Note that we'll only need the two trig substitutions (sine and tangent) that we used here. The third trig substitution (secant) that we used will not be needed here. Any quadratic that could use a secant substitution can be turned into a sine substitution simply by factoring a minus sign out of the quadratic. Note that we can do that for these types of problems because we don't have a root and so the minus sign can be completely factored out of the integrand while we couldn't do that with the roots we had in the problems back in the Trig Substitution section.

7.7 Integration Strategy

We've now seen a fair number of different integration techniques and so we should probably pause at this point and talk a little bit about a strategy to use for determining the correct technique to use when faced with an integral.

There are a couple of points that need to be made about this strategy. First, it isn't a hard and fast set of rules for determining the method that should be used. It is really nothing more than a general set of guidelines that will help us to identify techniques that may work. Some integrals can be done in more than one way and so depending on the path you take through the strategy you may end up with a different technique than somebody else who also went through this strategy.

Second, while the strategy is presented as a way to identify the technique that could be used on an integral also keep in mind that, for many integrals, it can also automatically exclude certain techniques as well. When going through the strategy keep two lists in mind. The first list is integration techniques that simply won't work and the second list is techniques that look like they might work. After going through the strategy and the second list has only one entry then that is the technique to use. If, on the other hand, there is more than one possible technique to use we will then have to decide on which is liable to be the best for us to use. Unfortunately, there is no way to teach which technique is the best as that usually depends upon the person and which technique they find to be the easiest.

Third, don't forget that many integrals can be evaluated in multiple ways and so more than one technique may be used on it. This has already been mentioned in each of the previous points but is important enough to warrant a separate mention. Sometimes one technique will be significantly easier than the others and so don't just stop at the first technique that appears to work. Always identify all possible techniques and then go back and determine which you feel will be the easiest for you to use.

Next, it's entirely possible that you will need to use more than one method to completely do an integral. For instance, a substitution may lead to using integration by parts or partial fractions integral.

Finally, in my class I will accept any valid integration technique as a solution. As already noted there is often more than one way to do an integral and just because I find one technique to be the easiest doesn't mean that you will as well. So, in my class, there is no one right way of doing an integral. You may use any integration technique that I've taught you in this class or you learned in Calculus I to evaluate integrals in this class. In other words, always take the approach that you find to be the easiest.

Note that this final point is more geared towards my class and it's completely possible that your instructor may not agree with this and so be careful in applying this point if you aren't in my class.

Okay, let's get on with the strategy.

Integration Strategy

1. **Simplify the integrand, if possible.** This step is very important in the integration process. Many integrals can be taken from impossible or very difficult to very easy with a little simplification or manipulation. Don't forget basic trig and algebraic identities as these can often be used to simplify the integral. We used this idea when we were looking at integrals involving trig functions. For example, consider the following integral.

$$\int \cos^2(x)\,dx$$

   This integral can't be done as is, however simply by recalling the identity,

$$\cos^2(x) = \frac{1}{2}\left(1 + \cos(2x)\right)$$

   the integral becomes very easy to do.

   Note that this example also shows that simplification does not necessarily mean that we'll write the integrand in a "simpler" form. It only means that we'll write the integrand into a form that we can deal with and this is often longer and/or "messier" than the original integral.

2. **See if a "simple", *i.e.* a u-substitution will work.** Look to see if a simple substitution can be used instead of the often more complicated methods from Calculus II. For example, consider both of the following integrals.

$$\int \frac{x}{x^2 - 1}\,dx \qquad\qquad \int x\sqrt{x^2 - 1}\,dx$$

   The first integral can be done with partial fractions and the second could be done with a trig substitution.

   However, both could also be evaluated using the substitution $u = x^2 - 1$ and the work involved in the substitution would be significantly less than the work involved in either partial fractions or trig substitution.

   So, always look for quick, simple substitutions before moving on to the more complicated Calculus II techniques.

3. **Identify the type of integral.** Note that any integral may fall into more than one of these types. Because of this fact it's usually best to go all the way through the list and identify all possible types since one may be easier than the other and it's entirely possible that the easier type is listed lower in the list.

   (a) Is the integrand a rational expression (*i.e* is the integrand a polynomial divided by a polynomial)? If so, then partial fractions may work on the integral.

(b) Is the integrand a polynomial times a trig function, exponential, or logarithm? If so, then integration by parts may work.

(c) Is the integrand a product of sines and cosines, secant and tangents, or cosecants and cotangents? If so, then the topics from the second section may work.

Likewise, don't forget that some quotients involving these functions can also be done using these techniques.

(d) Does the integrand involve $\sqrt{b^2x^2+a^2}$, $\sqrt{b^2x^2-a^2}$, or $\sqrt{a^2-b^2x^2}$? If so, then a trig substitution might work nicely.

(e) Does the integrand have roots other than those listed above in it? If so, then the substitution $u = \sqrt[n]{g(x)}$ might work.

(f) Does the integrand have a quadratic in it? If so, then completing the square on the quadratic might put it into a form that we can deal with.

4. **Can we relate the integral to an integral we already know how to do?** In other words, can we use a substitution or manipulation to write the integrand into a form that does fit into the forms we've looked at previously in this chapter. typical example here is the following integral.

$$\int \cos(x)\sqrt{1+\sin^2(x)}\,dx$$

This integral doesn't obviously fit into any of the forms we looked at in this chapter. However, with the substitution $u = \sin(x)$ we can reduce the integral to the form,

$$\int \sqrt{1+u^2}\,du$$

which is a trig substitution problem.

5. **Do we need to use multiple techniques?** In this step we need to ask ourselves if it is possible that we'll need to use multiple techniques. The example in the previous part is a good example. Using a substitution didn't allow us to actually do the integral. All it did was put the integral and put it into a form that we could use a different technique on. Don't ever get locked into the idea that an integral will only require one step to completely evaluate it. Many will require more than one step.

6. **Try again.** If everything that you've tried to this point doesn't work then go back through the process and try again. This time try a technique that you didn't use the first time around.

As noted above this strategy is not a hard and fast set of rules. It is only intended to guide you through the process of best determining how to do any given integral. Note as well that the only place Calculus II actually arises is in the third step. Steps 1, 2 and 4 involve nothing more than manipulation of the integrand either through direct manipulation of the integrand or by using a substitution. The last two steps are simply ideas to think about in going through this strategy.

Many students go through this process and concentrate almost exclusively on Step 3 (after all this is Calculus II, so it's easy to see why they might do that....) to the exclusion of the other steps. One very large consequence of that exclusion is that often a simple manipulation or substitution is overlooked that could make the integral very easy to do.

Before moving on to the next section we should work a couple of quick problems illustrating a couple of not so obvious simplifications/manipulations and a not so obvious substitution.

Example 1

Evaluate the following integral.

$$\int \frac{\tan(x)}{\sec^4(x)}\,dx$$

Solution

This integral almost falls into the form given in **3c**. It is a quotient of tangent and secant and we know that sometimes we can use the same methods for products of tangents and secants on quotients.

The process from that section tells us that if we have even powers of secant to strip two of them off and convert the rest to tangents. That won't work here. We can split two secants off, but they would be in the denominator and they won't do us any good there. Remember that the point of splitting them off is so they would be there for the substitution $u = \tan(x)$. That requires them to be in the numerator. So, that won't work and so we'll have to find another solution method.

There are in fact two solution methods to this integral depending on how you want to go about it. We'll take a look at both.

<u>**Solution 1**</u>

In this solution method we could just convert everything to sines and cosines and see if that gives us an integral we can deal with.

$$\begin{aligned}\int \frac{\tan(x)}{\sec^4(x)}\,dx &= \int \frac{\sin(x)}{\cos(x)}\cos^4(x)\,dx \\ &= \int \sin(x)\cos^3(x)\,dx \qquad\qquad u = \cos(x) \\ &= -\int u^3\,du \\ &= -\frac{1}{4}\cos^4(x) + c\end{aligned}$$

Note that just converting to sines and cosines won't always work and if it does it won't always work this nicely. Often there will be a lot more work that would need to be done to complete the integral.

Solution 2

This solution method goes back to dealing with secants and tangents. Let's notice that if we had a secant in the numerator we could just use $u = \sec(x)$ as a substitution and it would be a fairly quick and simple substitution to use. We don't have a secant in the numerator. However, we could very easily get a secant in the numerator simply by multiplying the numerator and denominator by secant.

$$\begin{aligned}\int \frac{\tan(x)}{\sec^4(x)}\,dx &= \int \frac{\tan(x)\sec(x)}{\sec^5(x)}\,dx \qquad\qquad u = \sec(x) \\ &= \int \frac{1}{u^5}\,du \\ &= -\frac{1}{4}\frac{1}{\sec^4(x)} + c \\ &= -\frac{1}{4}\cos^4(x) + c\end{aligned}$$

In the previous example we saw two "simplifications" that allowed us to do the integral. The first was using identities to rewrite the integral into terms we could deal with and the second involved multiplying the numerator and the denominator by something to again put the integral into terms we could deal with.

Using identities to rewrite an integral is an important "simplification" and we should not forget about it. Integrals can often be greatly simplified or at least put into a form that can be dealt with by using an identity.

The second "simplification" is not used as often, but does show up on occasion so again, it's best to not forget about it. In fact, let's take another look at an example in which multiplying the numerator and denominator by something will allow us to do an integral.

Example 2

Evaluate the following integral.

$$\int \frac{1}{1+\sin(x)}\,dx$$

Solution

This is an integral in which if we just concentrate on the third step we won't get anywhere. This integral doesn't appear to be any of the kinds of integrals that we worked in this chapter.

We can do the integral however, if we do the following,

$$\begin{aligned}\int \frac{1}{1+\sin(x)}\,dx &= \int \frac{1}{1+\sin(x)}\frac{1-\sin(x)}{1-\sin(x)}\,dx \\ &= \int \frac{1-\sin(x)}{1-\sin^2(x)}\,dx\end{aligned}$$

This does not appear to have done anything for us. However, if we now remember the first "simplification" we looked at above we will notice that we can use an identity to rewrite the denominator. Once we do that we can further reduce the integral into something we can deal with.

$$\begin{aligned}\int \frac{1}{1+\sin(x)}\,dx &= \int \frac{1-\sin(x)}{\cos^2(x)}\,dx \\ &= \int \frac{1}{\cos^2(x)} - \frac{\sin(x)}{\cos(x)}\frac{1}{\cos(x)}\,dx \\ &= \int \sec^2(x) - \tan(x)\sec(x)\,dx \\ &= \tan(x) - \sec(x) + c\end{aligned}$$

So, we've seen once again that multiplying the numerator and denominator by something can put the integral into a form that we can integrate. Notice as well that this example also showed that "simplifications" do not necessarily put an integral into a simpler form. They only put the integral into a form that is easier to integrate.

Let's now take a quick look at an example of a substitution that is not so obvious.

Example 3

Evaluate the following integral.

$$\int \cos\left(\sqrt{x}\right)\,dx$$

Solution

We introduced this example saying that the substitution was not so obvious. However, this is really an integral that falls into the form given by 3e in our strategy above. However, many people miss that form and so don't think about it. So, let's try the following substitution.

$$u = \sqrt{x} \qquad x = u^2 \qquad dx = 2u\,du$$

With this substitution the integral becomes,

$$\int \cos\left(\sqrt{x}\right)\,dx = 2\int u\cos(u)\,du$$

This is now an integration by parts integral. Remember that often we will need to use more than one technique to completely do the integral. This is a fairly simple integration by parts problem so we'll leave the remainder of the details to you to check.

$$\int \cos\left(\sqrt{x}\right)\,dx = 2\left(\cos\left(\sqrt{x}\right) + \sqrt{x}\,\sin\left(\sqrt{x}\right)\right) + c$$

Before leaving this section we should also point out that there are integrals out there in the world that just can't be done in terms of functions that we know. Some examples of these are.

$$\int \mathbf{e}^{-x^2}\,dx \qquad \int \cos\left(x^2\right)\,dx \qquad \int \frac{\sin(x)}{x}\,dx \qquad \int \cos\left(\mathbf{e}^x\right)\,dx$$

That doesn't mean that these integrals can't be done at some level. If you go to a computer algebra system such as Maple or Mathematica and have it do these integrals it will return the following.

$$\int \mathbf{e}^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}\mathrm{erf}(x)$$
$$\int \cos\left(x^2\right)\,dx = \sqrt{\frac{\pi}{2}}\mathrm{FresnelC}\left(x\sqrt{\frac{2}{\pi}}\right)$$
$$\int \frac{\sin(x)}{x}\,dx = \mathrm{Si}(x)$$
$$\int \cos\left(\mathbf{e}^x\right)\,dx = \mathrm{Ci}\left(\mathbf{e}^x\right)$$

So, it appears that these integrals can in fact be done. However, this is a little misleading. Here are the definitions of each of the functions given above.

Error Function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathbf{e}^{-t^2}\, dt$$

The Sine Integral

$$\operatorname{Si}(x) = \int_0^x \frac{\sin(t)}{t}\, dt$$

The Fresnel Cosine Integral

$$\operatorname{FresnelC}(x) = \int_0^x \cos\left(\frac{\pi}{2} t^2\right)\, dt$$

The Cosine Integral

$$\operatorname{Ci}(x) = \gamma + \ln(x) + \int_0^x \frac{\cos(t) - 1}{t}\, dt$$

Where γ is the Euler-Mascheroni constant.

Note that the first three are simply defined in terms of themselves and so when we say we can integrate them all we are really doing is renaming the integral. The fourth one is a little different and yet it is still defined in terms of an integral that can't be done in practice.

It will be possible to integrate every integral given in this class, but it is important to note that there are integrals that just can't be done. We should also note that after we look at Series we will be able to write down series representations of each of the integrals above.

7.8 Improper Integrals

In this section we need to take a look at a couple of different kinds of integrals. Both of these are examples of integrals that are called Improper Integrals.

Let's start with the first kind of improper integrals that we're going to take a look at.

Infinite Interval

In this kind of integral one or both of the limits of integration are infinity. In these cases, the interval of integration is said to be over an infinite interval.

Let's take a look at an example that will also show us how we are going to deal with these integrals.

Example 1

Evaluate the following integral.

$$\int_1^\infty \frac{1}{x^2}\,dx$$

Solution

This is an innocent enough looking integral. However, because infinity is not a real number we can't just integrate as normal and then "plug in" the infinity to get an answer.

To see how we're going to do this integral let's think of this as an area problem. So instead of asking what the integral is, let's instead ask what the area under $f(x) = \frac{1}{x^2}$ on the interval $[1, \infty)$ is.

We still aren't able to do this, however, let's step back a little and instead ask what the area under $f(x)$ is on the interval $[1, t]$ where $t > 1$ and t is finite. This is a problem that we can do.

$$A_t = \int_1^t \frac{1}{x^2}\,dx = -\frac{1}{x}\bigg|_1^t = 1 - \frac{1}{t}$$

Now, we can get the area under $f(x)$ on $[1, \infty)$ simply by taking the limit of A_t as t goes to infinity.

$$A = \lim_{t\to\infty} A_t = \lim_{t\to\infty}\left(1 - \frac{1}{t}\right) = 1$$

This is then how we will do the integral itself.

$$\begin{aligned}\int_1^\infty \frac{1}{x^2}\,dx &= \lim_{t\to\infty}\int_1^t \frac{1}{x^2}\,dx \\ &= \lim_{t\to\infty}\left(-\frac{1}{x}\right)\bigg|_1^t = \lim_{t\to\infty}\left(1 - \frac{1}{t}\right) = 1\end{aligned}$$

So, this is how we will deal with these kinds of integrals in general. We will replace the infinity with a variable (usually t), do the integral and then take the limit of the result as t goes to infinity.

On a side note, notice that the area under a curve on an infinite interval was not infinity as we might have suspected it to be. In fact, it was a surprisingly small number. Of course, this won't always be the case, but it is important enough to point out that not all areas on an infinite interval will yield infinite areas.

Let's now get some definitions out of the way. We will call these integrals **convergent** if the associated limit exists and is a finite number (*i.e.* it's not plus or minus infinity) and **divergent** if the associated limit either doesn't exist or is (plus or minus) infinity.

Let's now formalize up the method for dealing with infinite intervals. There are essentially three cases that we'll need to look at.

Integrals with Infinite Intervals

1. If $\int_a^t f(x)\,dx$ exists for every $t > a$ then,

$$\int_a^\infty f(x)\,dx = \lim_{t\to\infty}\int_a^t f(x)\,dx$$

provided the limit exists and is finite.

2. If $\int_t^b f(x)\,dx$ exists for every $t < b$ then,

$$\int_{-\infty}^b f(x)\,dx = \lim_{t\to-\infty}\int_t^b f(x)\,dx$$

provided the limits exists and is finite.

3. If $\int_{-\infty}^c f(x)\,dx$ and $\int_c^\infty f(x)\,dx$ are both convergent then,

$$\int_{-\infty}^\infty f(x)\,dx = \int_{-\infty}^c f(x)\,dx + \int_c^\infty f(x)\,dx$$

Where c is any number. Note as well that this requires BOTH of the integrals to be convergent in order for this integral to also be convergent. If either of the two integrals is divergent then so is this integral.

Let's take a look at a couple more examples.

Example 2

Determine if the following integral is convergent or divergent and if it's convergent find its value.

$$\int_1^\infty \frac{1}{x}\,dx$$

Solution

So, the first thing we do is convert the integral to a limit.

$$\int_1^\infty \frac{1}{x}\,dx = \lim_{t\to\infty}\int_1^t \frac{1}{x}\,dx$$

Now, do the integral and the limit.

$$\begin{aligned}\int_1^\infty \frac{1}{x}\,dx &= \lim_{t\to\infty} \ln(x)\Big|_1^t \\ &= \lim_{t\to\infty}\left(\ln(t) - \ln 1\right) \\ &= \infty\end{aligned}$$

So, the limit is infinite and so the integral is divergent.

If we go back to thinking in terms of area notice that the area under $g(x) = \frac{1}{x}$ on the interval $[1, \infty)$ is infinite. This is in contrast to the area under $f(x) = \frac{1}{x^2}$ which was quite small. There really isn't all that much difference between these two functions and yet there is a large difference in the area under them. We can actually extend this out to the following fact.

Fact

If $a > 0$ then

$$\int_a^\infty \frac{1}{x^p}\,dx$$

is convergent if $p > 1$ and divergent if $p \le 1$.

One thing to note about this fact is that it's in essence saying that if an integrand goes to zero fast enough then the integral will converge. How fast is fast enough? If we use this fact as a guide it looks like integrands that go to zero faster than $\frac{1}{x}$ goes to zero will probably converge.

Let's take a look at a couple more examples.

Example 3

Determine if the following integral is convergent or divergent. If it is convergent find its value.

$$\int_{-\infty}^{0} \frac{1}{\sqrt{3-x}}\, dx$$

Solution

There really isn't much to do with these problems once you know how to do them. We'll convert the integral to a limit/integral pair, evaluate the integral and then the limit.

$$\begin{aligned}
\int_{-\infty}^{0} \frac{1}{\sqrt{3-x}}\, dx &= \lim_{t\to-\infty} \int_{t}^{0} \frac{1}{\sqrt{3-x}}\, dx \\
&= \lim_{t\to-\infty} -2\sqrt{3-x}\Big|_{t}^{0} \\
&= \lim_{t\to-\infty} \left(-2\sqrt{3} + 2\sqrt{3-t}\right) \\
&= -2\sqrt{3} + \infty \\
&= \infty
\end{aligned}$$

So, the limit is infinite and so this integral is divergent.

Example 4

Determine if the following integral is convergent or divergent. If it is convergent find its value.

$$\int_{-\infty}^{\infty} x\mathbf{e}^{-x^2}\, dx$$

Solution

In this case we've got infinities in both limits. The process we are using to deal with the infinite limits requires only one infinite limit in the integral and so we'll need to split the integral up into two separate integrals. We can split the integral up at any point, so let's choose $x = 0$ since this will be a convenient point for the evaluation process. The integral is then,

$$\int_{-\infty}^{\infty} x\mathbf{e}^{-x^2}\, dx = \int_{-\infty}^{0} x\mathbf{e}^{-x^2}\, dx + \int_{0}^{\infty} x\mathbf{e}^{-x^2}\, dx$$

We've now got to look at each of the individual limits.

$$\begin{aligned}\int_{-\infty}^{0} x\mathbf{e}^{-x^2}\,dx &= \lim_{t\to-\infty}\int_{t}^{0} x\mathbf{e}^{-x^2}\,dx \\ &= \lim_{t\to-\infty}\left(-\frac{1}{2}\mathbf{e}^{-x^2}\right)\Big|_{t}^{0} \\ &= \lim_{t\to-\infty}\left(-\frac{1}{2}+\frac{1}{2}\mathbf{e}^{-t^2}\right) \\ &= -\frac{1}{2}\end{aligned}$$

So, the first integral is convergent. Note that this does NOT mean that the second integral will also be convergent. So, let's take a look at that one.

$$\begin{aligned}\int_{0}^{\infty} x\mathbf{e}^{-x^2}\,dx &= \lim_{t\to\infty}\int_{0}^{t} x\mathbf{e}^{-x^2}\,dx \\ &= \lim_{t\to\infty}\left(-\frac{1}{2}\mathbf{e}^{-x^2}\right)\Big|_{0}^{t} \\ &= \lim_{t\to\infty}\left(-\frac{1}{2}\mathbf{e}^{-t^2}+\frac{1}{2}\right) \\ &= \frac{1}{2}\end{aligned}$$

This integral is convergent and so since they are both convergent the integral we were actually asked to deal with is also convergent and its value is,

$$\int_{-\infty}^{\infty} x\mathbf{e}^{-x^2}\,dx = \int_{-\infty}^{0} x\mathbf{e}^{-x^2}\,dx + \int_{0}^{\infty} x\mathbf{e}^{-x^2}\,dx = -\frac{1}{2}+\frac{1}{2} = 0$$

Example 5

Determine if the following integral is convergent or divergent. If it is convergent find its value.

$$\int_{-2}^{\infty} \sin(x)\,dx$$

Solution

First convert to a limit.

$$\begin{aligned}\int_{-2}^{\infty} \sin(x)\,dx &= \lim_{t\to\infty}\int_{-2}^{t} \sin(x)\,dx \\ &= \lim_{t\to\infty}\left(-\cos(x)\right)\Big|_{-2}^{t} = \lim_{t\to\infty}\left(\cos(-2)-\cos(t)\right)\end{aligned}$$

This limit doesn't exist and so the integral is divergent.

In most examples in a Calculus II class that are worked over infinite intervals the limit either exists or is infinite. However, there are limits that don't exist, as the previous example showed, so don't forget about those.

Discontinuous Integrand

We now need to look at the second type of improper integrals that we'll be looking at in this section. These are integrals that have discontinuous integrands. The process here is basically the same with one subtle difference. Here are the general cases that we'll look at for these integrals.

Integrals with Discontinuous Integrands

1. If $f(x)$ is continuous on the interval $[a,b)$ and not continuous at $x=b$ then,

$$\int_{a}^{b} f(x)\,dx = \lim_{t\to b^-}\int_{a}^{t} f(x)\,dx$$

provided the limit exists and is finite. Note as well that we do need to use a left-hand limit here since the interval of integration is entirely on the left side of the upper limit.

2. If $f(x)$ is continuous on the interval $(a,b]$ and not continuous at $x=a$ then,

$$\int_{a}^{b} f(x)\,dx = \lim_{t\to a^+}\int_{t}^{b} f(x)\,dx$$

provided the limit exists and is finite. In this case we need to use a right-hand limit here since the interval of integration is entirely on the right side of the lower limit.

3. If $f(x)$ is not continuous at $x = c$ where $a < c < b$ and $\int_a^c f(x)\,dx$ and $\int_c^b f(x)\,dx$ are both convergent then,

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

As with the infinite interval case this requires BOTH of the integrals to be convergent in order for this integral to also be convergent. If either of the two integrals is divergent then so is this integral.

4. If $f(x)$ is not continuous at $x = a$ and $x = b$ and if $\int_a^c f(x)\,dx$ and $\int_c^b f(x)\,dx$ are both convergent then,

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

Where c is any number. Again, this requires BOTH of the integrals to be convergent in order for this integral to also be convergent.

Note that the limits in these cases really do need to be right or left-handed limits. Since we will be working inside the interval of integration we will need to make sure that we stay inside that interval. This means that we'll use one-sided limits to make sure we stay inside the interval.

Let's do a couple of examples of these kinds of integrals.

Example 6

Determine if the following integral is convergent or divergent. If it is convergent find its value.

$$\int_0^3 \frac{1}{\sqrt{3-x}}\,dx$$

Solution

The problem point is the upper limit so we are in the first case above.

$$\begin{aligned}\int_0^3 \frac{1}{\sqrt{3-x}}\,dx &= \lim_{t\to 3^-}\int_0^t \frac{1}{\sqrt{3-x}}\,dx \\ &= \lim_{t\to 3^-}\left(-2\sqrt{3-x}\right)\Big|_0^t = \lim_{t\to 3^-}\left(2\sqrt{3}-2\sqrt{3-t}\right) = 2\sqrt{3}\end{aligned}$$

The limit exists and is finite and so the integral converges and the integral's value is $2\sqrt{3}$.

Example 7

Determine if the following integral is convergent or divergent. If it is convergent find its value.

$$\int_{-2}^{3} \frac{1}{x^3}\,dx$$

Solution

This integrand is not continuous at $x = 0$ and so we'll need to split the integral up at that point.

$$\int_{-2}^{3} \frac{1}{x^3}\,dx = \int_{-2}^{0} \frac{1}{x^3}\,dx + \int_{0}^{3} \frac{1}{x^3}\,dx$$

Now we need to look at each of these integrals and see if they are convergent.

$$\begin{aligned}\int_{-2}^{0} \frac{1}{x^3}\,dx &= \lim_{t\to 0^-} \int_{-2}^{t} \frac{1}{x^3}\,dx \\ &= \lim_{t\to 0^-} \left(-\frac{1}{2x^2}\right)\Bigg|_{-2}^{t} \\ &= \lim_{t\to 0^-} \left(-\frac{1}{2t^2} + \frac{1}{8}\right) \\ &= -\infty\end{aligned}$$

At this point we're done. One of the integrals is divergent that means the integral that we were asked to look at is divergent. We don't even need to bother with the second integral.

Before leaving this section let's note that we can also have integrals that involve both of these cases. Consider the following integral.

Example 8

Determine if the following integral is convergent or divergent. If it is convergent find its value.

$$\int_{0}^{\infty} \frac{1}{x^2}\,dx$$

Solution

This is an integral over an infinite interval that also contains a discontinuous integrand. To do this integral we'll need to split it up into two integrals so each integral contains only one point of discontinuity. It is important to remember that all of the processes we are working with in this section so that each integral only contains one problem point.

We can split it up anywhere but pick a value that will be convenient for evaluation purposes.

$$\int_0^\infty \frac{1}{x^2}\,dx = \int_0^1 \frac{1}{x^2}\,dx + \int_1^\infty \frac{1}{x^2}\,dx$$

In order for the integral in the example to be convergent we will need BOTH of these to be convergent. If one or both are divergent then the whole integral will also be divergent.

We know that the second integral is convergent by the fact given in the infinite interval portion above. So, all we need to do is check the first integral.

$$\begin{aligned}\int_0^1 \frac{1}{x^2}\,dx &= \lim_{t\to 0^+}\int_t^1 \frac{1}{x^2}\,dx \\ &= \lim_{t\to 0^+}\left.\left(-\frac{1}{x}\right)\right|_t^1 \\ &= \lim_{t\to 0^+}\left(-1+\frac{1}{t}\right) \\ &= \infty\end{aligned}$$

So, the first integral is divergent and so the whole integral is divergent.

7.9 Comparison Test for Improper Integrals

Now that we've seen how to actually compute improper integrals we need to address one more topic about them. Often we aren't concerned with the actual value of these integrals. Instead we might only be interested in whether the integral is convergent or divergent. Also, there will be some integrals that we simply won't be able to integrate and yet we would still like to know if they converge or diverge.

To deal with this we've got a test for convergence or divergence that we can use to help us answer the question of convergence for an improper integral.

We will give this test only for a sub-case of the infinite interval integral, however versions of the test exist for the other sub-cases of the infinite interval integrals as well as integrals with discontinuous integrands.

Comparison Test

If $f(x) \geq g(x) \geq 0$ on the interval $[a, \infty)$ then,

1. If $\displaystyle\int_a^\infty f(x)\,dx$ converges then so does $\displaystyle\int_a^\infty g(x)\,dx$.
2. If $\displaystyle\int_a^\infty g(x)\,dx$ diverges then so does $\displaystyle\int_a^\infty f(x)\,dx$.

Note that if you think in terms of area the Comparison Test makes a lot of sense. If $f(x)$ is larger than $g(x)$ then the area under $f(x)$ must also be larger than the area under $g(x)$.

So, if the area under the larger function is finite (*i.e.* $\int_a^\infty f(x)\,dx$ converges) then the area under the smaller function must also be finite (*i.e.* $\int_a^\infty g(x)\,dx$ converges). Likewise, if the area under the smaller function is infinite (*i.e.* $\int_a^\infty g(x)\,dx$ diverges) then the area under the larger function must also be infinite (*i.e.* $\int_a^\infty f(x)\,dx$ diverges).

Be careful not to misuse this test. If the smaller function converges there is no reason to believe that the larger will also converge (after all infinity is larger than a finite number...) and if the larger function diverges there is no reason to believe that the smaller function will also diverge.

Let's work a couple of examples using the comparison test. Note that all we'll be able to do is determine the convergence of the integral. We won't be able to determine the value of the integrals and so won't even bother with that.

Example 1

Determine if the following integral is convergent or divergent.

$$\int_{2}^{\infty} \frac{\cos^2(x)}{x^2}\, dx$$

Solution

Let's take a second and think about how the Comparison Test works. If this integral is convergent then we'll need to find a larger function that also converges on the same interval. Likewise, if this integral is divergent then we'll need to find a smaller function that also diverges.

So, it seems like it would be nice to have some idea as to whether the integral converges or diverges ahead of time so we will know whether we will need to look for a larger (and convergent) function or a smaller (and divergent) function.

To get the guess for this function let's notice that the numerator is nice and bounded because we know that,

$$0 \le \cos^2(x) \le 1$$

Therefore, the numerator simply won't get too large.

So, it seems likely that the denominator will determine the convergence/divergence of this integral and we know that

$$\int_{2}^{\infty} \frac{1}{x^2}\, dx$$

converges since $p = 2 > 1$ by the fact in the previous section. So, let's guess that this integral will converge.

So we now know that we need to find a function that is larger than

$$\frac{\cos^2(x)}{x^2}$$

and also converges. Making a fraction larger is actually a fairly simple process. We can either make the numerator larger or we can make the denominator smaller. In this case we can't do a lot about the denominator in a way that will help. However, we can use the fact that $0 \le \cos^2(x) \le 1$ to make the numerator larger (*i.e.* we'll replace the cosine with something we know to be larger, namely 1). So,

$$\frac{\cos^2(x)}{x^2} \le \frac{1}{x^2}$$

Now, as we've already noted

$$\int_{2}^{\infty} \frac{1}{x^2}\, dx$$

converges and so by the Comparison Test we know that

$$\int_{2}^{\infty} \frac{\cos^2(x)}{x^2}\,dx$$

must also converge.

Example 2

Determine if the following integral is convergent or divergent.

$$\int_{3}^{\infty} \frac{1}{x + \mathbf{e}^x}\,dx$$

Solution

Let's first take a guess about the convergence of this integral. As noted after the fact in the last section about

$$\int_{a}^{\infty} \frac{1}{x^p}\,dx$$

if the integrand goes to zero faster than $\frac{1}{x}$ then the integral will probably converge. Now, we've got an exponential in the denominator which is approaching infinity much faster than the x and so it looks like this integral should probably converge.

So, we need a larger function that will also converge. In this case we can't really make the numerator larger and so we'll need to make the denominator smaller in order to make the function larger as a whole. We will need to be careful however. There are two ways to do this and only one, in this case, of them will work for us.

First, notice that since the lower limit of integration is 3 we can say that $x \geq 3 > 0$ and we know that exponentials are always positive. So, the denominator is the sum of two positive terms and if we were to drop one of them the denominator would get smaller. This would in turn make the function larger.

The question then is which one to drop? Let's first drop the exponential. Doing this gives,

$$\frac{1}{x + \mathbf{e}^x} < \frac{1}{x}$$

This is a problem however, since

$$\int_{3}^{\infty} \frac{1}{x}\,dx$$

diverges by the fact. We've got a larger function that is divergent. This doesn't say anything about the smaller function. Therefore, we chose the wrong one to drop.

Let's try it again and this time let's drop the x.

$$\frac{1}{x + \mathbf{e}^x} < \frac{1}{\mathbf{e}^x} = \mathbf{e}^{-x}$$

Also,

$$\begin{aligned}\int_3^\infty \mathbf{e}^{-x}\,dx &= \lim_{t\to\infty}\int_3^t \mathbf{e}^{-x}\,dx \\ &= \lim_{t\to\infty}\left(-\mathbf{e}^{-t} + \mathbf{e}^{-3}\right) \\ &= \mathbf{e}^{-3}\end{aligned}$$

So, $\int_3^\infty \mathbf{e}^{-x}\,dx$ is convergent. Therefore, by the Comparison test

$$\int_3^\infty \frac{1}{x + \mathbf{e}^x}\,dx$$

is also convergent.

Example 3

Determine if the following integral is convergent or divergent.

$$\int_3^\infty \frac{1}{x - \mathbf{e}^{-x}}\,dx$$

Solution

This is very similar to the previous example with a couple of very important differences. First, notice that the exponential now goes to zero as x increases instead of growing larger as it did in the previous example (because of the negative in the exponent). Also note that the exponential is now subtracted off the x instead of added onto it.

The fact that the exponential goes to zero means that this time the x in the denominator will probably dominate the term and that means that the integral probably diverges. We will therefore need to find a smaller function that also diverges.

Making fractions smaller is pretty much the same as making fractions larger. In this case we'll need to either make the numerator smaller or the denominator larger.

This is where the second change will come into play. As before we know that both x and the exponential are positive. However, this time since we are subtracting the exponential from the x if we were to drop the exponential the denominator will become larger (we will no longer be subtracting a positive number off the x) and so the fraction will become smaller.

In other words,

$$\frac{1}{x - \mathbf{e}^{-x}} > \frac{1}{x}$$

and we know that

$$\int_{3}^{\infty} \frac{1}{x}\, dx$$

diverges and so by the Comparison Test we know that

$$\int_{3}^{\infty} \frac{1}{x - \mathbf{e}^{-x}}\, dx$$

must also diverge.

Example 4

Determine if the following integral is convergent or divergent.

$$\int_{1}^{\infty} \frac{1 + 3\sin^4(2x)}{\sqrt{x}}\, dx$$

Solution

First notice that as with the first example, the numerator in this function is going to be bounded since the sine is never larger than 1. Therefore, since the exponent on the denominator is less than 1 we can guess that the integral will probably diverge. We will need a smaller function that also diverges.

We know that $0 \le 3\sin^4(2x) \le 3$. In particular, this term is positive and so if we drop it from the numerator the numerator will get smaller. This gives,

$$\frac{1 + 3\sin^4(2x)}{\sqrt{x}} > \frac{1}{\sqrt{x}}$$

and

$$\int_{1}^{\infty} \frac{1}{\sqrt{x}}\, dx$$

diverges so by the Comparison Test

$$\int_{1}^{\infty} \frac{1 + 3\sin^4(2x)}{\sqrt{x}}\, dx$$

also diverges.

Up to this point all the examples used on manipulation of either the numerator or the denominator in order to use the Comparison Test. Don't get so locked into that idea that you decide that is all you will ever have to do. Sometimes you will need to manipulate both the numerator and the denominator.

Let's do an example like that.

Example 5

Determine if the following integral is convergent or divergent.

$$\int_2^\infty \frac{1+\cos^2(x)}{\sqrt{x}\left[2-\sin^4(x)\right]}\,dx$$

Solution

In this case we can notice that because the cosine in the numerator is bounded the numerator will never get too large. Likewise, the sine in the denominator is bounded and so again that term will not get too large or too small.

That leaves only the square root in the denominator and because the exponent is less than one we can guess that the integral will probably diverge. Therefore, we will need a smaller function that also diverges.

We know that $0 \le \cos^2(x) \le 1$. In particular, this term is positive and so if we drop it from the numerator the numerator will get smaller. This gives,

$$\frac{1+\cos^2(x)}{\sqrt{x}\left[2-\sin^4(x)\right]} > \frac{1}{\sqrt{x}\left[2-\sin^4(x)\right]}$$

Next, we also know that $0 \le \sin^4(x) \le 1$. Again, this is a positive term and so if we no longer subtract this off from the 2 the term in the brackets will get larger and so the rational expression will get smaller. This gives,

$$\frac{1+\cos^2(x)}{\sqrt{x}\left[2-\sin^4(x)\right]} > \frac{1}{\sqrt{x}\left[2-\sin^4(x)\right]} > \frac{1}{2\sqrt{x}}$$

Finally, we know that

$$\int_2^\infty \frac{1}{2\sqrt{x}}\,dx$$

Diverges (the 2 in the denominator will not affect this) so by the Comparison Test

$$\int_2^\infty \frac{1+\cos^2(x)}{\sqrt{x}\left[2-\sin^4(x)\right]}\,dx$$

also diverges.

Okay, we've seen a few examples of the Comparison Test now. However, most of them worked pretty much the same way. All the functions were rational and all we did for most of them was add or subtract something from the numerator and/or the denominator to get what we want.

Let's take a look at an example that works a little differently so we don't get too locked into these ideas.

Example 6

Determine if the following integral is convergent or divergent.

$$\int_1^\infty \frac{\mathbf{e}^{-x}}{x}\,dx$$

Solution

Normally, the presence of just an x in the denominator would lead us to guess divergent for this integral. However, the exponential in the numerator will approach zero so fast that instead we'll need to guess that this integral converges.

To get a larger function we'll use the fact that we know from the limits of integration that $x > 1$. This means that if we just replace the x in the denominator with 1 (which is always smaller than x) we will make the denominator smaller and so the function will get larger.

$$\frac{\mathbf{e}^{-x}}{x} < \frac{\mathbf{e}^{-x}}{1} = \mathbf{e}^{-x}$$

and we can show that

$$\int_1^\infty \mathbf{e}^{-x}\,dx$$

converges. In fact, we've already done this for a lower limit of 3 and changing that to a 1 won't change the convergence of the integral. Therefore, by the Comparison Test

$$\int_1^\infty \frac{\mathbf{e}^{-x}}{x}\,dx$$

also converges.

We should also really work an example that doesn't involve a rational function since there is no reason to assume that we'll always be working with rational functions.

Example 7

Determine if the following integral is convergent or divergent.

$$\int_1^\infty \mathbf{e}^{-x^2}\,dx$$

Solution

We know that exponentials with negative exponents die down to zero very fast so it makes sense to guess that this integral will be convergent. We need a larger function, but this time we don't have a fraction to work with so we'll need to do something different.

We'll take advantage of the fact that $\mathbf{e}^{-x}$ is a decreasing function. This means that

$$x_1 > x_2 \qquad \Rightarrow \qquad \mathbf{e}^{-x_1} < \mathbf{e}^{-x_2}$$

In other words, plug in a larger number and the function gets smaller.

From the limits of integration we know that $x > 1$ and this means that if we square x it will get larger. Or,

$$x^2 > x \qquad \text{provided } x > 1$$

Note that we can only say this since $x > 1$. This won't be true if $x \leq 1$! We can now use the fact that $\mathbf{e}^{-x}$ is a decreasing function to get,

$$\mathbf{e}^{-x^2} < \mathbf{e}^{-x}$$

So, $\mathbf{e}^{-x}$ is a larger function than $\mathbf{e}^{-x^2}$ and we know that

$$\int_1^\infty \mathbf{e}^{-x}\,dx$$

converges so by the Comparison Test we also know that

$$\int_1^\infty \mathbf{e}^{-x^2}\,dx$$

is convergent.

The last two examples made use of the fact that $x > 1$. Let's take a look at an example to see how we would have to go about these if the lower limit had been smaller than 1.

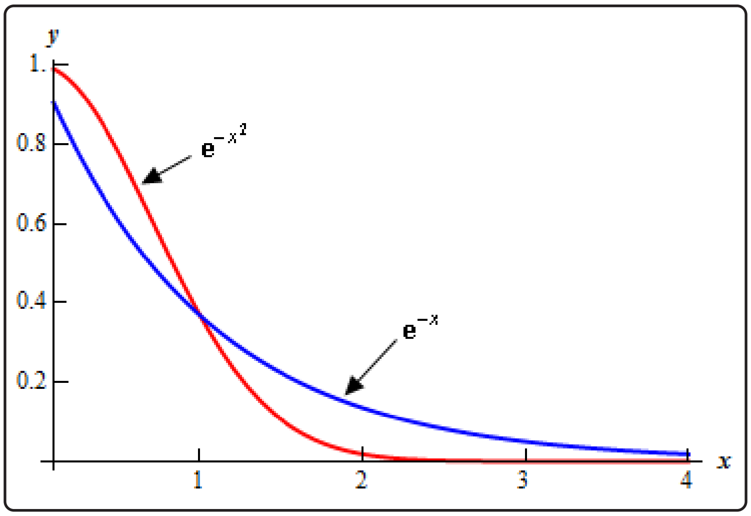
Example 8

Determine if the following integral is convergent or divergent.

$$\int_{\frac{1}{2}}^{\infty} \mathbf{e}^{-x^2}\, dx$$

Solution

First, we need to note that $\mathbf{e}^{-x^2} \le \mathbf{e}^{-x}$ is only true on the interval $[1, \infty)$ as is illustrated in the graph below.



So, we can't just proceed as we did in the previous example with the Comparison Test on the interval $\left[\frac{1}{2}, \infty\right)$. However, this isn't the problem it might at first appear to be. We can always write the integral as follows,

$$\begin{aligned}\int_{\frac{1}{2}}^{\infty} \mathbf{e}^{-x^2}\, dx &= \int_{\frac{1}{2}}^{1} \mathbf{e}^{-x^2}\, dx + \int_{1}^{\infty} \mathbf{e}^{-x^2}\, dx \\ &= 0.28554 + \int_{1}^{\infty} \mathbf{e}^{-x^2}\, dx\end{aligned}$$

We used Mathematica to get the value of the first integral. Now, if the second integral converges it will have a finite value and so the sum of two finite values will also be finite and so the original integral will converge. Likewise, if the second integral diverges it will either be infinite or not have a value at all and adding a finite number onto this will not all of a sudden make it finite or exist and so the original integral will diverge. Therefore, this integral will converge or diverge depending only on the convergence of the second integral.

As we saw in Example 7 the second integral does converge and so the whole integral must also converge.

As we saw in this example, if we need to, we can split the integral up into one that doesn't involve any problems and can be computed and one that may contain a problem that we can use the Comparison Test on to determine its convergence.

7.10 Approximating Definite Integrals

In this chapter we've spent quite a bit of time on computing the values of integrals. However, not all integrals can be computed. A perfect example is the following definite integral.

$$\int_0^2 \mathbf{e}^{x^2}\,dx$$

We now need to talk a little bit about estimating values of definite integrals. We will look at three different methods, although one should already be familiar to you from your Calculus I days. We will develop all three methods for estimating

$$\int_a^b f(x)\,dx$$

by thinking of the integral as an area problem and using known shapes to estimate the area under the curve.

Let's get first develop the methods and then we'll try to estimate the integral shown above.

Midpoint Rule

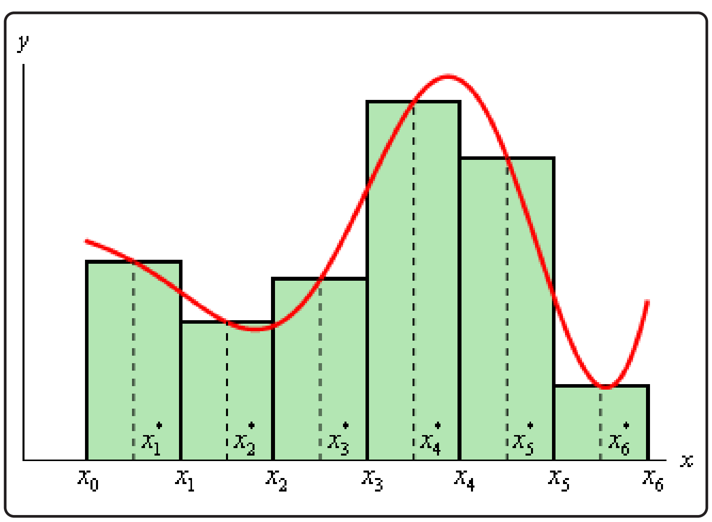
This is the rule that should be somewhat familiar to you. We will divide the interval $[a, b]$ into n subintervals of equal width,

$$\Delta x = \frac{b-a}{n}$$

We will denote each of the intervals as follows,

$$[x_0, x_1],\ [x_1, x_2], \ldots, [x_{n-1}, x_n] \qquad \text{where } x_0 = a \text{ and } x_n = b$$

Then for each interval let x_i^* be the midpoint of the interval. We then sketch in rectangles for each subinterval with a height of $f(x_i^*)$. Here is a graph showing the set up using $n = 6$.



We can easily find the area for each of these rectangles and so for a general n we get that,

$$\int_a^b f(x)\,dx \approx \Delta x\, f(x_1^*) + \Delta x\, f(x_2^*) + \cdots + \Delta x\, f(x_n^*)$$

Or, upon factoring out a Δx we get the general Midpoint Rule.

Midpoint Rule

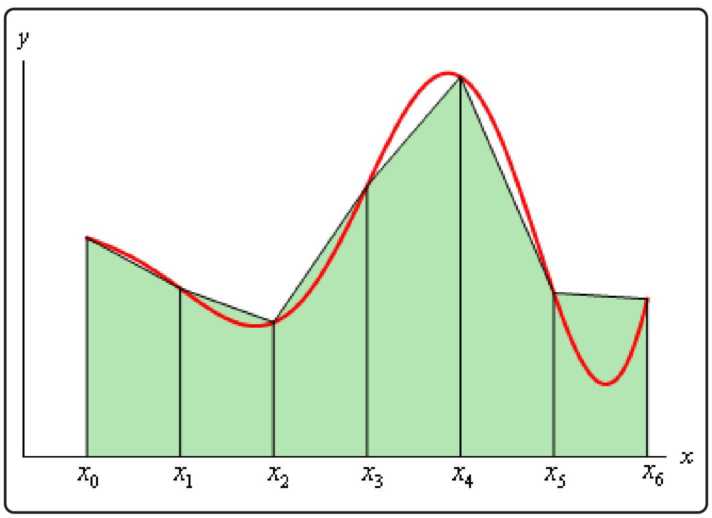
$$\int_a^b f(x)\,dx \approx \Delta x\left[f(x_1^*) + f(x_2^*) + \cdots + f(x_n^*)\right]$$

Trapezoid Rule

For this rule we will do the same set up as for the Midpoint Rule. We will break up the interval $[a, b]$ into n subintervals of width,

$$\Delta x = \frac{b-a}{n}$$

Then on each subinterval we will approximate the function with a straight line that is equal to the function values at either endpoint of the interval. Here is a sketch of this case for $n = 6$.



Each of these objects is a trapezoid (hence the rule's name...) and as we can see some of them do a very good job of approximating the actual area under the curve and others don't do such a good job.

The area of the trapezoid in the interval $[x_{i-1}, x_i]$ is given by,

$$A_i = \frac{\Delta x}{2}\Big(f(x_{i-1}) + f(x_i)\Big)$$

So, if we use n subintervals the integral is approximately,

$$\int_a^b f(x)\,dx \approx \frac{\Delta x}{2}\left(f(x_0)+f(x_1)\right)+\frac{\Delta x}{2}\left(f(x_1)+f(x_2)\right)+\cdots+\frac{\Delta x}{2}\left(f(x_{n-1})+f(x_n)\right)$$

Upon doing a little simplification we arrive at the general Trapezoid Rule.

Trapezoid Rule

$$\int_a^b f(x)\,dx \approx \frac{\Delta x}{2}\left[f(x_0)+2f(x_1)+2f(x_2)+\cdots+2f(x_{n-1})+f(x_n)\right]$$

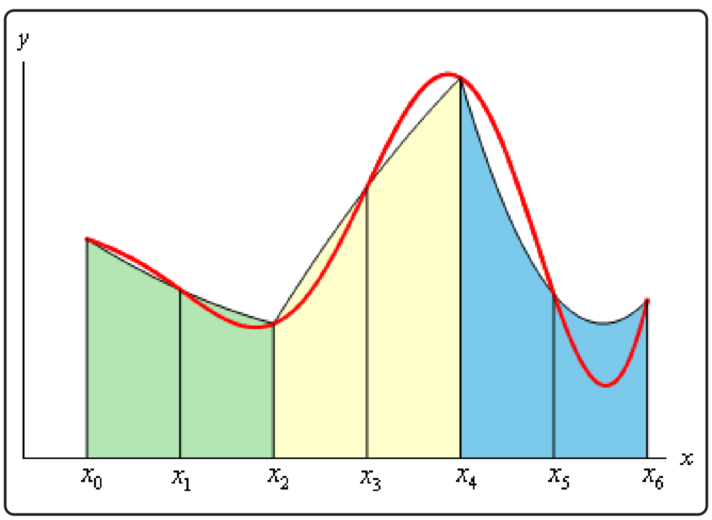
Note that all the function evaluations, with the exception of the first and last, are multiplied by 2.

Simpson's Rule

This is the final method we're going to take a look at and in this case we will again divide up the interval $[a,b]$ into n subintervals. However, unlike the previous two methods we need to require that n be even. The reason for this will be evident in a bit. The width of each subinterval is,

$$\Delta x = \frac{b-a}{n}$$

In the Trapezoid Rule we approximated the curve with a straight line. For Simpson's Rule we are going to approximate the function with a quadratic and we're going to require that the quadratic agree with three of the points from our subintervals. Below is a sketch of this using $n = 6$. Each of the approximations is colored differently so we can see how they actually work.



Notice that each approximation actually covers two of the subintervals. This is the reason for requiring n to be even. Some of the approximations look more like a line than a quadratic, but they really are quadratics. Also note that some of the approximations do a better job than others. It can be shown that the area under the approximation on the intervals $[x_{i-1}, x_i]$ and $[x_i, x_{i+1}]$ is,

$$A_i = \frac{\Delta x}{3}\Big(f(x_{i-1}) + 4f(x_i) + f(x_{i+1})\Big)$$

If we use n subintervals the integral is then approximately,

$$\int_a^b f(x)\,dx \approx \frac{\Delta x}{3}(f(x_0) + 4f(x_1) + f(x_2)) + \frac{\Delta x}{3}(f(x_2) + 4f(x_3) + f(x_4)) + \cdots + \frac{\Delta x}{3}(f(x_{n-2}) + 4f(x_{n-1}) + f(x_n))$$

Upon simplifying we arrive at the general Simpson's Rule.

Simpson's Rule

$$\int_a^b f(x)\,dx \approx \frac{\Delta x}{3}\Big[f(x_0) + 4f(x_1) + 2f(x_2) + \cdots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)\Big]$$

In this case notice that all the function evaluations at points with odd subscripts are multiplied by 4 and all the function evaluations at points with even subscripts (except for the first and last) are multiplied by 2. If you can remember this, this is a fairly easy rule to remember.

Okay, it's time to work an example and see how these rules work.

Example 1

Using $n = 4$ and all three rules to approximate the value of the following integral.

$$\int_0^2 \mathbf{e}^{x^2}\,dx$$

Solution

First, for reference purposes, Mathematica gives the following value for this integral.

$$\int_0^2 \mathbf{e}^{x^2}\,dx = 16.45262776$$

In each case the width of the subintervals will be,

$$\Delta x = \frac{2-0}{4} = \frac{1}{2}$$

and so the subintervals will be,

$$[0,\ 0.5],\ [0.5,\ 1],\ [1,\ 1.5],\ [1.5,\ 2]$$

Let's go through each of the methods.

Midpoint Rule

$$\int_0^2 \mathbf{e}^{x^2}\,dx \approx \frac{1}{2}\left(\mathbf{e}^{(0.25)^2} + \mathbf{e}^{(0.75)^2} + \mathbf{e}^{(1.25)^2} + \mathbf{e}^{(1.75)^2}\right) = 14.48561253$$

Remember that we evaluate at the midpoints of each of the subintervals here! The Midpoint Rule has an error of 1.96701523.

Trapezoid Rule

$$\int_0^2 \mathbf{e}^{x^2}\,dx \approx \frac{1/2}{2}\left(\mathbf{e}^{(0)^2} + 2\mathbf{e}^{(0.5)^2} + 2\mathbf{e}^{(1)^2} + 2\mathbf{e}^{(1.5)^2} + \mathbf{e}^{(2)^2}\right) = 20.64455905$$

The Trapezoid Rule has an error of 4.19193129

Simpson's Rule

$$\int_0^2 \mathbf{e}^{x^2}\,dx \approx \frac{1/2}{3}\left(\mathbf{e}^{(0)^2} + 4\mathbf{e}^{(0.5)^2} + 2\mathbf{e}^{(1)^2} + 4\mathbf{e}^{(1.5)^2} + \mathbf{e}^{(2)^2}\right) = 17.35362645$$

The Simpson's Rule has an error of 0.90099869.

None of the estimations in the previous example are all that good. The best approximation in this case is from the Simpson's Rule and yet it still had an error of almost 1. To get a better estimation we would need to use a larger n. So, for completeness sake here are the estimates for some larger value of n.

	Midpoint		Trapezoid		Simpson's	
n	Approx.	Error	Approx.	Error	Approx.	Error
8	15.9056767	0.5469511	17.5650858	1.1124580	16.5385947	0.0859669
16	16.3118539	0.1407739	16.7353812	0.2827535	16.4588131	0.0061853
32	16.4171709	0.0354568	16.5236176	0.0709898	16.4530297	0.0004019
64	16.4437469	0.0088809	16.4703942	0.0177665	16.4526531	0.0000254
128	16.4504065	0.0022212	16.4570706	0.0044428	16.4526294	0.0000016

In this case we were able to determine the error for each estimate because we could get our hands on the exact value. Often this won't be the case and so we'd next like to look at error bounds for each estimate.

These bounds will give the largest possible error in the estimate, but it should also be pointed out that the actual error may be significantly smaller than the bound. The bound is only there so we can say that we know the actual error will be less than the bound.

So, suppose that $|f''(x)| \le K$ and $\left|f^{(4)}(x)\right| \le M$ for $a \le x \le b$ then if E_M, E_T, and E_S are the actual errors for the Midpoint, Trapezoid and Simpson's Rule we have the following bounds,

$$|E_M| \le \frac{K(b-a)^3}{24n^2} \qquad |E_T| \le \frac{K(b-a)^3}{12n^2} \qquad |E_S| \le \frac{M(b-a)^5}{180n^4}$$

Example 2

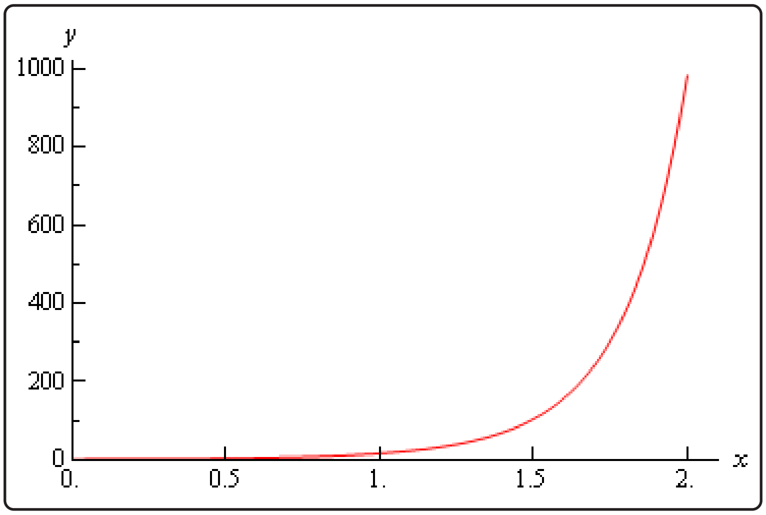
Determine the error bounds for the estimations in the last example.

Solution

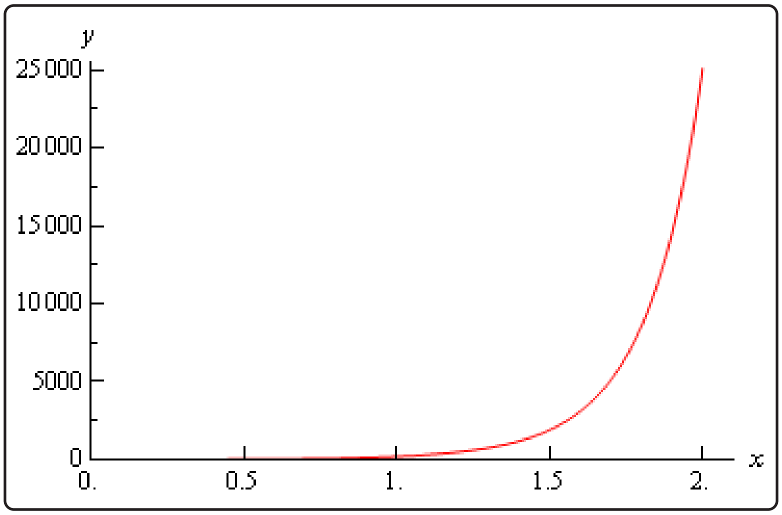
We already know that $n = 4$, $a = 0$, and $b = 2$ so we just need to compute K (the largest value of the second derivative) and M (the largest value of the fourth derivative). This means that we'll need the second and fourth derivative of $f(x)$.

$$f''(x) = 2\mathbf{e}^{x^2}\left(1+2x^2\right)$$
$$f^{(4)}(x) = 4\mathbf{e}^{x^2}\left(3+12x^2+4x^4\right)$$

Here is a graph of the second derivative.



Here is a graph of the fourth derivative.



So, from these graphs it's clear that the largest value of both of these are at $x = 2$. So,

$$f''(2) = 982.7667 \qquad \Rightarrow \qquad K = 983$$
$$f^{(4)}(2) = 25115.14901 \qquad \Rightarrow \qquad M = 25116$$

We rounded to make the computations simpler. Note however, that this does not need to be done.

Here are the bounds for each rule.

$$|E_M| \leq \frac{983(2-0)^3}{24(4)^2} = 20.4791666667$$

$$|E_T| \leq \frac{983(2-0)^3}{12(4)^2} = 40.9583333333$$

$$|E_S| \leq \frac{25116(2-0)^5}{180(4)^4} = 17.4416666667$$

In each case we can see that the errors are significantly smaller than the actual bounds.

8 More Applications of Integrals

It is now time to take a look at some more applications of integrals. As noted the last time we looked at applications of integrals many, although, not all of these new applications in this chapter have a fairly high chance of needing some of the integration techniques from the last chapter.

The first application, Arc Length can be kept to only u-substitutions at the worst, but most of those problems tend to be very simple. Once we start moving into more complicated problems arc length problems they tend to involve trig substitutions.

The next application, Surface Area tends to be u-substitutions but the notation used here is also used in the Arc Length section and so the surface area section is also here because of the shared notation.

Center of Mass and Probability are applications that will, in almost every case, involve integration by parts. In addition, the Probability section has the potential for improper integrals to show up.

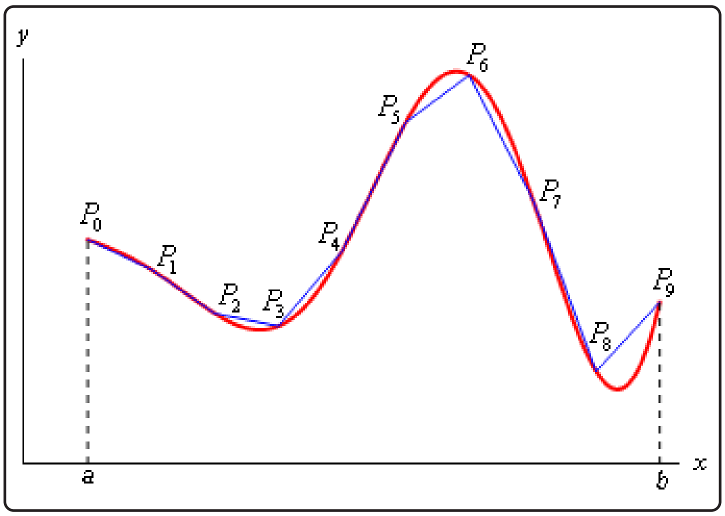
The other application we'll be looking at in this chapter, Hydrostatic Pressure and Force, will typically involve fairly simple integrals that could have been done in the earlier chapter. The reason the topic is here is because we have to derive up the integral using the definition of the definite integral in every problem. In addition, more complicated problems could lead to much more complicated integrals. The integrals in this section are kept simple mostly to keep the derivation work simpler.

8.1 Arc Length

In this section we are going to look at computing the arc length of a function. Because it's easy enough to derive the formulas that we'll use in this section we will derive one of them and leave the other to you to derive.

We want to determine the length of the continuous function $y = f(x)$ on the interval $[a, b]$. We'll also need to assume that the derivative is continuous on $[a, b]$.

Initially we'll need to estimate the length of the curve. We'll do this by dividing the interval up into n equal subintervals each of width Δx and we'll denote the point on the curve at each point by P_i. We can then approximate the curve by a series of straight lines connecting the points. Here is a sketch of this situation for $n = 9$.



Now denote the length of each of these line segments by $|P_{i-1}\ P_i|$ and the length of the curve will then be approximately,

$$L \approx \sum_{i=1}^{n} |P_{i-1}\ P_i|$$

and we can get the exact length by taking n larger and larger. In other words, the exact length will be,

$$L = \lim_{n\to\infty} \sum_{i=1}^{n} |P_{i-1}\ P_i|$$

Now, let's get a better grasp on the length of each of these line segments. First, on each segment let's define $\Delta y_i = y_i - y_{i-1} = f(x_i) - f(x_{i-1})$. We can then compute directly the length of the line segments as follows.

$$|P_{i-1}\ P_i| = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} = \sqrt{\Delta x^2 + \Delta y_i^2}$$

By the Mean Value Theorem we know that on the interval $[x_{i-1}, x_i]$ there is a point x_i^* so that,

$$\begin{aligned} f(x_i) - f(x_{i-1}) &= f'(x_i^*)(x_i - x_{i-1}) \\ \Delta y_i &= f'(x_i^*)\,\Delta x \end{aligned}$$

Therefore, the length can now be written as,

$$\begin{aligned} |P_{i-1}\ P_i| &= \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \\ &= \sqrt{\Delta x^2 + \left[f'(x_i^*)\right]^2 \Delta x^2} \\ &= \sqrt{1 + \left[f'(x_i^*)\right]^2}\ \ \Delta x \end{aligned}$$

The exact length of the curve is then,

$$\begin{aligned} L &= \lim_{n\to\infty} \sum_{i=1}^{n} |P_{i-1}\ P_i| \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \sqrt{1 + \left[f'(x_i^*)\right]^2}\ \ \Delta x \end{aligned}$$

However, using the definition of the definite integral, this is nothing more than,

$$L = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx$$

A slightly more convenient notation (in our opinion anyway) is the following.

$$L = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx$$

In a similar fashion we can also derive a formula for $x = h(y)$ on $[c, d]$. This formula is,

$$L = \int_c^d \sqrt{1 + \left[h'(y)\right]^2}\, dy = \int_c^d \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy$$

Again, the second form is probably a little more convenient.

Note the difference in the derivative under the square root! Don't get too confused. With one we differentiate with respect to x and with the other we differentiate with respect to y. One way to keep the two straight is to notice that the differential in the "denominator" of the derivative will match up with the differential in the integral. This is one of the reasons why the second form is a little more convenient.

Before we work any examples we need to make a small change in notation. Instead of having two formulas for the arc length of a function we are going to reduce it, in part, to a single formula.

From this point on we are going to use the following formula for the length of the curve.

Arc Length Formula(s)

$$L = \int ds$$

where,

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \qquad \text{if } y = f(x),\ a \le x \le b$$

$$ds = \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy \qquad \text{if } x = h(y),\ c \le y \le d$$

Note that no limits were put on the integral as the limits will depend upon the ds that we're using. Using the first ds will require x limits of integration and using the second ds will require y limits of integration.

Thinking of the arc length formula as a single integral with different ways to define ds will be convenient when we run across arc lengths in future sections. Also, this ds notation will be a nice notation for the next section as well.

Now that we've derived the arc length formula let's work some examples.

Example 1

Determine the length of $y = \ln\left(\sec(x)\right)$ between $0 \le x \le \frac{\pi}{4}$.

Solution

In this case we'll need to use the first ds since the function is in the form $y = f(x)$. So, let's get the derivative out of the way.

$$\frac{dy}{dx} = \frac{\sec(x)\tan(x)}{\sec(x)} = \tan(x) \qquad \left(\frac{dy}{dx}\right)^2 = \tan^2(x)$$

Let's also get the root out of the way since there is often simplification that can be done and there's no reason to do that inside the integral.

$$\sqrt{1 + \left(\frac{dy}{dx}\right)^2} = \sqrt{1 + \tan^2(x)} = \sqrt{\sec^2(x)} = |\sec(x)| = \sec(x)$$

Note that we could drop the absolute value bars here since secant is positive in the range given.

The arc length is then,

$$\begin{aligned} L &= \int_0^{\frac{\pi}{4}} \sec(x)\,dx \\ &= \ln\left|\sec(x) + \tan(x)\right|\Big|_0^{\frac{\pi}{4}} \\ &= \ln\left(\sqrt{2}+1\right) \end{aligned}$$

Example 2

Determine the length of $x = \frac{2}{3}(y-1)^{\frac{3}{2}}$ between $1 \le y \le 4$.

Solution

There is a very common mistake that students make in problems of this type. Many students see that the function is in the form $x = h(y)$ and they immediately decide that it will be too difficult to work with it in that form so they solve for y to get the function into the form $y = f(x)$. While that can be done here it will lead to a messier integral for us to deal with.

Sometimes it's just easier to work with functions in the form $x = h(y)$. In fact, if you can work with functions in the form $y = f(x)$ then you can work with functions in the form $x = h(y)$. There really isn't a difference between the two so don't get excited about functions in the form $x = h(y)$.

Let's compute the derivative and the root.

$$\frac{dx}{dy} = (y-1)^{\frac{1}{2}} \qquad \Rightarrow \qquad \sqrt{1+\left(\frac{dx}{dy}\right)^2} = \sqrt{1+y-1} = \sqrt{y}$$

As you can see keeping the function in the form $x = h(y)$ is going to lead to a very easy integral. To see what would happen if we tried to work with the function in the form $y = f(x)$ see the next example.

Let's get the length.

$$\begin{aligned} L &= \int_1^4 \sqrt{y}\,dy \\ &= \frac{2}{3}y^{\frac{3}{2}}\Big|_1^4 \\ &= \frac{14}{3} \end{aligned}$$

As noted in the last example we really do have a choice as to which ds we use. Provided we can get the function in the form required for a particular ds we can use it. However, as also noted above, there will often be a significant difference in difficulty in the resulting integrals. Let's take a quick look at what would happen in the previous example if we did put the function into the form $y = f(x)$.

Example 3

Redo the previous example using the function in the form $y = f(x)$ instead.

Solution

In this case the function and its derivative would be,

$$y = \left(\frac{3x}{2}\right)^{\frac{2}{3}} + 1 \qquad \frac{dy}{dx} = \left(\frac{3x}{2}\right)^{-\frac{1}{3}}$$

The root in the arc length formula would then be.

$$\sqrt{1 + \left(\frac{dy}{dx}\right)^2} = \sqrt{1 + \frac{1}{\left(\frac{3x}{2}\right)^{\frac{2}{3}}}} = \sqrt{\frac{\left(\frac{3x}{2}\right)^{\frac{2}{3}} + 1}{\left(\frac{3x}{2}\right)^{\frac{2}{3}}}} = \frac{\sqrt{\left(\frac{3x}{2}\right)^{\frac{2}{3}} + 1}}{\left(\frac{3x}{2}\right)^{\frac{1}{3}}}$$

All the simplification work above was just to put the root into a form that will allow us to do the integral.

Now, before we write down the integral we'll also need to determine the limits. This particular ds requires x limits of integration and we've got y limits. They are easy enough to get however. Since we know x as a function of y all we need to do is plug in the original y limits of integration and get the x limits of integration. Doing this gives,

$$0 \le x \le \frac{2}{3}(3)^{\frac{3}{2}}$$

Not easy limits to deal with, but there they are.

Let's now write down the integral that will give the length.

$$L = \int_0^{\frac{2}{3}(3)^{\frac{3}{2}}} \frac{\sqrt{\left(\frac{3x}{2}\right)^{\frac{2}{3}} + 1}}{\left(\frac{3x}{2}\right)^{\frac{1}{3}}}\, dx$$

That's a really unpleasant looking integral. It can be evaluated however using the following substitution.

$$u = \left(\frac{3x}{2}\right)^{\frac{2}{3}} + 1 \qquad du = \left(\frac{3x}{2}\right)^{-\frac{1}{3}} dx$$

$$x = 0 \quad \Rightarrow \quad u = 1$$
$$x = \frac{2}{3}(3)^{\frac{3}{2}} \quad \Rightarrow \quad u = 4$$

Using this substitution the integral becomes,

$$\begin{aligned} L &= \int_1^4 \sqrt{u}\, du \\ &= \left.\frac{2}{3}u^{\frac{3}{2}}\right|_1^4 \\ &= \frac{14}{3} \end{aligned}$$

So, we got the same answer as in the previous example. Although that shouldn't really be all that surprising since we were dealing with the same curve.

From a technical standpoint the integral in the previous example was not that difficult. It was just a Calculus I substitution. However, from a practical standpoint the integral was significantly more difficult than the integral we evaluated in Example 2. So, the moral of the story here is that we can use either formula (provided we can get the function in the correct form of course) however one will often be significantly easier to actually evaluate.

Okay, let's work one more example.

Example 4

Determine the length of $x = \frac{1}{2}y^2$ for $0 \le x \le \frac{1}{2}$. Assume that y is positive.

Solution

We'll use the second ds for this one as the function is already in the correct form for that one. Also, the other ds would again lead to a particularly difficult integral. The derivative and root will then be,

$$\frac{dx}{dy} = y \qquad \Rightarrow \qquad \sqrt{1 + \left(\frac{dx}{dy}\right)^2} = \sqrt{1 + y^2}$$

Before writing down the length notice that we were given x limits and we will need y limits for this ds. With the assumption that y is positive these are easy enough to get. All we need to do is plug x into our equation and solve for y. Doing this gives,

$$0 \le y \le 1$$

The integral for the arc length is then,

$$L = \int_0^1 \sqrt{1+y^2}\,dy$$

This integral will require the following trig substitution.

$$y = \tan(\theta) \qquad\qquad dy = \sec^2(\theta)\,d\theta$$

$$\begin{aligned} y = 0 \quad &\Rightarrow \quad 0 = \tan(\theta) \quad \Rightarrow \quad \theta = 0 \\ y = 1 \quad &\Rightarrow \quad 1 = \tan(\theta) \quad \Rightarrow \quad \theta = \frac{\pi}{4} \end{aligned}$$

$$\sqrt{1+y^2} = \sqrt{1+\tan^2(\theta)} = \sqrt{\sec^2(\theta)} = |\sec(\theta)| = \sec(\theta)$$

The length is then,

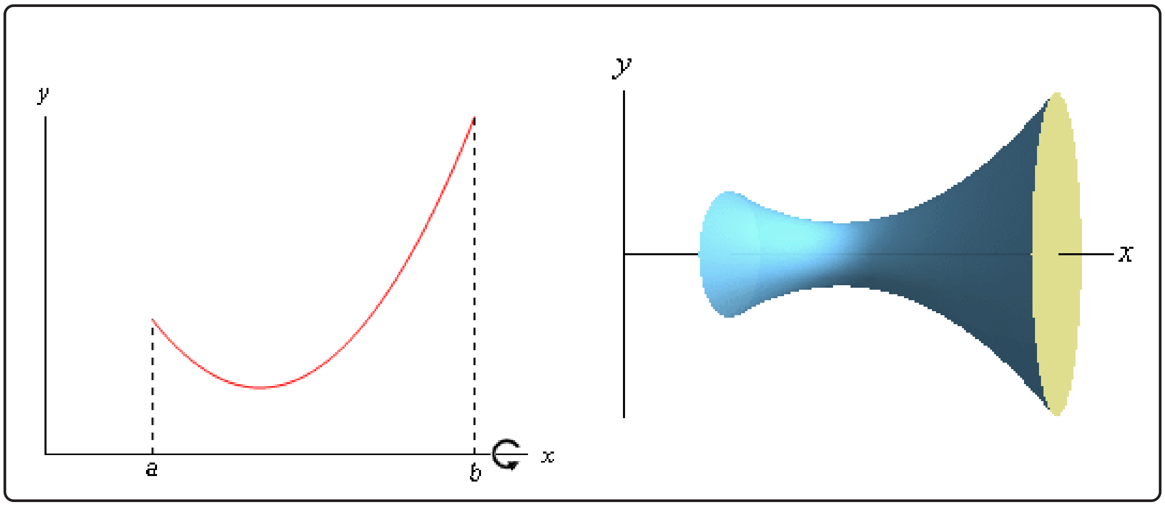
$$\begin{aligned} L &= \int_0^{\frac{\pi}{4}} \sec^3(\theta)\,d\theta \\ &= \frac{1}{2}\Big(\sec(\theta)\tan(\theta) + \ln|\sec(\theta)+\tan(\theta)|\Big)\Big|_0^{\frac{\pi}{4}} \\ &= \frac{1}{2}\left(\sqrt{2} + \ln\left(1+\sqrt{2}\right)\right) \end{aligned}$$

The first couple of examples ended up being fairly simple Calculus I substitutions. However, as this last example had shown we can end up with trig substitutions as well for these integrals.

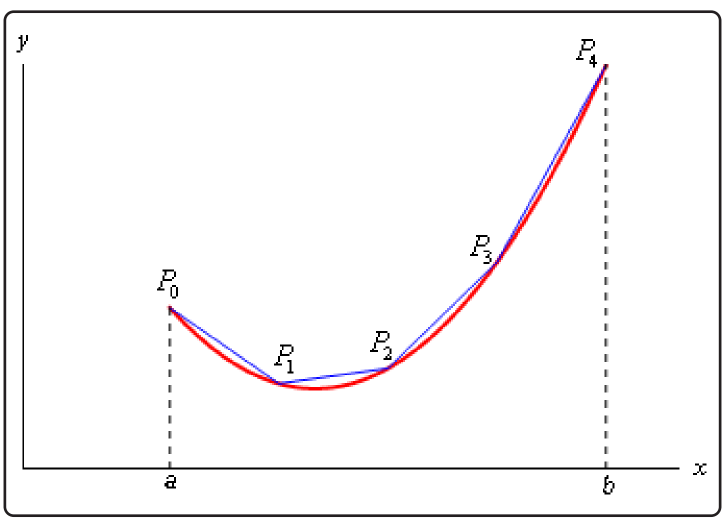
8.2 Surface Area

In this section we are going to look once again at solids of revolution. We first looked at them back in Calculus I when we found the volume of the solid of revolution. In this section we want to find the surface area of this region.

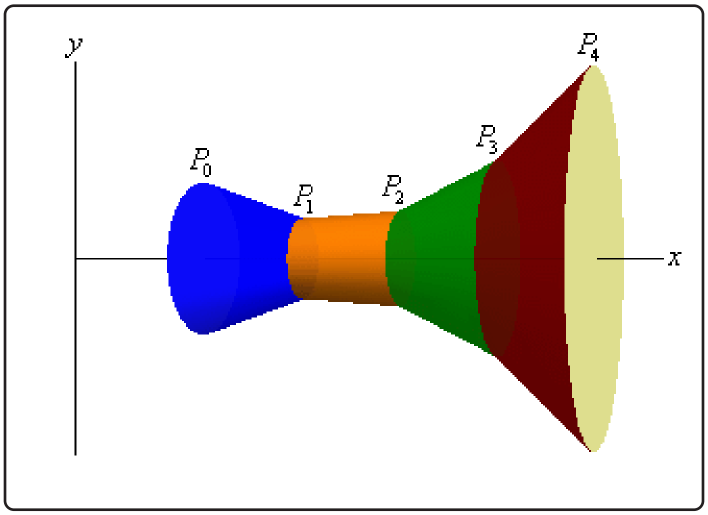
So, for the purposes of the derivation of the formula, let's look at rotating the continuous function $y = f(x)$ in the interval $[a, b]$ about the x-axis. We'll also need to assume that the derivative is continuous on $[a, b]$. Below is a sketch of a function and the solid of revolution we get by rotating the function about the x-axis.



We can derive a formula for the surface area much as we derived the formula for arc length. We'll start by dividing the interval into n equal subintervals of width Δx. On each subinterval we will approximate the function with a straight line that agrees with the function at the endpoints of each interval. Here is a sketch of that for our representative function using $n = 4$.



Now, rotate the approximations about the x-axis and we get the following solid.



The approximation on each interval gives a distinct portion of the solid and to make this clear each portion is colored differently. Each of these portions are called frustums and we know how to find the surface area of frustums.

The surface area of a frustum is given by,

$$A = 2\pi r l$$

where,

$$r = \frac{1}{2}\left(r_1 + r_2\right) \qquad r_1 = \text{radius of right end}$$
$$r_2 = \text{radius of left end}$$

and l is the length of the slant of the frustum.

For the frustum on the interval $[x_{i-1}, x_i]$ we have,

$$\begin{aligned} r_1 &= f\left(x_i\right) \\ r_2 &= f\left(x_{i-1}\right) \\ l &= \left|P_{i-1}\ P_i\right| \qquad \text{(length of the line segment connecting } P_i \text{ and } P_{i-1}\text{)} \end{aligned}$$

and we know from the previous section that,

$$\left|P_{i-1}\ P_i\right| = \sqrt{1 + \left[f'\left(x_i^*\right)\right]^2}\ \Delta x \quad \text{where } x_i^* \text{ is some point in } \left[x_{i-1}, x_i\right]$$

Before writing down the formula for the surface area we are going to assume that Δx is "small" and since $f(x)$ is continuous we can then assume that,

$$f\left(x_i\right) \approx f\left(x_i^*\right) \qquad \text{and} \qquad f\left(x_{i-1}\right) \approx f\left(x_i^*\right)$$

So, the surface area of the frustum on the interval $[x_{i-1}, x_i]$ is approximately,

$$\begin{aligned} A_i &= 2\pi \left(\frac{f(x_i) + f(x_{i-1})}{2} \right) |P_{i-1}\, P_i| \\ &\approx 2\pi f(x_i^*) \sqrt{1 + \left[f'(x_i^*) \right]^2}\ \Delta x \end{aligned}$$

The surface area of the whole solid is then approximately,

$$S \approx \sum_{i=1}^{n} 2\pi f(x_i^*) \sqrt{1 + \left[f'(x_i^*) \right]^2}\ \Delta x$$

and we can get the exact surface area by taking the limit as n goes to infinity.

$$\begin{aligned} S &= \lim_{n \to \infty} \sum_{i=1}^{n} 2\pi f(x_i^*) \sqrt{1 + \left[f'(x_i^*) \right]^2}\ \Delta x \\ &= \int_a^b 2\pi f(x) \sqrt{1 + \left[f'(x) \right]^2}\, dx \end{aligned}$$

If we wanted to we could also derive a similar formula for rotating $x = h(y)$ on $[c, d]$ about the y-axis. This would give the following formula.

$$S = \int_c^d 2\pi\, h(y) \sqrt{1 + \left[h'(y) \right]^2}\, dy$$

These are not the "standard" formulas however. Notice that the roots in both of these formulas are nothing more than the two ds's we used in the previous section. Also, we will replace $f(x)$ with y and $h(y)$ with x. Doing this gives the following two formulas for the surface area.

Surface Area Formulas

$$S = \int 2\pi y\, ds \qquad \text{rotation about } x - \text{axis}$$

$$S = \int 2\pi x\, ds \qquad \text{rotation about } y - \text{axis}$$

where,

$$ds = \sqrt{1 + \left(\frac{dy}{dx} \right)^2}\, dx \qquad \text{if } y = f(x),\ a \le x \le b$$

$$ds = \sqrt{1 + \left(\frac{dx}{dy} \right)^2}\, dy \qquad \text{if } x = h(y),\ c \le y \le d$$

There are a couple of things to note about these formulas. First, notice that the variable in the integral itself is always the opposite variable from the one we're rotating about. Second, we are allowed to use either ds in either formula. This means that there are, in some way, four formulas here. We will choose the ds based upon which is the most convenient for a given function and problem.

Now let's work a couple of examples.

Example 1

Determine the surface area of the solid obtained by rotating $y = \sqrt{9 - x^2}$, $-2 \le x \le 2$ about the x-axis.

Solution

The formula that we'll be using here is,

$$S = \int 2\pi y \, ds$$

since we are rotating about the x-axis and we'll use the first ds in this case because our function is in the correct form for that ds and we won't gain anything by solving it for x.

Let's first get the derivative and the root taken care of.

$$\frac{dy}{dx} = \frac{1}{2}\left(9 - x^2\right)^{-\frac{1}{2}}(-2x) = -\frac{x}{\left(9 - x^2\right)^{\frac{1}{2}}}$$

$$\sqrt{1 + \left(\frac{dy}{dx}\right)^2} = \sqrt{1 + \frac{x^2}{9 - x^2}} = \sqrt{\frac{9}{9 - x^2}} = \frac{3}{\sqrt{9 - x^2}}$$

Here's the integral for the surface area,

$$S = \int_{-2}^{2} 2\pi y \frac{3}{\sqrt{9 - x^2}} \, dx$$

There is a problem however. The dx means that we shouldn't have any y's in the integral. So, before evaluating the integral we'll need to substitute in for y as well.

The surface area is then,

$$\begin{aligned} S &= \int_{-2}^{2} 2\pi \sqrt{9 - x^2} \frac{3}{\sqrt{9 - x^2}} \, dx \\ &= \int_{-2}^{2} 6\pi \, dx \\ &= 24\pi \end{aligned}$$

Previously we made the comment that we could use either ds in the surface area formulas. Let's work an example in which using either ds won't create integrals that are too difficult to evaluate and so we can check both ds's.

Example 2

Determine the surface area of the solid obtained by rotating $y = \sqrt[3]{x}$, $1 \le y \le 2$ about the y-axis. Use both ds's to compute the surface area.

Solution

Note that we've been given the function set up for the first ds and limits that work for the second ds.

Solution 1

This solution will use the first ds listed above. We'll start with the derivative and root.

$$\frac{dy}{dx} = \frac{1}{3}x^{-\frac{2}{3}}$$

$$\sqrt{1 + \left(\frac{dy}{dx}\right)^2} = \sqrt{1 + \frac{1}{9x^{\frac{4}{3}}}} = \sqrt{\frac{9x^{\frac{4}{3}} + 1}{9x^{\frac{4}{3}}}} = \frac{\sqrt{9x^{\frac{4}{3}} + 1}}{3x^{\frac{2}{3}}}$$

We'll also need to get new limits. That isn't too bad however. All we need to do is plug in the given y's into our equation and solve to get that the range of x's is $1 \le x \le 8$. The integral for the surface area is then,

$$\begin{aligned} S &= \int_1^8 2\pi x \frac{\sqrt{9x^{\frac{4}{3}} + 1}}{3x^{\frac{2}{3}}}\, dx \\ &= \frac{2\pi}{3}\int_1^8 x^{\frac{1}{3}}\sqrt{9x^{\frac{4}{3}} + 1}\, dx \end{aligned}$$

Note that this time we didn't need to substitute in for the x as we did in the previous example. In this case we picked up a dx from the ds and so we don't need to do a substitution for the x. In fact, if we had substituted for x we would have put y's into the integral which would have caused problems.

Using the substitution

$$u = 9x^{\frac{4}{3}} + 1 \qquad\qquad du = 12x^{\frac{1}{3}}\, dx$$

the integral becomes,

$$\begin{aligned} S &= \frac{\pi}{18}\int_{10}^{145} \sqrt{u}\, du \\ &= \frac{\pi}{27}u^{\frac{3}{2}}\Big|_{10}^{145} \\ &= \frac{\pi}{27}\left(145^{\frac{3}{2}} - 10^{\frac{3}{2}}\right) = 199.48 \end{aligned}$$

Solution 2

This time we'll use the second ds. So, we'll first need to solve the equation for x. We'll also go ahead and get the derivative and root while we're at it.

$$x = y^3 \qquad\qquad \frac{dx}{dy} = 3y^2$$

$$\sqrt{1 + \left(\frac{dx}{dy}\right)^2} = \sqrt{1 + 9y^4}$$

The surface area is then,

$$S = \int_1^2 2\pi x \sqrt{1 + 9y^4}\, dy$$

We used the original y limits this time because we picked up a dy from the ds. Also note that the presence of the dy means that this time, unlike the first solution, we'll need to substitute in for the x. Doing that gives,

$$\begin{aligned} S &= \int_1^2 2\pi y^3 \sqrt{1 + 9y^4}\, dy \qquad\qquad u = 1 + 9y^4 \\ &= \frac{\pi}{18} \int_{10}^{145} \sqrt{u}\, du \\ &= \frac{\pi}{27} \left(145^{\frac{3}{2}} - 10^{\frac{3}{2}}\right) = 199.48 \end{aligned}$$

Note that after the substitution the integral was identical to the first solution and so the work was skipped.

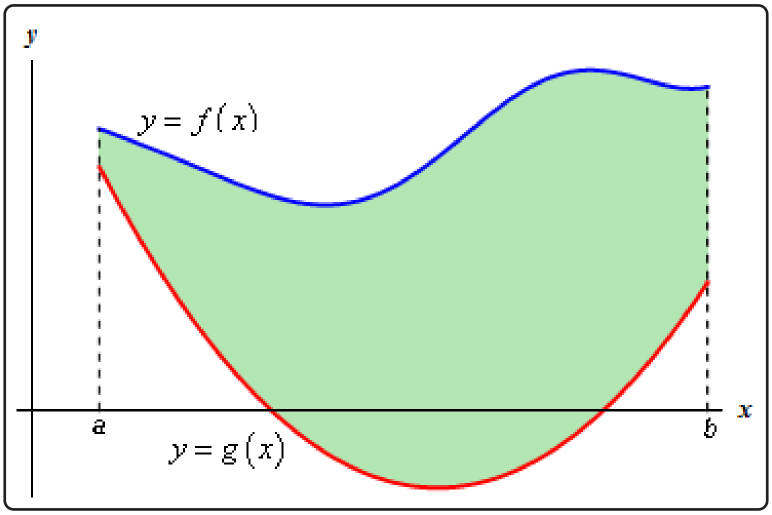
As this example has shown we can use either ds to get the surface area. It is important to point out as well that with one ds we had to do a substitution for the x and with the other we didn't. This will always work out that way.

Note as well that in the case of the last example it was just as easy to use either ds. That often won't be the case. In many examples only one of the ds will be convenient to work with so we'll always need to determine which ds is liable to be the easiest to work with before starting the problem.

8.3 Center Of Mass

In this section we are going to find the **center of mass** or **centroid** of a thin plate with uniform density ρ. The center of mass or centroid of a region is the point in which the region will be perfectly balanced horizontally if suspended from that point.

So, let's suppose that the plate is the region bounded by the two curves $f(x)$ and $g(x)$ on the interval $[a, b]$. So, we want to find the center of mass of the region below.



We'll first need the mass of this plate. The mass is,

$$\begin{aligned} M &= \rho\,(\text{Area of plate}) \\ &= \rho \int_a^b f(x) - g(x)\, dx \end{aligned}$$

Next, we'll need the **moments** of the region. There are two moments, denoted by M_x and M_y. The moments measure the tendency of the region to rotate about the x and y-axis respectively. The moments are given by,

Equations of Moments

$$M_x = \rho \int_a^b \frac{1}{2}\left(\left[f(x)\right]^2 - \left[g(x)\right]^2\right) dx$$
$$M_y = \rho \int_a^b x\big(f(x) - g(x)\big)\, dx$$

The coordinates of the center of mass, $(\overline{x}, \overline{y})$, are then,

Center of Mass Coordinates

$$\overline{x} = \frac{M_y}{M} = \frac{\int_a^b x\left(f(x) - g(x)\right)\,dx}{\int_a^b f(x) - g(x)\,dx} = \frac{1}{A}\int_a^b x\left(f(x) - g(x)\right)\,dx$$

$$\overline{y} = \frac{M_x}{M} = \frac{\int_a^b \frac{1}{2}\left([f(x)]^2 - [g(x)]^2\right)\,dx}{\int_a^b f(x) - g(x)\,dx} = \frac{1}{A}\int_a^b \frac{1}{2}\left([f(x)]^2 - [g(x)]^2\right)\,dx$$

where,

$$A = \int_a^b f(x) - g(x)\,dx$$

Note that the density, ρ, of the plate cancels out and so isn't really needed.

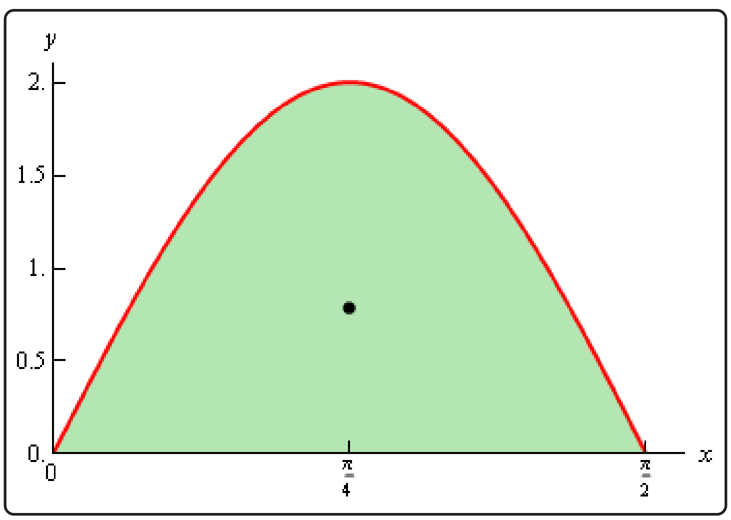
Let's work a couple of examples.

Example 1

Determine the center of mass for the region bounded by $y = 2\sin(2x)$, $y = 0$ on the interval $\left[0, \frac{\pi}{2}\right]$.

Solution

Here is a sketch of the region with the center of mass denoted with a dot.



Let's first get the area of the region.

$$\begin{aligned} A &= \int_0^{\frac{\pi}{2}} 2\sin(2x)\ dx \\ &= -\cos(2x)\Big|_0^{\frac{\pi}{2}} \\ &= 2 \end{aligned}$$

Now, the moments (without density since it will just drop out) are,

$$\begin{aligned} M_x &= \int_0^{\frac{\pi}{2}} 2\sin^2(2x)\ dx & M_y &= \int_0^{\frac{\pi}{2}} 2x\sin(2x)\ dx \quad \text{integrating by parts...} \\ &= \int_0^{\frac{\pi}{2}} 1-\cos(4x)\ dx & &= -x\cos(2x)\Big|_0^{\frac{\pi}{2}} + \int_0^{\frac{\pi}{2}} \cos(2x)\ dx \\ &= \left(x - \frac{1}{4}\sin(4x)\right)\Big|_0^{\frac{\pi}{2}} & &= -x\cos(2x)\Big|_0^{\frac{\pi}{2}} + \frac{1}{2}\sin(2x)\Big|_0^{\frac{\pi}{2}} \\ &= \frac{\pi}{2} & &= \frac{\pi}{2} \end{aligned}$$

The coordinates of the center of mass are then,

$$\begin{aligned} \overline{x} &= \frac{{}^{\pi}/_{2}}{2} = \frac{\pi}{4} \\ \overline{y} &= \frac{{}^{\pi}/_{2}}{2} = \frac{\pi}{4} \end{aligned}$$

Again, note that we didn't put in the density since it will cancel out.

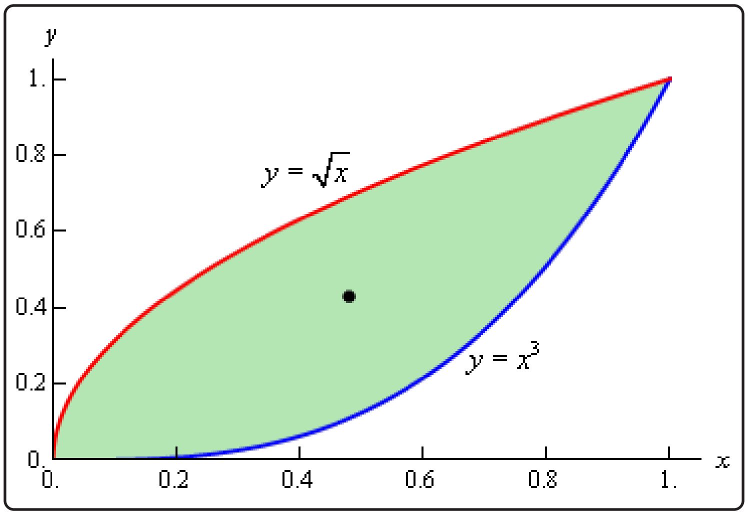
So, the center of mass for this region is $\left(\frac{\pi}{4}, \frac{\pi}{4}\right)$.

Example 2

Determine the center of mass for the region bounded by $y = x^3$ and $y = \sqrt{x}$.

Solution

The two curves intersect at $x = 0$ and $x = 1$ and here is a sketch of the region with the center of mass marked with a box.



We'll first get the area of the region.

$$\begin{aligned} A &= \int_0^1 \sqrt{x} - x^3 \, dx \\ &= \left(\frac{2}{3}x^{\frac{3}{2}} - \frac{1}{4}x^4 \right) \Big|_0^1 \\ &= \frac{5}{12} \end{aligned}$$

Now the moments, again without density, are

$$\begin{aligned} M_x &= \int_0^1 \frac{1}{2} \left(x - x^6 \right) \, dx \\ &= \frac{1}{2} \left(\frac{1}{2}x^2 - \frac{1}{7}x^7 \right) \Big|_0^1 \\ &= \frac{5}{28} \end{aligned}$$

$$\begin{aligned} M_y &= \int_0^1 x \left(\sqrt{x} - x^3 \right) \, dx \\ &= \int_0^1 x^{\frac{3}{2}} - x^4 \, dx \\ &= \left(\frac{2}{5}x^{\frac{5}{2}} - \frac{1}{5}x^5 \right) \Big|_0^1 \\ &= \frac{1}{5} \end{aligned}$$

The coordinates of the center of mass is then,

$$\begin{aligned} \overline{x} &= \frac{1/5}{5/12} = \frac{12}{25} \\ \overline{y} &= \frac{5/28}{5/12} = \frac{3}{7} \end{aligned}$$

The coordinates of the center of mass are then, $\left(\frac{12}{25}, \frac{3}{7}\right)$.

8.4 Hydrostatic Pressure and Force

In this section we are going to submerge a vertical plate in water and we want to know the force that is exerted on the plate due to the pressure of the water. This force is often called the hydrostatic force.

There are two basic formulas that we'll be using here. First, if we are d meters below the surface then the hydrostatic pressure is given by,

$$P = \rho g d$$

where, ρ is the density of the fluid and g is the gravitational acceleration. We are going to assume that the fluid in question is water and since we are going to be using the metric system these quantities become,

$$\rho = 1000 \text{ kg/m}^3 \qquad\qquad g = 9.81 \text{ m/s}^2$$

The second formula that we need is the following. Assume that a constant pressure P is acting on a surface with area A. Then the hydrostatic force that acts on the area is,

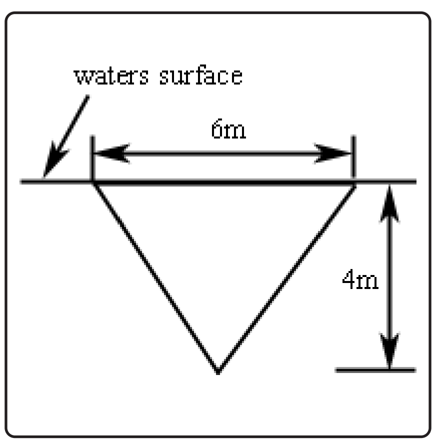
$$F = PA$$

Note that we won't be able to find the hydrostatic force on a vertical plate using this formula since the pressure will vary with depth and hence will not be constant as required by this formula. We will however need this for our work.

The best way to see how these problems work is to do an example or two.

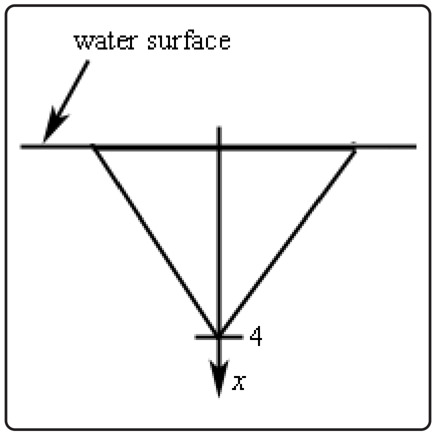
Example 1

Determine the hydrostatic force on the following triangular plate that is submerged in water as shown.



Solution

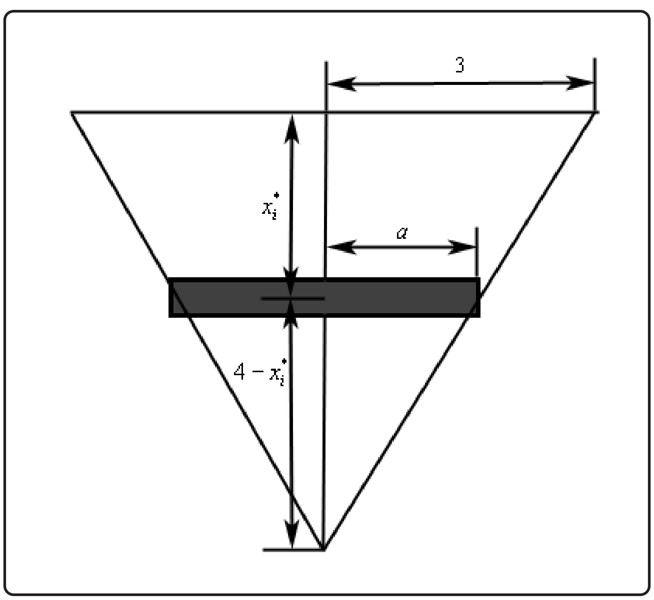
The first thing to do here is set up an axis system. So, let's redo the sketch above with the following axis system added in.



So, we are going to orient the x-axis so that positive x is downward, $x = 0$ corresponds to the water surface and $x = 4$ corresponds to the depth of the tip of the triangle.

Next we break up the triangle into n horizontal strips each of equal width Δx and in each interval $[x_{i-1}, x_i]$ choose any point x_i^*. In order to make the computations easier we are going to make two assumptions about these strips. First, we will ignore the fact that the ends are actually going to be slanted and assume the strips are rectangular. If Δx is sufficiently small this will not affect our computations much. Second, we will assume that Δx is small enough that the hydrostatic pressure on each strip is essentially constant.

Below is a representative strip.



The height of this strip is Δx and the width is $2a$. We can use similar triangles to determine a as follows,

$$\frac{3}{4} = \frac{a}{4 - x_i^*} \qquad \Rightarrow \qquad a = 3 - \frac{3}{4}x_i^*$$

Now, since we are assuming the pressure on this strip is constant, the pressure is given by,

$$P_i = \rho g d = 1000\,(9.81)\,x_i^* = 9810x_i^*$$

and the hydrostatic force on each strip is,

$$F_i = P_i\,A = P_i\,(2a\Delta x) = 9810x_i^*\,(2)\left(3 - \frac{3}{4}x_i^*\right)\Delta x = 19620x_i^*\left(3 - \frac{3}{4}x_i^*\right)\,\Delta x$$

The approximate hydrostatic force on the plate is then the sum of the forces on all the strips or,

$$F \approx \sum_{i=1}^{n} 19620x_i^*\left(3 - \frac{3}{4}x_i^*\right)\Delta x$$

Taking the limit will get the exact hydrostatic force,

$$F = \lim_{n\to\infty}\sum_{i=1}^{n} 19620x_i^*\left(3 - \frac{3}{4}x_i^*\right)\Delta x$$

Using the definition of the definite integral this is nothing more than,

$$F = \int_0^4 19620\left(3x - \frac{3}{4}x^2\right)\,dx$$

The hydrostatic force is then,

$$\begin{aligned} F &= \int_0^4 19620\left(3x - \frac{3}{4}x^2\right)\,dx \\ &= 19620\left(\frac{3}{2}x^2 - \frac{1}{4}x^3\right)\bigg|_0^4 \\ &= 156960\,N \end{aligned}$$

Let’s take a look at another example.

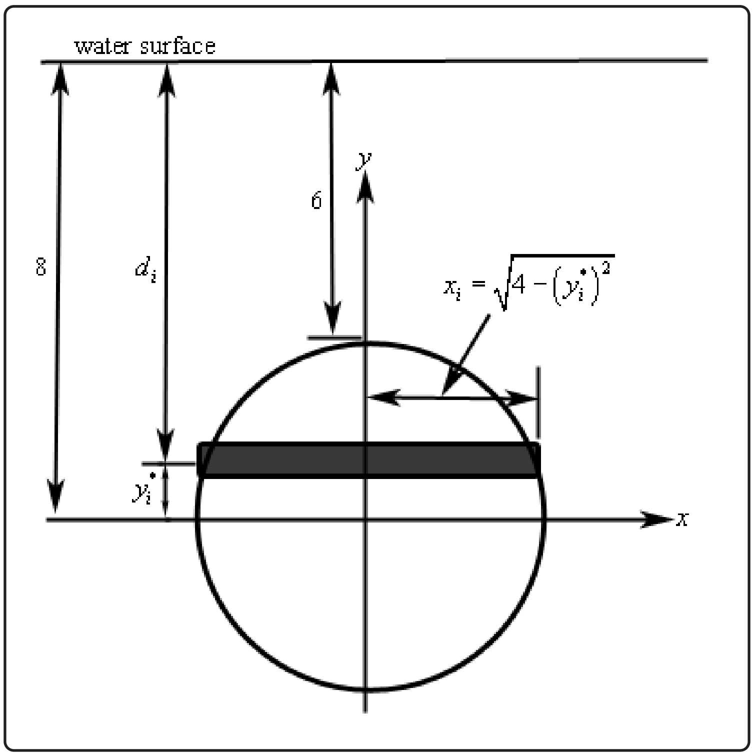
Example 2

Find the hydrostatic force on a circular plate of radius 2 that is submerged 6 meters in the water.

Solution

First, we're going to assume that the top of the circular plate is 6 meters under the water. Next, we will set up the axis system so that the origin of the axis system is at the center of the plate. Setting the axis system up in this way will greatly simplify our work.

Finally, we will again split up the plate into n horizontal strips each of width Δy and we'll choose a point y_i^* from each strip. We'll also assume that the strips are rectangular again to help with the computations. Here is a sketch of the setup.



The depth below the water surface of each strip is,

$$d_i = 8 - y_i^*$$

and that in turn gives us the pressure on the strip,

$$P_i = \rho g d_i = 9810\,(8 - y_i^*)$$

The area of each strip is,

$$A_i = 2\sqrt{4 - (y_i^*)^2}\ \Delta y$$

The hydrostatic force on each strip is,

$$F_i = P_i A_i = 9810\,(8 - y_i^*)\,(2)\,\sqrt{4 - (y_i^*)^2}\;\Delta y$$

The total force on the plate is,

$$\begin{aligned} F &= \lim_{n \to \infty} \sum_{i=1}^{n} 19620\,(8 - y_i^*)\,\sqrt{4 - (y_i^*)^2}\;\Delta y \\ &= 19620 \int_{-2}^{2} (8 - y)\,\sqrt{4 - y^2}\,dy \end{aligned}$$

To do this integral we'll need to split it up into two integrals.

$$F = 19620 \int_{-2}^{2} 8\sqrt{4 - y^2}\,dy - 19620 \int_{-2}^{2} y\sqrt{4 - y^2}\,dy$$

The first integral requires the trig substitution $y = 2\sin(\theta)$ and the second integral needs the substitution $v = 4 - y^2$. After using these substitutions we get,

$$\begin{aligned} F &= 627840 \int_{-\pi/2}^{\pi/2} \cos^2(\theta)\,d\theta + 9810 \int_{0}^{0} \sqrt{v}\,dv \\ &= 313920 \int_{-\pi/2}^{\pi/2} 1 + \cos(2\theta)\;d\theta + 0 \\ &= 313920 \left(\theta + \frac{1}{2}\sin(2\theta)\right)\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \\ &= 313920\pi \end{aligned}$$

Note that after the substitution we know the second integral will be zero because the upper and lower limit is the same.

8.5 Probability

In this last application of integrals that we'll be looking at we're going to look at probability. Before actually getting into the applications we need to get a couple of definitions out of the way.

Suppose that we wanted to look at the age of a person, the height of a person, the amount of time spent waiting in line, or maybe the lifetime of a battery. Each of these quantities have values that will range over an interval of real numbers. Because of this these are called **continuous random variables**. Continuous random variables are often represented by X.

Every continuous random variable, X, has a **probability density function**, $f(x)$. Probability density functions satisfy the following conditions.

1. $f(x) \geq 0$ for all x.
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$

Probability density functions can be used to determine the probability that a continuous random variable lies between two values, say a and b. This probability is denoted by $P(a \leq X \leq b)$ and is given by,

Fact

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

Let's take a look at an example of this.

Example 1

Let $f(x) = \frac{x^3}{5000}(10 - x)$ for $0 \leq x \leq 10$ and $f(x) = 0$ for all other values of x. Answer each of the following questions about this function.

(a) Show that $f(x)$ is a probability density function.

(b) Find $P(1 \leq X \leq 4)$

(c) Find $P(x \geq 6)$

Solution

(a) Show that $f(x)$ is a probability density function.

First note that in the range $0 \leq x \leq 10$ is clearly positive and outside of this range we've defined it to be zero.

So, to show this is a probability density function we'll need to show that $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\, dx &= \int_{0}^{10} \frac{x^3}{5000}(10-x)\, dx \\ &= \left.\left(\frac{x^4}{2000} - \frac{x^5}{25000}\right)\right|_{0}^{10} \\ &= 1\end{aligned}$$

Note the change in limits on the integral. The function is only non-zero in these ranges and so the integral can be reduced down to only the interval where the function is not zero.

(b) Find $P(1 \le X \le 4)$

In this case we need to evaluate the following integral.

$$\begin{aligned}P(1 \le X \le 4) &= \int_{1}^{4} \frac{x^3}{5000}(10-x)\, dx \\ &= \left.\left(\frac{x^4}{2000} - \frac{x^5}{25000}\right)\right|_{1}^{4} \\ &= 0.08658\end{aligned}$$

So the probability of X being between 1 and 4 is 8.658

(c) Find $P(x \ge 6)$

Note that in this case $P(x \ge 6)$ is equivalent to $P(6 \le X \le 10)$ since 10 is the largest value that X can be. So the probability that X is greater than or equal to 6 is,

$$\begin{aligned}P(X \ge 6) &= \int_{6}^{10} \frac{x^3}{5000}(10-x)\, dx \\ &= \left.\left(\frac{x^4}{2000} - \frac{x^5}{25000}\right)\right|_{6}^{10} \\ &= 0.66304\end{aligned}$$

This probability is then 66.304%.

Probability density functions can also be used to determine the mean of a continuous random variable. The mean is given by,

Fact

$$\mu = \int_{-\infty}^{\infty} x f(x)\, dx$$

Let's work one more example.

Example 2

It has been determined that the probability density function for the wait in line at a counter is given by,

$$f(t) = \begin{cases} 0 & \text{if } t < 0 \\ 0.1\mathbf{e}^{-\frac{t}{10}} & \text{if } t \geq 0 \end{cases}$$

(a) Verify that this is in fact a probability density function.

(b) Determine the probability that a person will wait in line for at least 6 minutes.

(c) Determine the mean wait in line.

Solution

(a) Verify that this is in fact a probability density function.

This function is clearly positive or zero and so there's not much to do here other than compute the integral.

$$\begin{aligned} \int_{-\infty}^{\infty} f(t)\, dt &= \int_{0}^{\infty} 0.1\mathbf{e}^{-\frac{t}{10}}\, dt \\ &= \lim_{u\to\infty} \int_{0}^{u} 0.1\mathbf{e}^{-\frac{t}{10}}\, dt \\ &= \lim_{u\to\infty} \left(-\mathbf{e}^{-\frac{t}{10}}\right)\Big|_{0}^{u} \\ &= \lim_{u\to\infty} \left(1 - \mathbf{e}^{-\frac{u}{10}}\right) = 1 \end{aligned}$$

So it is a probability density function.

(b) Determine the probability that a person will wait in line for at least 6 minutes.

The probability that we're looking for here is $P(X \geq 6)$.

$$\begin{aligned} P(X \geq 6) &= \int_6^\infty 0.1\mathbf{e}^{-\frac{t}{10}}\,dt \\ &= \lim_{u\to\infty}\int_6^u 0.1\mathbf{e}^{-\frac{t}{10}}\,dt \\ &= \lim_{u\to\infty}\left(-\mathbf{e}^{-\frac{t}{10}}\right)\Big|_6^u \\ &= \lim_{u\to\infty}\left(\mathbf{e}^{-\frac{6}{10}} - \mathbf{e}^{-\frac{u}{10}}\right) = \mathbf{e}^{-\frac{3}{5}} = 0.548812 \end{aligned}$$

So the probability that a person will wait in line for more than 6 minutes is 54.8811%.

(c) Determine the mean wait in line.

Here's the mean wait time.

$$\begin{aligned} \mu &= \int_{-\infty}^{\infty} t\,f(t)\,dt \\ &= \int_0^\infty 0.1t\,\mathbf{e}^{-\frac{t}{10}}\,dt \\ &= \lim_{u\to\infty}\int_0^u 0.1t\,\mathbf{e}^{-\frac{t}{10}}\,dt \qquad \text{integrating by parts....} \\ &= \lim_{u\to\infty}\left(-(t+10)\,\mathbf{e}^{-\frac{t}{10}}\right)\Big|_0^u \\ &= \lim_{u\to\infty}\left(10-(u+10)\,\mathbf{e}^{-\frac{u}{10}}\right) = 10 \end{aligned}$$

So, it looks like the average wait time is 10 minutes.

9 Parametric Equations and Polar Coordinates

We are now going to take a look at a couple of topics that are completely different from anything we've seen to this point. That does not mean, however, that we can just forget everything that we've seen to this point. As we will see before too long we will still need to be able to do a large part of the material (both Calculus I and Calculus II material) that we've looked at to this point.

The first major topic that we'll look at in this chapter will be that of Parametric Equations. Parametric Equations will allow us to work with and perform Calculus operations on equations that cannot be (easily) solved into the form $y = f(x)$ or $x = h(y)$ (assuming we are using x and y as our variables). Also, as we'll see we can write some equations that can be solved for y or x as a set of easier to work with parametric equations.

Once we've got an idea of what parametric equations are and how to sketch graphs of them we will revisit some of the Calculus topics we've looked at to this point. Specifically we'll take a look at how to use only parametric equations to get the equation of tangent lines, where the graph is increasing/decreasing and the concavity of the graph. In addition, we'll revisit the idea of using a definite integral to find the area between the graph of a set of parametric equation and the x-axis. We will close out the Calculus topics by discussing arc length and surface area for a set of parametric equations.

We will then move into the other major topic of this chapter, namely Polar Coordinates. Once we've defined polar coordinates and gotten comfortable with them we will, again, revisit the same Calculus topics we looked at in terms of parametric equations only now we will look at how to work them in terms of polar coordinates.

On the surface it will appear that polar coordinates has nothing in common with parametric equations. We will see however that several topics in Polar Coordinates can be easily done, in some way, if we first set them up in terms of parametric equations.

In addition, we should point out that the purpose of the topics in this chapter is in preparation for multi-variable Calculus (*i.e.* the material that is usually taught in Calculus III). As we will see when we get to that point there are a lot of topics that involve and/or require parametric equations. In addition, polar coordinates will pop up every so often so keep that in mind as we go through this stuff. It is easy sometimes to get the idea that the topics in this chapter don't have a lot of use but once we hit multi-variable Calculus they will start to pop up with some regularity.

9.1 Parametric Equations and Curves

To this point (in both Calculus I and Calculus II) we've looked almost exclusively at functions in the form $y = f(x)$ or $x = h(y)$ and almost all of the formulas that we've developed require that functions be in one of these two forms. The problem is that not all curves or equations that we'd like to look at fall easily into this form.

Take, for example, a circle. It is easy enough to write down the equation of a circle centered at the origin with radius r.

$$x^2 + y^2 = r^2$$

However, we will never be able to write the equation of a circle down as a single equation in either of the forms above. Sure we can solve for x or y as the following two formulas show

$$y = \pm\sqrt{r^2 - x^2} \qquad\qquad x = \pm\sqrt{r^2 - y^2}$$

but there are in fact two functions in each of these. Each formula gives a portion of the circle.

$$\begin{array}{llll} y = \sqrt{r^2 - x^2} & \text{(top)} & x = \sqrt{r^2 - y^2} & \text{(right side)} \\ y = -\sqrt{r^2 - x^2} & \text{(bottom)} & x = -\sqrt{r^2 - y^2} & \text{(left side)} \end{array}$$

Unfortunately, we usually are working on the whole circle, or simply can't say that we're going to be working only on one portion of it. Even if we can narrow things down to only one of these portions the function is still often fairly unpleasant to work with.

There are also a great many curves out there that we can't even write down as a single equation in terms of only x and y. So, to deal with some of these problems we introduce **parametric equations**. Instead of defining y in terms of x ($y = f(x)$) or x in terms of y ($x = h(y)$) we define both x and y in terms of a third variable called a parameter as follows,

$$x = f(t) \qquad\qquad y = g(t)$$

This third variable is usually denoted by t (as we did here) but doesn't have to be of course. Sometimes we will restrict the values of t that we'll use and at other times we won't. This will often be dependent on the problem and just what we are attempting to do.

Each value of t defines a point $(x, y) = (f(t), g(t))$ that we can plot. The collection of points that we get by letting t be all possible values is the graph of the parametric equations and is called the **parametric curve**.

To help visualize just what a parametric curve is pretend that we have a big tank of water that is in constant motion and we drop a ping pong ball into the tank. The point $(x, y) = (f(t), g(t))$ will then represent the location of the ping pong ball in the tank at time t and the parametric curve will be a trace of all the locations of the ping pong ball. Note that this is not always a correct analogy but it is useful initially to help visualize just what a parametric curve is.

Sketching a parametric curve is not always an easy thing to do. Let's take a look at an example to see one way of sketching a parametric curve. This example will also illustrate why this method is usually not the best.

Example 1

Sketch the parametric curve for the following set of parametric equations.

$$x = t^2 + t \qquad y = 2t - 1$$

Solution

At this point our only option for sketching a parametric curve is to pick values of t, plug them into the parametric equations and then plot the points. So, let's plug in some t's.

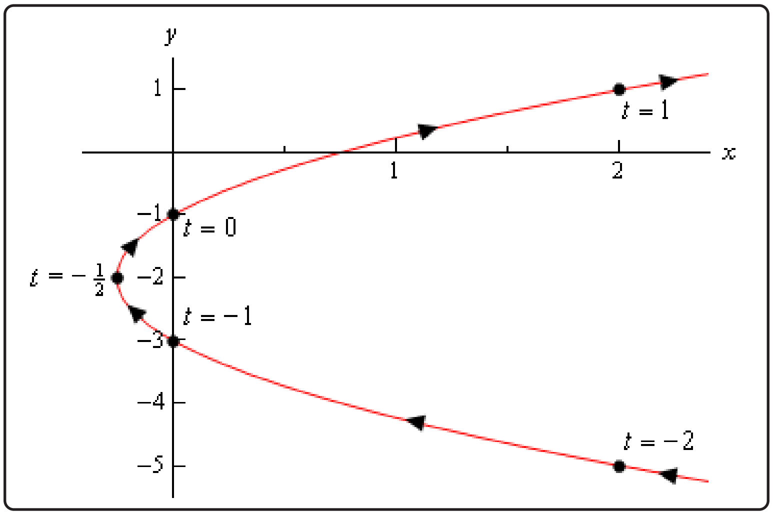
t	x	y
-2	2	-5
-1	0	-3
$-\frac{1}{2}$	$-\frac{1}{4}$	-2
0	0	-1
1	2	1

The first question that should be asked at this point is, how did we know to use the values of t that we did, especially the third choice? Unfortunately, there is no real answer to this question at this point. We simply pick t's until we are fairly confident that we've got a good idea of what the curve looks like. It is this problem with picking "good" values of t that make this method of sketching parametric curves one of the poorer choices. Sometimes we have no choice, but if we do have a choice we should avoid it.

We'll discuss an alternate graphing method in later examples that will help to explain how these values of t were chosen.

We have one more idea to discuss before we actually sketch the curve. Parametric curves have a **direction of motion**. The direction of motion is given by increasing t. So, when plotting parametric curves, we also include arrows that show the direction of motion. We will often give the value of t that gave specific points on the graph as well to make it clear the value of t that gave that particular point.

Here is the sketch of this parametric curve.



So, it looks like we have a parabola that opens to the right.

Before we end this example there is a somewhat important and subtle point that we need to discuss first. Notice that we made sure to include a portion of the sketch to the right of the points corresponding to $t = -2$ and $t = 1$ to indicate that there are portions of the sketch there. Had we simply stopped the sketch at those points we are indicating that there was no portion of the curve to the right of those points and there clearly will be. We just didn't compute any of those points.

This may seem like an unimportant point, but as we'll see in the next example it's more important than we might think.

Before addressing a much easier way to sketch this graph let's first address the issue of limits on the parameter. In the previous example we didn't have any limits on the parameter. Without limits on the parameter the graph will continue in both directions as shown in the sketch above.

We will often have limits on the parameter however and this will affect the sketch of the parametric equations. To see this effect let's look a slight variation of the previous example.

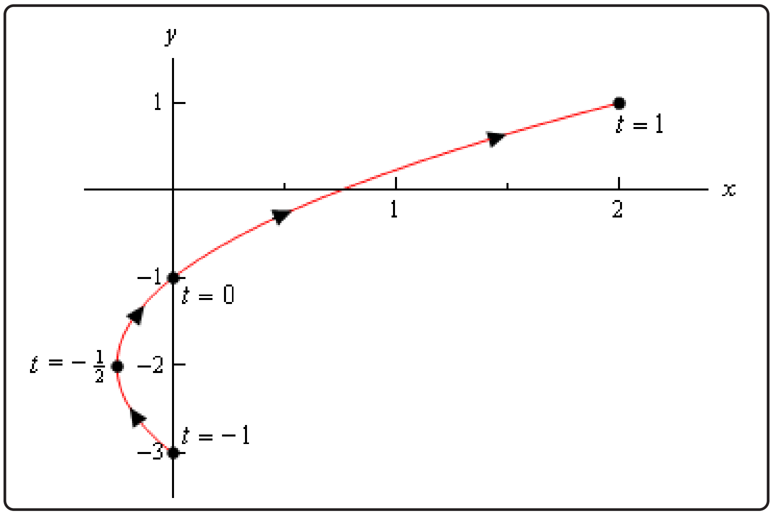
Example 2

Sketch the parametric curve for the following set of parametric equations.

$$x = t^2 + t \qquad y = 2t - 1 \qquad -1 \le t \le 1$$

Solution

Note that the only difference here is the presence of the limits on t. All these limits do is tell us that we can't take any value of t outside of this range. Therefore, the parametric curve will only be a portion of the curve above. Here is the parametric curve for this example.



Notice that with this sketch we started and stopped the sketch right on the points originating from the end points of the range of t's. Contrast this with the sketch in the previous example where we had a portion of the sketch to the right of the "start" and "end" points that we computed.

In this case the curve starts at $t = -1$ and ends at $t = 1$, whereas in the previous example the curve didn't really start at the right most points that we computed. We need to be clear in our sketches if the curve starts/ends right at a point, or if that point was simply the first/last one that we computed.

It is now time to take a look at an easier method of sketching this parametric curve. This method uses the fact that in many, but not all, cases we can actually eliminate the parameter from the parametric equations and get a function involving only x and y. We will sometimes call this the *algebraic equation* to differentiate it from the original parametric equations. There will be two small problems with this method, but it will be easy to address those problems. It is important to note however that we won't always be able to do this.

Just how we eliminate the parameter will depend upon the parametric equations that we've got. Let's see how to eliminate the parameter for the set of parametric equations that we've been working with to this point.

Example 3

Eliminate the parameter from the following set of parametric equations.

$$x = t^2 + t \qquad y = 2t - 1$$

Solution

One of the easiest ways to eliminate the parameter is to simply solve one of the equations for the parameter (t, in this case) and substitute that into the other equation. Note that while this may be the easiest to eliminate the parameter, it's usually not the best way as we'll see soon enough.

In this case we can easily solve y for t.

$$t = \frac{1}{2}(y+1)$$

Plugging this into the equation for x gives the following algebraic equation,

$$x = \left(\frac{1}{2}(y+1)\right)^2 + \frac{1}{2}(y+1) = \frac{1}{4}y^2 + y + \frac{3}{4}$$

Sure enough from our Algebra knowledge we can see that this is a parabola that opens to the right and will have a vertex at $\left(-\frac{1}{4}, -2\right)$.

We won't bother with a sketch for this one as we've already sketched this once and the point here was more to eliminate the parameter anyway.

Before we leave this example let's address one quick issue.

In the first example we just, seemingly randomly, picked values of t to use in our table, especially the third value. There really was no apparent reason for choosing $t = -\frac{1}{2}$. It is however probably the most important choice of t as it is the one that gives the vertex.

The reality is that when writing this material up we actually did this problem first then went back and did the first problem. Plotting points is generally the way most people first learn how to construct graphs and it does illustrate some important concepts, such as direction, so it made sense to do that first in the notes. In practice however, this example is often done first.

So, how did we get those values of t? Well let's start off with the vertex as that is probably the most important point on the graph. We have the x and y coordinates of the vertex and we also have x and y parametric equations for those coordinates. So, plug in the coordinates for the vertex into the parametric equations and solve for t. Doing this gives,

$$\begin{aligned} -\tfrac{1}{4} &= t^2 + t \\ -2 &= 2t - 1 \end{aligned} \qquad \Rightarrow \qquad \begin{aligned} t &= -\tfrac{1}{2} \quad \text{(double root)} \\ t &= -\tfrac{1}{2} \end{aligned}$$

So, as we can see, the value of t that will give both of these coordinates is $t = -\frac{1}{2}$. Note that the x parametric equation gave a double root and this will often not happen. Often we would have gotten two distinct roots from that equation. In fact, it won't be unusual to get multiple values of t from each of the equations.

However, what we can say is that there will be a value(s) of t that occurs in both sets of solutions and that is the t that we want for that point. We'll eventually see an example where this happens in a later section.

Now, from this work we can see that if we use $t = -\frac{1}{2}$ we will get the vertex and so we included that value of t in the table in Example 1. Once we had that value of t we chose two integer values of t on either side to finish out the table.

As we will see in later examples in this section determining values of t that will give specific points is something that we'll need to do on a fairly regular basis. It is fairly simple however as this example has shown. All we need to be able to do is solve a (usually) fairly basic equation which by this point in time shouldn't be too difficult.

Getting a sketch of the parametric curve once we've eliminated the parameter seems fairly simple. All we need to do is graph the equation that we found by eliminating the parameter. As noted already however, there are two small problems with this method. The first is direction of motion. The equation involving only x and y will NOT give the direction of motion of the parametric curve. This is generally an easy problem to fix however. Let's take a quick look at the derivatives of the parametric equations from the last example. They are,

$$\frac{dx}{dt} = 2t + 1$$
$$\frac{dy}{dt} = 2$$

Now, all we need to do is recall our Calculus I knowledge. The derivative of y with respect to t is clearly always positive. Recalling that one of the interpretations of the first derivative is rate of change we now know that as t increases y must also increase. Therefore, we must be moving up the curve from bottom to top as t increases as that is the only direction that will always give an increasing y as t increases.

Note that the x derivative isn't as useful for this analysis as it will be both positive and negative and hence x will be both increasing and decreasing depending on the value of t. That doesn't help with direction much as following the curve in either direction will exhibit both increasing and decreasing x.

In some cases, only one of the equations, such as this example, will give the direction while in other cases either one could be used. It is also possible that, in some cases, both derivatives would be needed to determine direction. It will always be dependent on the individual set of parametric equations.

The second problem with eliminating the parameter is best illustrated in an example as we'll be running into this problem in the remaining examples.

Example 4

Sketch the parametric curve for the following set of parametric equations. Clearly indicate direction of motion.

$$x = 5\cos(t) \qquad y = 2\sin(t) \qquad 0 \le t \le 2\pi$$

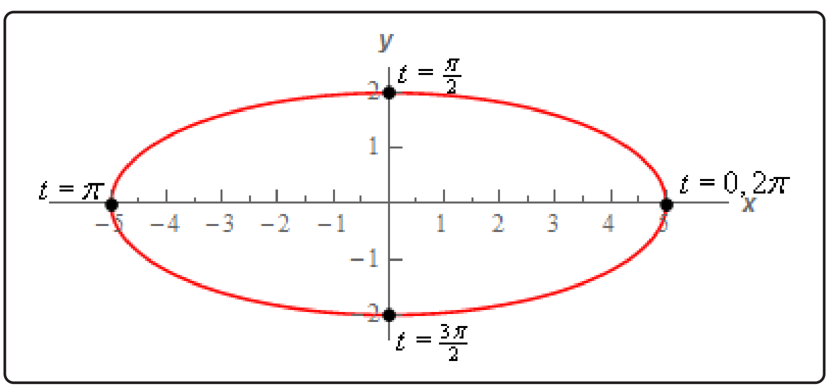
Solution

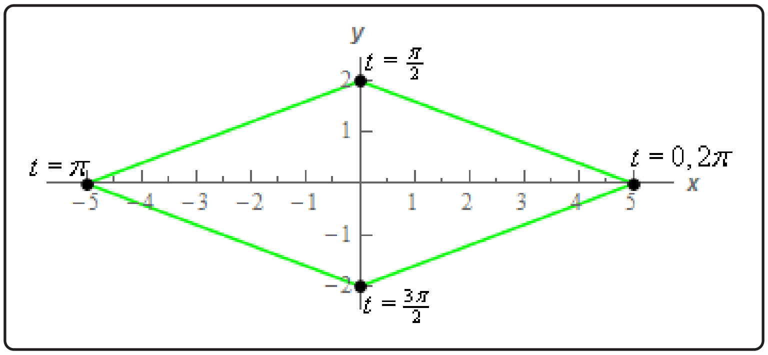
Before we proceed with eliminating the parameter for this problem let's first address again why just picking t's and plotting points is not really a good idea.

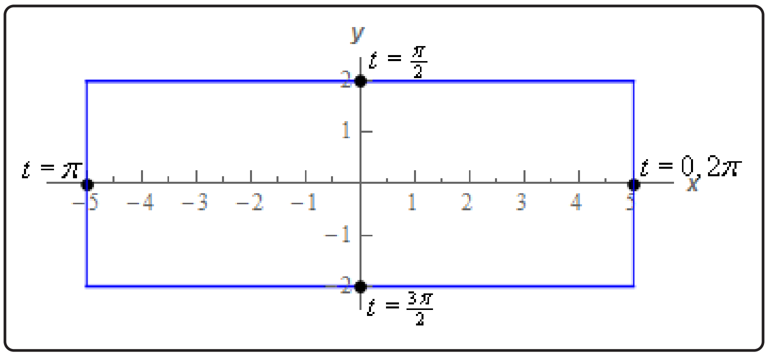
Given the range of t's in the problem statement let's use the following set of t's.

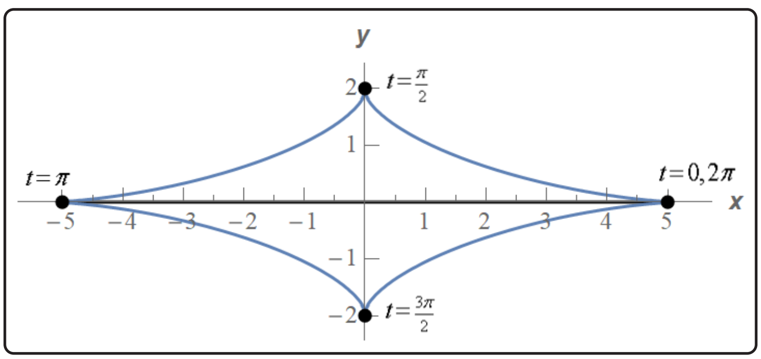
t	x	y
0	5	0
$\frac{\pi}{2}$	0	2
π	-5	0
$\frac{3\pi}{2}$	0	-2
2π	5	0

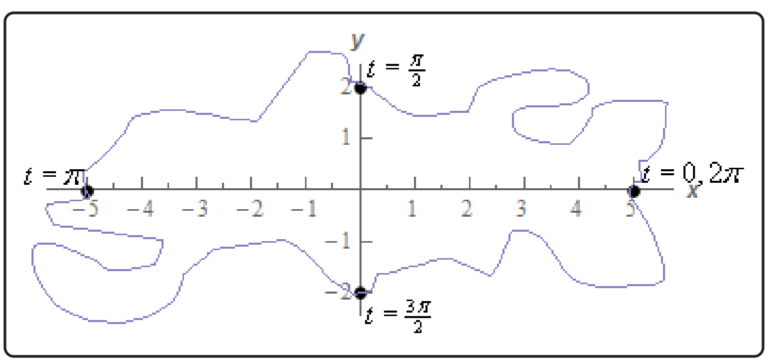
The question that we need to ask now is do we have enough points to accurately sketch the graph of this set of parametric equations? Below are some sketches of some possible graphs of the parametric equation based only on these five points.




y
$t = \frac{\pi}{2}$
$t = \pi$
$t = 0, 2\pi$
x
$t = \frac{3\pi}{2}$


y
$t = \frac{\pi}{2}$
$t = \pi$
$t = 0, 2\pi$
x
$t = \frac{3\pi}{2}$


y
$t=\frac{\pi}{2}$
$t=\pi$
$t=0,2\pi$
x
$t=\frac{3\pi}{2}$


y
$t = \frac{\pi}{2}$
$t = \pi$
$t = 0, 2\pi$
x
$t = \frac{3\pi}{2}$

Given the nature of sine/cosine you might be tempted to eliminate the diamond and the square but there is no denying that they are graphs that go through the given points. The first and third graphs both have some curvature to them and so you might be tempted to assume that one of those is the correct one given the sine/cosine in the equations. The last graph is also a little silly but it does show a graph going through the given points.

Again, given the nature of sine/cosine you are probably guessing that the correct graph is the the first or third graph. However, that is all that would be at this point. A guess. Nothing actually says unequivocally that the parametric curve is an will be one of those two just from those five points. That is the danger of sketching parametric curves based on a handful of points. Unless we know what the graph will be ahead of time we are really just making a guess.

It is important to note at this point that it is very easy to construct a set of parametric equations both containing sines and/or cosines and yet have the graph not have any curvature at all. You can often make some guesses as to the shape of the curve from the parametric equations but you won't always guess correctly unfortunately. Care must be taken when graphing parametric equations to not take the behavior of the individual parametric equations and just assume that behavior will translate to the curve of the set of parametric equations.

Also, in general, we should avoid plotting points to sketch parametric curves as that will, on occasion, lead to incorrect graphs. The best method, provided it can be done, is to eliminate the parameter. As noted just prior to starting this example there is still a potential problem with eliminating the parameter that we'll need to deal with. We will eventually discuss this issue. For now, let's just proceed with eliminating the parameter.

We'll start by eliminating the parameter as we did in the previous section. We'll solve one of the of the equations for t and plug this into the other equation. For example, we could do the following,

$$t = \cos^{-1}\left(\frac{x}{5}\right) \qquad \Rightarrow \qquad y = 2\sin\left(\cos^{-1}\left(\frac{x}{5}\right)\right)$$

Can you see the problem with doing this? This is definitely easy to do but we have a greater chance of correctly graphing the original parametric equations by plotting points than we do graphing this!

There are many ways to eliminate the parameter from the parametric equations and solving for t is usually not the best way to do it. While it is often easy to do we will, in most cases, end up with an equation that is almost impossible to deal with.

So, how can we eliminate the parameter here? In this case all we need to do is recall a very nice trig identity and the equation of an ellipse. Recall,

$$\cos^2(t) + \sin^2(t) = 1$$

Then from the parametric equations we get,

$$\cos(t) = \frac{x}{5} \qquad \sin(t) = \frac{y}{2}$$

Then, using the trig identity from above and these equations we get,

$$1 = \cos^2(t) + \sin^2(t) = \left(\frac{x}{5}\right)^2 + \left(\frac{y}{2}\right)^2 = \frac{x^2}{25} + \frac{y^2}{4}$$

So we now know that we will have an ellipse.

Now, let's continue on with the example. We've identified that the parametric equations describe an ellipse, but we can't just sketch an ellipse and be done with it.

First, just because the algebraic equation was an ellipse doesn't actually mean that the parametric curve is the full ellipse. It is always possible that the parametric curve is only a portion of the ellipse. In order to identify just how much of the ellipse the parametric curve will cover let's go back to the parametric equations and see what they tell us about any limits on x and y. Based on our knowledge of sine and cosine we have the following,

$$\begin{aligned} -1 \le \cos(t) \le 1 \quad &\Rightarrow \quad -5 \le 5\cos(t) \le 5 \quad \Rightarrow \quad -5 \le x \le 5 \\ -1 \le \sin(t) \le 1 \quad &\Rightarrow \quad -2 \le 2\sin(t) \le 2 \quad \Rightarrow \quad -2 \le y \le 2 \end{aligned}$$

So, by starting with sine/cosine and "building up" the equation for x and y using basic algebraic manipulations we get that the parametric equations enforce the above limits on x and y. In this case, these also happen to be the full limits on x and y we get by graphing the full ellipse.

This is the second potential issue alluded to above. The parametric curve may not always trace out the full graph of the algebraic curve. We should always find limits on x and y enforced upon us by the parametric curve to determine just how much of the algebraic curve is actually sketched out by the parametric equations.

Therefore, in this case, we now know that we get a full ellipse from the parametric equations. Before we proceed with the rest of the example be careful to not always just assume we will get the full graph of the algebraic equation. There are definitely times when we will not get the full graph and we'll need to do a similar analysis to determine just how much of the graph we actually get. We'll see an example of this later.

Note as well that any limits on t given in the problem statement can also affect how much of the graph of the algebraic equation we get. In this case however, based on the table of values we computed at the start of the problem we can see that we do indeed get the full ellipse in the range $0 \le t \le 2\pi$. That won't always be the case however, so pay attention to any restrictions on t that might exist!

Next, we need to determine a direction of motion for the parametric curve. Recall that all parametric curves have a direction of motion and the equation of the ellipse simply tells us nothing about the direction of motion.

To get the direction of motion it is tempting to just use the table of values we computed above to get the direction of motion. In this case, we would guess (and yes that is all it is - a guess) that the curve traces out in a counter-clockwise direction. We'd be correct. In this case, we'd be correct! The problem is that tables of values can be misleading when determining a direction of motion as we'll see in the next example.

Therefore, it is best to not use a table of values to determine the direction of motion. To correctly determine the direction of motion we'll use the same method of determining the direction that we discussed after Example 3. In other words, we'll take the derivative of the parametric equations and use our knowledge of Calculus I and trig to determine the direction of motion.

The derivatives of the parametric equations are,

$$\frac{dx}{dt} = -5\sin(t) \qquad\qquad \frac{dy}{dt} = 2\cos(t)$$

Now, at $t = 0$ we are at the point $(5, 0)$ and let's see what happens if we start increasing t. Let's increase t from $t = 0$ to $t = \frac{\pi}{2}$. In this range of t's we know that sine is always positive and so from the derivative of the x equation we can see that x must be decreasing in this range of t's.

This, however, doesn't really help us determine a direction for the parametric curve. Starting at $(5, 0)$ no matter if we move in a clockwise or counter-clockwise direction x will have to decrease so we haven't really learned anything from the x derivative.

The derivative from the y parametric equation on the other hand will help us. Again, as we increase t from $t = 0$ to $t = \frac{\pi}{2}$ we know that cosine will be positive and so y must be increasing in this range. That however, can only happen if we are moving in a counter-clockwise direction. If we were moving in a clockwise direction from the point $(5, 0)$ we can see that y would have to decrease!

Therefore, in the first quadrant we must be moving in a counter-clockwise direction. Let's move on to the second quadrant.

So, we are now at the point $(0, 2)$ and we will increase t from $t = \frac{\pi}{2}$ to $t = \pi$. In this range of t we know that cosine will be negative and sine will be positive. Therefore, from the derivatives of the parametric equations we can see that x is still decreasing and y will now be decreasing as well.

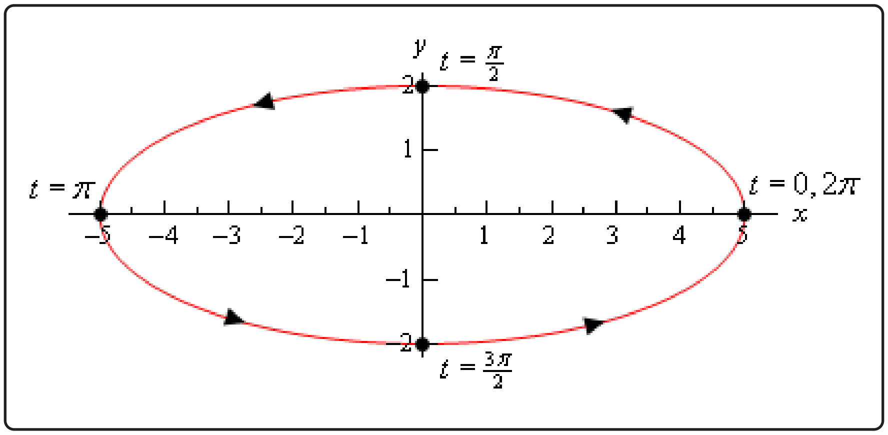
In this quadrant the y derivative tells us nothing as y simply must decrease to move from $(0, 2)$. However, in order for x to decrease, as we know it does in this quadrant, the direction must still be moving a counter-clockwise rotation.

We are now at $(-5, 0)$ and we will increase t from $t = \pi$ to $t = \frac{3\pi}{2}$. In this range of t we know that cosine is negative (and hence y will be decreasing) and sine is also negative (and hence x will be increasing). Therefore, we will continue to move in a counter-clockwise motion.

For the 4^{th} quadrant we will start at $(0, -2)$ and increase t from $t = \frac{3\pi}{2}$ to $t = 2\pi$. In this range of t we know that cosine is positive (and hence y will be increasing) and sine is negative (and hence x will be increasing). So, as in the previous three quadrants, we continue to move in a counter-clockwise motion.

At this point we covered the range of t's we were given in the problem statement and during the full range the motion was in a counter-clockwise direction.

We can now fully sketch the parametric curve so, here is the sketch.



Okay, that was a really long example. Most of these types of problems aren't as long. We just had a lot to discuss in this one so we could get a couple of important ideas out of the way. The rest of the examples in this section shouldn't take as long to go through.

Now, let's take a look at another example that will illustrate an important idea about parametric equations.

Example 5

Sketch the parametric curve for the following set of parametric equations. Clearly indicate direction of motion.

$$x = 5\cos(3t) \qquad y = 2\sin(3t) \qquad 0 \le t \le 2\pi$$

Solution

Note that the only difference in between these parametric equations and those in Example 4 is that we replaced the t with 3 t. We can eliminate the parameter here in the same manner as we did in the previous example.

$$\cos(3t) = \frac{x}{5} \qquad \sin(3t) = \frac{y}{2}$$

We then get,

$$1 = \cos^2(3t) + \sin^2(3t) = \left(\frac{x}{5}\right)^2 + \left(\frac{y}{2}\right)^2 = \frac{x^2}{25} + \frac{y^2}{4}$$

So, we get the same ellipse that we did in the previous example. Also note that we can do the same analysis on the parametric equations to determine that we have exactly the same limits on x and y. Namely,

$$-5 \le x \le 5 \qquad\qquad -2 \le y \le 2$$

It's starting to look like changing the t into a $3t$ in the trig equations will not change the parametric curve in any way. That is not correct however. The curve does change in a small but important way which we will be discussing shortly.

Before discussing that small change the $3t$ brings to the curve let's discuss the direction of motion for this curve. Despite the fact that we said in the last example that picking values of t and plugging in to the equations to find points to plot is a bad idea let's do it any way.

Given the range of t's from the problem statement the following set looks like a good choice of t's to use.

t	x	y
0	5	0
$\frac{\pi}{2}$	0	-2
π	-5	0
$\frac{3\pi}{2}$	0	2
2π	5	0

So, the only change to this table of values/points from the last example is all the nonzero y values changed sign. From a quick glance at the values in this table it would look like the curve, in this case, is moving in a clockwise direction. But is that correct? Recall we said that these tables of values can be misleading when used to determine direction and that's why we don't use them.

Let's see if our first impression is correct. We can check our first impression by doing the derivative work to get the correct direction. Let's work with just the y parametric equation as the x will have the same issue that it had in the previous example. The derivative of the y parametric equation is,

$$\frac{dy}{dt} = 6\cos(3t)$$

Now, if we start at $t = 0$ as we did in the previous example and start increasing t. At $t = 0$ the derivative is clearly positive and so increasing t (at least initially) will force y to also be increasing. The only way for this to happen is if the curve is in fact tracing out in a counter-clockwise direction initially.

Now, we could continue to look at what happens as we further increase t, but when dealing with a parametric curve that is a full ellipse (as this one is) and the argument of the trig functions is of the form *nt* for any constant n the direction will not change so once we know the initial direction we know that it will always move in that direction. Note that this is only true for parametric equations in the form that we have here. We'll see in later examples that for different kinds of parametric equations this may no longer be true.

Okay, from this analysis we can see that the curve must be traced out in a counter-clockwise direction. This is directly counter to our guess from the tables of values above and so we can see that, in this case, the table would probably have led us to the wrong direction. So, once again, tables are generally not very reliable for getting pretty much any real information about a parametric curve other than a few points that must be on the curve. Outside of that the tables are rarely useful and will generally not be dealt with in further examples.

So, why did our table give an incorrect impression about the direction? Well recall that we mentioned earlier that the $3t$ will lead to a small but important change to the curve versus just a t? Let's take a look at just what that change is as it will also answer what "went wrong" with our table of values.

Let's start by looking at $t = 0$. At $t = 0$ we are at the point $(5, 0)$ and let's ask ourselves what values of t put us back at this point. We saw in Example 3 how to determine value(s) of t that put us at certain points and the same process will work here with a minor modification.

Instead of looking at both the x and y equations as we did in that example let's just look at the x equation. The reason for this is that we'll note that there are two points on the ellipse that will have a y coordinate of zero, $(5, 0)$ and $(-5, 0)$. If we set the y coordinate equal to zero we'll find all the t's that are at both of these points when we only want the values of t that are at $(5, 0)$.

So, because the x coordinate of five will only occur at this point we can simply use the x parametric equation to determine the values of t that will put us at this point. Doing this gives the following equation and solution,

$$\begin{aligned} 5 &= 5\cos(3t) \\ 3t &= \cos^{-1}(1) = 0 + 2\pi n \qquad \to \qquad t = \frac{2}{3}\pi n \quad n = 0, \pm1, \pm2, \pm3, \ldots \end{aligned}$$

Don't forget that when solving a trig equation we need to add on the "$+2\pi n$" where n represents the number of full revolutions in the counter-clockwise direction (positive n) and clockwise direction (negative n) that we rotate from the first solution to get all possible solutions to the equation.

Now, let's plug in a few values of n starting at $n = 0$. We don't need negative n in this case since all of those would result in negative t and those fall outside of the range of t's we were given in the problem statement. The first few values of t are then,

$$\begin{aligned} n = 0 \quad &: \quad t = 0 \\ n = 1 \quad &: \quad t = \frac{2\pi}{3} \\ n = 2 \quad &: \quad t = \frac{4\pi}{3} \\ n = 3 \quad &: \quad t = \frac{6\pi}{3} = 2\pi \end{aligned}$$

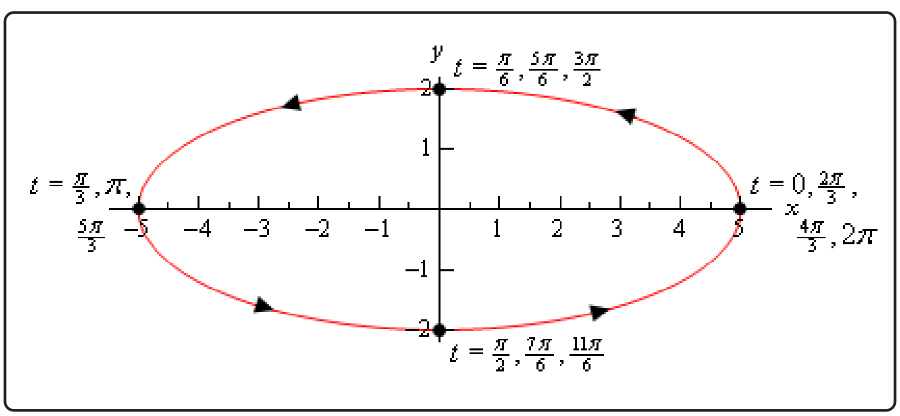
We can stop here as all further values of t will be outside the range of t's given in this problem.

So, what is this telling us? Well back in Example 4 when the argument was just t the ellipse was traced out exactly once in the range $0 \le t \le 2\pi$. However, when we change the argument to 3 t (and recalling that the curve will always be traced out in a counter-clockwise direction for this problem) we are going through the "starting" point of $(5, 0)$ two more times than we did in the previous example.

In fact, this curve is tracing out three separate times. The first trace is completed in the range $0 \le t \le \frac{2\pi}{3}$. The second trace is completed in the range $\frac{2\pi}{3} \le t \le \frac{4\pi}{3}$ and the third and final trace is completed in the range $\frac{4\pi}{3} \le t \le 2\pi$. In other words, changing the argument from t to 3 t increase the speed of the trace and the curve will now trace out three times in the range $0 \le t \le 2\pi$!

This is why the table gives the wrong impression. The speed of the tracing has increased leading to an incorrect impression from the points in the table. The table seems to suggest that between each pair of values of t a quarter of the ellipse is traced out in the clockwise direction when in reality it is tracing out three quarters of the ellipse in the counter-clockwise direction.

Here's a final sketch of the curve and note that it really isn't all that different from the previous sketch. The only differences are the values of t and the various points we included. We did include a few more values of t at various points just to illustrate where the curve is at for various values of t but in general these really aren't needed.



So, we saw in the last two examples two sets of parametric equations that in some way gave the same graph. Yet, because they traced out the graph a different number of times we really do need to think of them as different parametric curves at least in some manner. This may seem like a difference that we don't need to worry about, but as we will see in later sections this can be a very important difference. In some of the later sections we are going to need a curve that is traced out exactly once.

Before we move on to other problems let's briefly acknowledge what happens by changing the t to an nt in these kinds of parametric equations. When we are dealing with parametric equations involving only sines and cosines and they both have the same argument if we change the argument from t to nt we simply change the speed with which the curve is traced out. If $n > 1$ we will increase the speed and if $n < 1$ we will decrease the speed.

Let's take a look at a couple more examples.

Example 6

Sketch the parametric curve for the following set of parametric equations. Clearly identify the direction of motion. If the curve is traced out more than once give a range of the parameter for which the curve will trace out exactly once.

$$x = \sin^2(t) \qquad y = 2\cos(t)$$

Solution

We can eliminate the parameter much as we did in the previous two examples. However, we'll need to note that the x already contains a $\sin^2(t)$ and so we won't need to square the x. We will however, need to square the y as we need in the previous two examples.

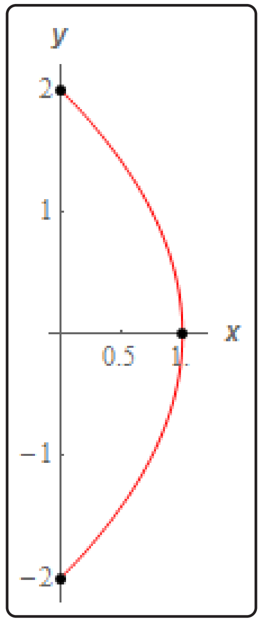
$$x + \frac{y^2}{4} = \sin^2(t) + \cos^2(t) = 1 \qquad \Rightarrow \qquad x = 1 - \frac{y^2}{4}$$

In this case the algebraic equation is a parabola that opens to the left.

We will need to be very, very careful however in sketching this parametric curve. We will NOT get the whole parabola. A sketch of the algebraic form parabola will exist for all possible values of y. However, the parametric equations have defined both x and y in terms of sine and cosine and we know that the ranges of these are limited and so we won't get all possible values of x and y here. So, first let's get limits on x and y as we did in previous examples. Doing this gives,

$$\begin{array}{lclcl} -1 \le \sin(t) \le 1 & \Rightarrow & 0 \le \sin^2(t) \le 1 & \Rightarrow & 0 \le x \le 1 \\ -1 \le \cos(t) \le 1 & \Rightarrow & -2 \le 2\cos(t) \le 2 & \Rightarrow & -2 \le y \le 2 \end{array}$$

So, it is clear from this that we will only get a portion of the parabola that is defined by the algebraic equation. Below is a quick sketch of the portion of the parabola that the parametric curve will cover.



To finish the sketch of the parametric curve we also need the direction of motion for the curve. Before we get to that however, let's jump forward and determine the range of t's for one trace. To do this we'll need to know the t's that put us at each end point and we can follow the same procedure we used in the previous example. The only difference is this time let's use the y parametric equation instead of the x because the y coordinates of the two end points of the curve are different whereas the x coordinates are the same.

So, for the top point we have,

$$\begin{aligned} 2 &= 2\cos(t) \\ t &= \cos^{-1}(1) = 0 + 2\pi n = 2\pi n, \qquad n = 0, \pm 1, \pm 2, \pm 3, \ldots \end{aligned}$$

For, plugging in some values of n we get that the curve will be at the top point at,

$$t = \ldots, -4\pi, -2\pi, 0, 2\pi, 4\pi, \ldots$$

Similarly, for the bottom point we have,

$$\begin{aligned} -2 &= 2\cos(t) \\ t &= \cos^{-1}(-1) = \pi + 2\pi n, \qquad n = 0, \pm 1, \pm 2, \pm 3, \ldots \end{aligned}$$

So, we see that we will be at the bottom point at,

$$t = \ldots, -3\pi, -\pi, \pi, 3\pi, \ldots$$

So, if we start at say, $t = 0$, we are at the top point and we increase t we have to move along the curve downwards until we reach $t = \pi$ at which point we are now at the bottom point. This means that we will trace out the curve exactly once in the range $0 \le t \le \pi$.

This is not the only range that will trace out the curve however. Note that if we further increase t from $t = \pi$ we will now have to travel back up the curve until we reach $t = 2\pi$ and we are now back at the top point. Increasing t again until we reach $t = 3\pi$ will take us back down the curve until we reach the bottom point again, *etc*. From this analysis we can get two more ranges of t for one trace,

$$\pi \le t \le 2\pi \qquad\qquad 2\pi \le t \le 3\pi$$

As you can probably see there are an infinite number of ranges of t we could use for one trace of the curve. Any of them would be acceptable answers for this problem.

Note that in the process of determining a range of t's for one trace we also managed to determine the direction of motion for this curve. In the range $0 \le t \le \pi$ we had to travel downwards along the curve to get from the top point at $t = 0$ to the bottom point at $t = \pi$. However, at $t = 2\pi$ we are back at the top point on the curve and to get there we must travel along the path. We can't just jump back up to the top point or take a different path to get there. All travel must be done on the path sketched out. This means that we had to go back up the path. Further increasing t takes us back down the path, then up the path again *etc*.

In other words, this path is sketched out in both directions because we are not putting any restrictions on the t's and so we have to assume we are using all possible values of t. If we had put restrictions on which t's to use we might really have ended up only moving in one direction. That however would be a result only of the range of t's we are using and not the parametric equations themselves.

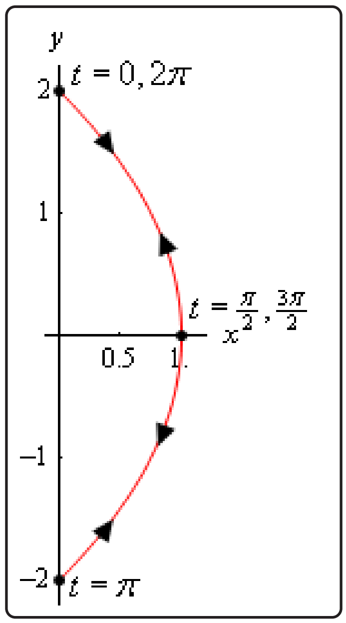
Note that we didn't really need to do the above work to determine that the curve traces out in both directions.in this case. Both the x and y parametric equations involve sine or cosine and we know both of those functions oscillate. This, in turn means that both x and y will oscillate as well. The only way for that to happen on this particular this curve will be for the curve to be traced out in both directions.

Be careful with the above reasoning that the oscillatory nature of sine/cosine forces the curve to be traced out in both directions. It can only be used in this example because the "starting" point and "ending" point of the curves are in different places. The only way to get from one of the "end" points on the curve to the other is to travel back along the curve in the opposite direction.

Contrast this with the ellipse in Example 4. In that case we had sine/cosine in the parametric equations as well. However, the curve only traced out in one direction, not in both directions. In Example 4 we were graphing the full ellipse and so no matter where we start sketching the graph we will eventually get back to the "starting" point without ever retracing any portion of the graph. In Example 4 as we trace out the full ellipse both x and y do in fact oscillate between their two "endpoints" but the curve itself does not trace out in both directions for this to happen.

Basically, we can only use the oscillatory nature of sine/cosine to determine that the curve traces out in both directions if the curve starts and ends at different points. If the starting/ending point is the same then we generally need to go through the full derivative argument to determine the actual direction of motion.

So, to finish this problem out, below is a sketch of the parametric curve. Note that we put direction arrows in both directions to clearly indicate that it would be traced out in both directions. We also put in a few values of t just to help illustrate the direction of motion.



To this point we've seen examples that would trace out the complete graph that we got by eliminating the parameter if we took a large enough range of t's. However, in the previous example we've now seen that this will not always be the case. It is more than possible to have a set of parametric equations which will continuously trace out just a portion of the curve. We can usually determine if this will happen by looking for limits on x and y that are imposed up us by the parametric equation.

We will often use parametric equations to describe the path of an object or particle. Let's take a look at an example of that.

Example 7

The path of a particle is given by the following set of parametric equations.

$$x = 3\cos(2t) \qquad y = 1 + \cos^2(2t)$$

Completely describe the path of this particle. Do this by sketching the path, determining limits on x and y and giving a range of t's for which the path will be traced out exactly once (provide it traces out more than once of course).

Solution

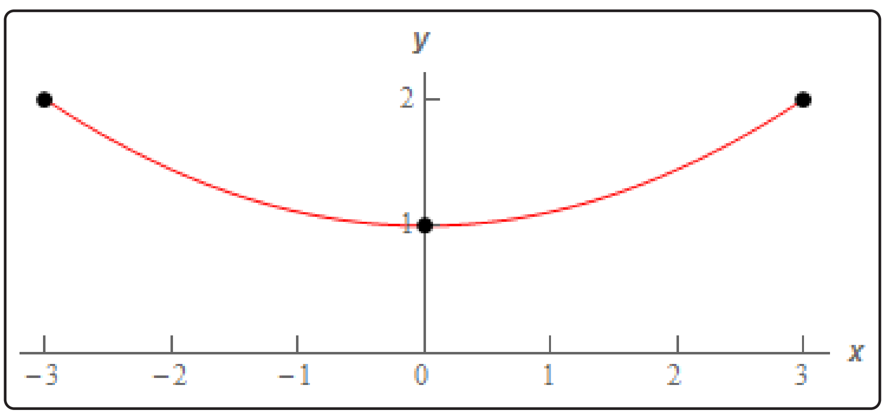
Eliminating the parameter this time will be a little different. We only have cosines this time and we'll use that to our advantage. We can solve the x equation for cosine and plug that into the equation for y. This gives,

$$\cos(2t) = \frac{x}{3} \qquad y = 1 + \left(\frac{x}{3}\right)^2 = 1 + \frac{x^2}{9}$$

This time the algebraic equation is a parabola that opens upward. We also have the following limits on x and y.

$$\begin{array}{ccc} -1 \le \cos(2t) \le 1 & -3 \le 3\cos(2t) \le 3 & -3 \le x \le 3 \\ 0 \le \cos^2(2t) \le 1 & 1 \le 1 + \cos^2(2t) \le 2 & 1 \le y \le 2 \end{array}$$

So, again we only trace out a portion of the curve. Here is a quick sketch of the portion of the parabola that the parametric curve will cover.



Now, as we discussed in the previous example because both the x and y parametric equations involve cosine we know that both x and y must oscillate and because the "start" and "end" points of the curve are not the same the only way x and y can oscillate is for the curve to trace out in both directions.

To finish the problem then all we need to do is determine a range of t's for one trace. Because the "end" points on the curve have the same y value and different x values we can use the x parametric equation to determine these values. Here is that work.

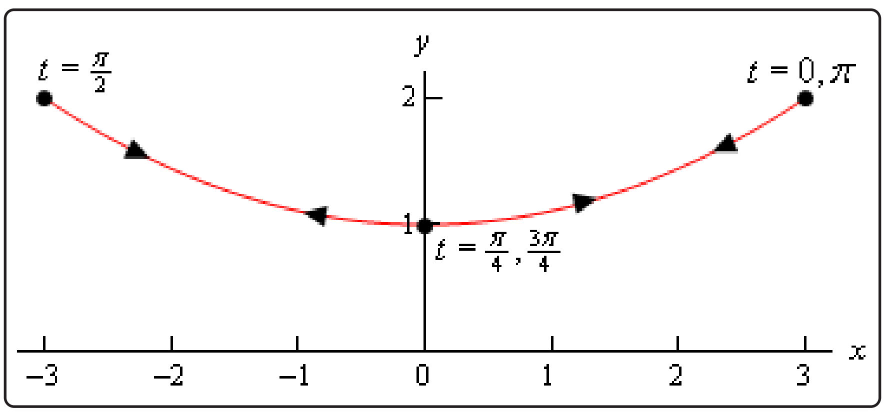
$$\begin{aligned} x = 3: \quad & 3 = 3\cos(2t) \\ & 1 = \cos(2t) \\ & 2t = 0 + 2\pi n \qquad \to \qquad t = \pi n \qquad n = 0, \pm 1, \pm 2, \pm 3, \ldots \end{aligned}$$

$$\begin{aligned} x = -3: \quad & -3 = 3\cos(2t) \\ & -1 = \cos(2t) \\ & 2t = \pi + 2\pi n \qquad \to \qquad t = \frac{1}{2}\pi + \pi n \qquad n = 0, \pm 1, \pm 2, \pm 3, \ldots \end{aligned}$$

So, we will be at the right end point at $t = \ldots, -2\pi, -\pi, 0, \pi, 2\pi, \ldots$ and we'll be at the left end point at $t = \ldots, -\frac{3}{2}\pi, -\frac{1}{2}\pi, \frac{1}{2}\pi, \frac{3}{2}\pi, \ldots$. So, in this case there are an infinite number of ranges of t's for one trace. Here are a few of them.

$$-\frac{1}{2}\pi \le t \le 0 \qquad 0 \le t \le \frac{1}{2}\pi \qquad \frac{1}{2}\pi \le t \le \pi$$

Here is a final sketch of the particle's path with a few values of t on it.



We should give a small warning at this point. Because of the ideas involved in them we concentrated on parametric curves that retraced portions of the curve more than once. Do not, however, get too locked into the idea that this will always happen. Many, if not most parametric curves will only trace out once. The first one we looked at is a good example of this. That parametric curve will never repeat any portion of itself.

There is one final topic to be discussed in this section before moving on. So far we've started with parametric equations and eliminated the parameter to determine the parametric curve.

However, there are times in which we want to go the other way. Given a function or equation we might want to write down a set of parametric equations for it. In these cases we say that we **parameterize** the function.

If we take Examples 4 and 5 as examples we can do this for ellipses (and hence circles). Given the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

a set of parametric equations for it would be,

$$x = a\cos(t) \qquad y = b\sin(t)$$

This set of parametric equations will trace out the ellipse starting at the point $(a, 0)$ and will trace in a counter-clockwise direction and will trace out exactly once in the range $0 \le t \le 2\pi$. This is a fairly important set of parametric equations as it used continually in some subjects with dealing with ellipses and/or circles.

Every curve can be parameterized in more than one way. Any of the following will also parameterize the same ellipse.

$$\begin{array}{ll} x = a\cos(\omega\, t) & y = b\sin(\omega\, t) \\ x = a\sin(\omega\, t) & y = b\cos(\omega\, t) \\ x = a\cos(\omega\, t) & y = -b\sin(\omega\, t) \end{array}$$

The presence of the ω will change the speed that the ellipse rotates as we saw in Example 5. Note as well that the last two will trace out ellipses with a clockwise direction of motion (you might want to verify this). Also note that they won't all start at the same place (if we think of $t = 0$ as the starting point that is).

There are many more parameterizations of an ellipse of course, but you get the idea. It is important to remember that each parameterization will trace out the curve once with a potentially different range of t's. Each parameterization may rotate with different directions of motion and may start at different points.

You may find that you need a parameterization of an ellipse that starts at a particular place and has a particular direction of motion and so you now know that with some work you can write down a set of parametric equations that will give you the behavior that you're after.

Now, let's write down a couple of other important parameterizations and all the comments about direction of motion, starting point, and range of t's for one trace (if applicable) are still true.

First, because a circle is nothing more than a special case of an ellipse we can use the parameterization of an ellipse to get the parametric equations for a circle centered at the origin of radius r as well. One possible way to parameterize a circle is,

$$x = r\cos(t) \qquad y = r\sin(t)$$

Finally, even though there may not seem to be any reason to, we can also parameterize functions in the form $y = f(x)$ or $x = h(y)$. In these cases we parameterize them in the following way,

$$\begin{aligned} x &= t & \qquad x &= h(t) \\ y &= f(t) & \qquad y &= t \end{aligned}$$

At this point it may not seem all that useful to do a parameterization of a function like this, but there are many instances where it will actually be easier, or it may even be required, to work with the parameterization instead of the function itself. Unfortunately, almost all of these instances occur in a Calculus III course.

9.2 Tangents with Parametric Equations

In this section we want to find the tangent lines to the parametric equations given by,

$$x = f(t) \qquad\qquad y = g(t)$$

To do this let's first recall how to find the tangent line to $y = F(x)$ at $x = a$. Here the tangent line is given by,

$$y = F(a) + m(x - a)\,, \text{ where } m = \left.\frac{dy}{dx}\right|_{x=a} = F'(a)$$

Now, notice that if we could figure out how to get the derivative $\frac{dy}{dx}$ from the parametric equations we could simply reuse this formula since we will be able to use the parametric equations to find the x and y coordinates of the point.

So, just for a second let's suppose that we were able to eliminate the parameter from the parametric form and write the parametric equations in the form $y = F(x)$. Now, plug the parametric equations in for x and y. Yes, it seems silly to eliminate the parameter, then immediately put it back in, but it's what we need to do in order to get our hands on the derivative. Doing this gives,

$$g(t) = F\big(f(t)\big)$$

Now, differentiate with respect to t and notice that we'll need to use the Chain Rule on the right-hand side.

$$g'(t) = F'\big(f(t)\big)\, f'(t)$$

Let's do another change in notation. We need to be careful with our derivatives here. Derivatives of the lower case function are with respect to t while derivatives of upper case functions are with respect to x. So, to make sure that we keep this straight let's rewrite things as follows.

$$\frac{dy}{dt} = F'(x)\,\frac{dx}{dt}$$

At this point we should remind ourselves just what we are after. We needed a formula for $\frac{dy}{dx}$ or $F'(x)$ that is in terms of the parametric formulas. Notice however that we can get that from the above equation.

Derivative for Parametric Equations, $\frac{dy}{dx}$

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}}, \text{ provided } \frac{dx}{dt} \neq 0$$

Notice as well that this will be a function of t and not x.

As an aside, notice that we could also get the following formula with a similar derivation if we needed to,

Derivative for Parametric Equations, $\dfrac{dx}{dy}$

$$\frac{dx}{dy} = \frac{\frac{dx}{dt}}{\frac{dy}{dt}}, \text{ provided } \frac{dy}{dt} \neq 0$$

Why would we want to do this? Well, recall that in the arc length section of the Applications of Integral section we actually needed this derivative on occasion.

So, let's find a tangent line.

Example 1

Find the tangent line(s) to the parametric curve given by

$$x = t^5 - 4t^3 \qquad\qquad y = t^2$$

at $(0, 4)$.

Solution

Note that there is apparently the potential for more than one tangent line here! We will look into this more after we're done with the example.

The first thing that we should do is find the derivative so we can get the slope of the tangent line.

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{2t}{5t^4 - 12t^2} = \frac{2}{5t^3 - 12t}$$

At this point we've got a small problem. The derivative is in terms of t and all we've got is an *x-y* coordinate pair. The next step then is to determine the value(s) of t which will give this point. We find these by plugging the x and y values into the parametric equations and solving for t.

$$\begin{aligned} 0 &= t^5 - 4t^3 = t^3\left(t^2 - 4\right) && \Rightarrow && t = 0, \pm 2 \\ 4 &= t^2 && \Rightarrow && t = \pm 2 \end{aligned}$$

Any value of t which appears in both lists will give the point. So, since there are two values of t that give the point we will in fact get two tangent lines. That's definitely not something that happened back in Calculus I and we're going to need to look into this a little more. However, before we do that let's actually get the tangent lines.

$\underline{t = -2:}$

Since we already know the x and y-coordinates of the point all that we need to do is find the slope of the tangent line.

$$m = \left.\frac{dy}{dx}\right|_{t=-2} = -\frac{1}{8}$$

The tangent line (at $t = -2$) is then,

$$y = 4 - \frac{1}{8}x$$

$\underline{t = 2:}$

Again, all we need is the slope.

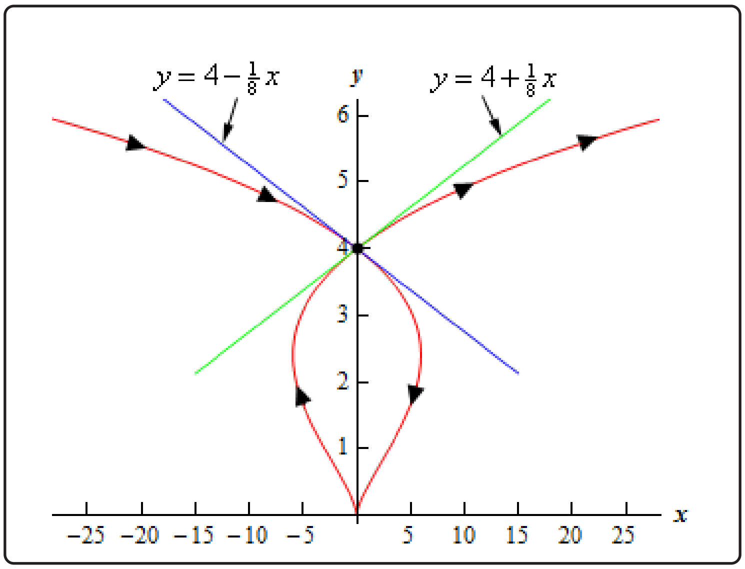
$$m = \left.\frac{dy}{dx}\right|_{t=2} = \frac{1}{8}$$

The tangent line (at $t = 2$) is then,

$$y = 4 + \frac{1}{8}x$$

Before we leave this example let's take a look at just how we could possibly get two tangents lines at a point. This was definitely not possible back in Calculus I where we first ran across tangent lines.

A quick graph of the parametric curve will explain what is going on here.



So, the parametric curve crosses itself! That explains how there can be more than one tangent line. There is one tangent line for each instance that the curve goes through the point.

The next topic that we need to discuss in this section is that of horizontal and vertical tangents. We can easily identify where these will occur (or at least the t's that will give them) by looking at the derivative formula.

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}}$$

Horizontal tangents will occur where the derivative is zero and that means that we'll get horizontal tangent at values of t for which we have,

Horizontal Tangent for Parametric Equations

$$\frac{dy}{dt} = 0, \text{ provided } \frac{dx}{dt} \neq 0$$

Vertical tangents will occur where the derivative is not defined and so we'll get vertical tangents at values of t for which we have,

Vertical Tangent for Parametric Equations

$$\frac{dx}{dt} = 0, \text{ provided } \frac{dy}{dt} \neq 0$$

Let's take a quick look at an example of this.

Example 2

Determine the x-y coordinates of the points where the following parametric equations will have horizontal or vertical tangents.

$$x = t^3 - 3t \qquad y = 3t^2 - 9$$

Solution

We'll first need the derivatives of the parametric equations.

$$\frac{dx}{dt} = 3t^2 - 3 = 3\left(t^2 - 1\right) \qquad \frac{dy}{dt} = 6t$$

Horizontal Tangents

We'll have horizontal tangents where,

$$6t = 0 \qquad \Rightarrow \qquad t = 0$$

Now, this is the value of t which gives the horizontal tangents and we were asked to find the *x-y* coordinates of the point. To get these we just need to plug t into the parametric equations. Therefore, the only horizontal tangent will occur at the point $(0, -9)$.

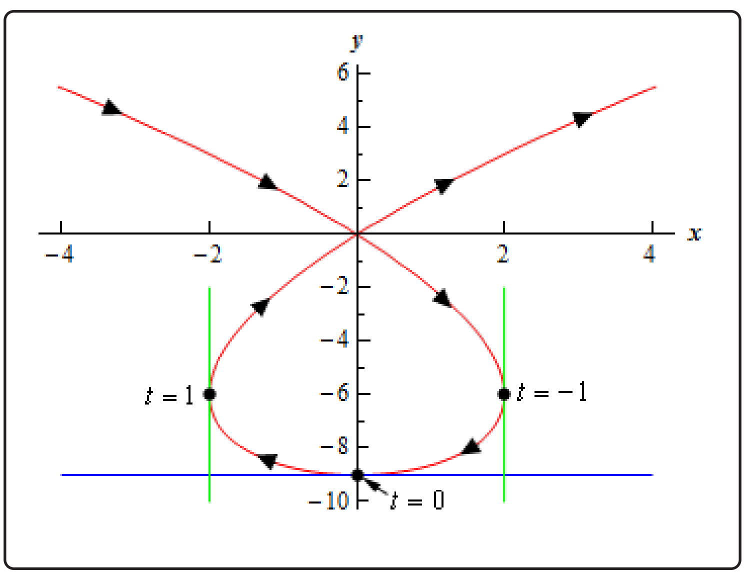
Vertical Tangents

In this case we need to solve,

$$3\left(t^2 - 1\right) = 0 \qquad \Rightarrow \qquad t = \pm 1$$

The two vertical tangents will occur at the points $(2, -6)$ and $(-2, -6)$.

For the sake of completeness and at least partial verification here is the sketch of the parametric curve.



The final topic that we need to discuss in this section really isn't related to tangent lines but does fit in nicely with the derivation of the derivative that we needed to get the slope of the tangent line.

Before moving into the new topic let's first remind ourselves of the formula for the first derivative and in the process rewrite it slightly.

$$\frac{dy}{dx} = \frac{d}{dx}(y) = \frac{\frac{d}{dt}(y)}{\frac{dx}{dt}}$$

Written in this way we can see that the formula actually tells us how to differentiate a function y (as a function of t) with respect to x (when x is also a function of t) when we are using parametric equations.

Now let's move onto the final topic of this section. We would also like to know how to get the second derivative of y with respect to x.

$$\frac{d^2y}{dx^2}$$

Getting a formula for this is fairly simple if we remember the rewritten formula for the first derivative above.

Second Derivative for Parametric Equations, $\frac{d^2y}{dx^2}$

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{\frac{dx}{dt}}$$

It is important to note that,

$$\frac{d^2y}{dx^2} \neq \frac{\frac{d^2y}{dt^2}}{\frac{d^2x}{dt^2}}$$

Let's work a quick example.

Example 3

Find the second derivative for the following set of parametric equations.

$$x = t^5 - 4t^3 \qquad y = t^2$$

Solution

This is the set of parametric equations that we used in the first example and so we already have the following computations completed.

$$\frac{dy}{dt} = 2t \qquad \frac{dx}{dt} = 5t^4 - 12t^2 \qquad \frac{dy}{dx} = \frac{2}{5t^3 - 12t}$$

We will first need the following,

$$\frac{d}{dt}\left(\frac{2}{5t^3 - 12t}\right) = \frac{-2\left(15t^2 - 12\right)}{\left(5t^3 - 12t\right)^2} = \frac{24 - 30t^2}{\left(5t^3 - 12t\right)^2}$$

The second derivative is then,

$$\begin{aligned}
\frac{d^2y}{dx^2} &= \frac{\frac{d}{dt}\left(\frac{dy}{dx}\right)}{\frac{dx}{dt}} \\
&= \frac{\frac{24-30t^2}{\left(5t^3-12t\right)^2}}{5t^4 - 12t^2} \\
&= \frac{24 - 30t^2}{\left(5t^4 - 12t^2\right)\left(5t^3 - 12t\right)^2} \\
&= \frac{24 - 30t^2}{t\left(5t^3 - 12t\right)^3}
\end{aligned}$$

So, why would we want the second derivative? Well, recall from your Calculus I class that with the second derivative we can determine where a curve is concave up and concave down. We could do the same thing with parametric equations if we wanted to.

Example 4

Determine the values of t for which the parametric curve given by the following set of parametric equations is concave up and concave down.

$$x = 1 - t^2 \qquad y = t^7 + t^5$$

Solution

To compute the second derivative we'll first need the following.

$$\frac{dy}{dt} = 7t^6 + 5t^4 \qquad \frac{dx}{dt} = -2t \qquad \frac{dy}{dx} = \frac{7t^6 + 5t^4}{-2t} = -\frac{1}{2}\left(7t^5 + 5t^3\right)$$

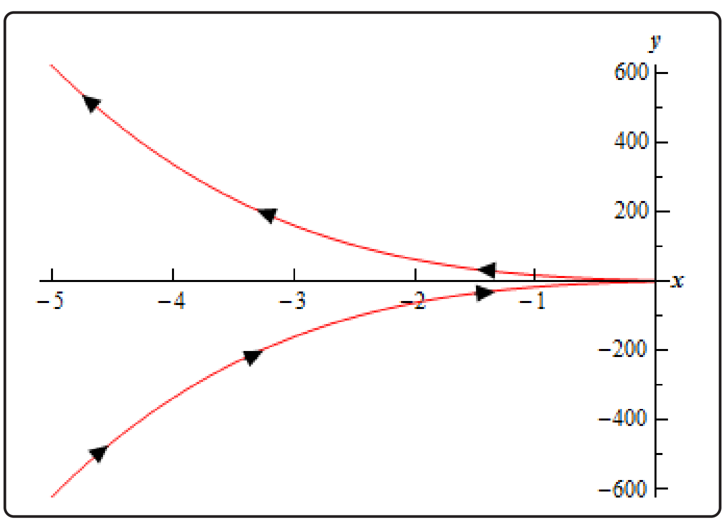
Note that we can also use the first derivative above to get some information about the increasing/decreasing nature of the curve as well. In this case it looks like the parametric curve will be increasing if $t < 0$ and decreasing if $t > 0$.

Now let's move on to the second derivative.

$$\frac{d^2y}{dx^2} = \frac{-\frac{1}{2}\left(35t^4 + 15t^2\right)}{-2t} = \frac{1}{4}\left(35t^3 + 15t\right)$$

It's clear, hopefully, that the second derivative will only be zero at $t = 0$. Using this we can see that the second derivative will be negative if $t < 0$ and positive if $t > 0$. So the parametric curve will be concave down for $t < 0$ and concave up for $t > 0$.

Here is a sketch of the curve for completeness sake.



9.3 Area with Parametric Equations

In this section we will find a formula for determining the area under a parametric curve given by the parametric equations,

$$x = f(t) \qquad y = g(t)$$

We will also need to further add in the assumption that the curve is traced out exactly once as t increases from α to β.

We will do this in much the same way that we found the first derivative in the previous section. We will first recall how to find the area under $y = F(x)$ on $a \le x \le b$.

$$A = \int_a^b F(x)\,dx$$

We will now think of the parametric equation $x = f(t)$ as a substitution in the integral. We will also assume that $a = f(\alpha)$ and $b = f(\beta)$ for the purposes of this formula. There is actually no reason to assume that this will always be the case and so we'll give a corresponding formula later if it's the opposite case ($b = f(\alpha)$ and $a = f(\beta)$).

So, if this is going to be a substitution we'll need,

$$dx = f'(t)\,dt$$

Plugging this into the area formula above and making sure to change the limits to their corresponding t values gives us,

$$A = \int_\alpha^\beta F(f(t))\,f'(t)\,dt$$

Since we don't know what $F(x)$ is we'll use the fact that

$$y = F(x) = F(f(t)) = g(t)$$

and we arrive at the formula that we want.

Area Under Parametric Curve, Formula I

$$A = \int_\alpha^\beta g(t)\,f'(t)\,dt$$

Now, if we should happen to have $b = f(\alpha)$ and $a = f(\beta)$ the formula would be,

Area Under Parametric Curve, Formula II

$$A = \int_\beta^\alpha g(t)\,f'(t)\,dt$$

Let's work an example.

Example 1

Determine the area under the parametric curve given by the following parametric equations.

$$x = 6\left(\theta - \sin(\theta)\right) \qquad y = 6\left(1 - \cos(\theta)\right) \qquad 0 \le \theta \le 2\pi$$

Solution

First, notice that we've switched the parameter to θ for this problem. This is to make sure that we don't get too locked into always having t as the parameter.

Now, we could graph this to verify that the curve is traced out exactly once for the given range if we wanted to. We are going to be looking at this curve in more detail after this example so we won't sketch its graph here.

There really isn't too much to this example other than plugging the parametric equations into the formula. We'll first need the derivative of the parametric equation for x however.

$$\frac{dx}{d\theta} = 6\left(1 - \cos(\theta)\right)$$

The area is then,

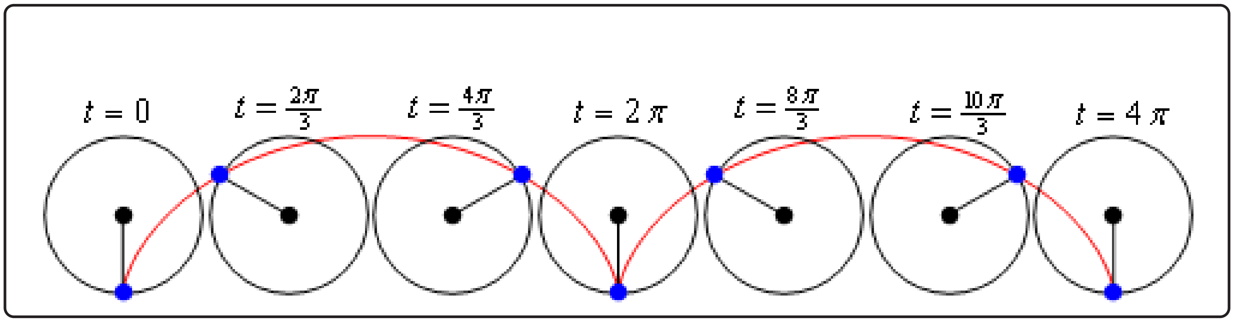
$$\begin{aligned} A &= \int_0^{2\pi} 36\left(1 - \cos(\theta)\right)^2 d\theta \\ &= 36\int_0^{2\pi} 1 - 2\cos(\theta) + \cos^2(\theta)\, d\theta \\ &= 36\int_0^{2\pi} \frac{3}{2} - 2\cos(\theta) + \frac{1}{2}\cos\left(2\theta\right)\, d\theta \\ &= 36\left(\frac{3}{2}\theta - 2\sin(\theta) + \frac{1}{4}\sin\left(2\theta\right)\right)\Bigg|_0^{2\pi} \\ &= 108\pi \end{aligned}$$

The parametric curve (without the limits) we used in the previous example is called a **cycloid**. In its general form the cycloid is,

$$x = r\left(\theta - \sin(\theta)\right) \qquad y = r\left(1 - \cos(\theta)\right)$$

The cycloid represents the following situation. Consider a wheel of radius r. Let the point where the wheel touches the ground initially be called P. Then start rolling the wheel to the right. As the wheel rolls to the right trace out the path of the point P. The path that the point P traces out is called a cycloid and is given by the equations above. In these equations we can think of θ as the angle through which the point P has rotated.

Here is a cycloid sketched out with the wheel shown at various places. The blue dot is the point P on the wheel that we're using to trace out the curve.



From this sketch we can see that one arch of the cycloid is traced out in the range $0 \le \theta \le 2\pi$. This makes sense when you consider that the point P will be back on the ground after it has rotated through an angle of 2π.

9.4 Arc Length with Parametric Equations

In the previous two sections we've looked at a couple of Calculus I topics in terms of parametric equations. We now need to look at a couple of Calculus II topics in terms of parametric equations.

In this section we will look at the arc length of the parametric curve given by,

$$x = f(t) \qquad y = g(t) \qquad \alpha \le t \le \beta$$

We will also be assuming that the curve is traced out exactly once as t increases from α to β. We will also need to assume that the curve is traced out from left to right as t increases. This is equivalent to saying,

$$\frac{dx}{dt} \ge 0 \qquad \text{for } \alpha \le t \le \beta$$

So, let's start out the derivation by recalling the arc length formula as we first derived it in the arc length section of the Applications of Integrals chapter.

$$L = \int ds$$

where,

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,dx \qquad \text{if } y = f(x),\ a \le x \le b$$
$$ds = \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\,dy \qquad \text{if } x = h(y),\ c \le y \le d$$

We will use the first ds above because we have a nice formula for the derivative in terms of the parametric equations (see the Tangents with Parametric Equations section). To use this we'll also need to know that,

$$dx = f'(t)\,dt = \frac{dx}{dt}\,dt$$

The arc length formula then becomes,

$$L = \int_{\alpha}^{\beta} \sqrt{1 + \left(\frac{\frac{dy}{dt}}{\frac{dx}{dt}}\right)^2}\ \frac{dx}{dt}dt = \int_{\alpha}^{\beta} \sqrt{1 + \frac{\left(\frac{dy}{dt}\right)^2}{\left(\frac{dx}{dt}\right)^2}}\ \frac{dx}{dt}dt$$

This is a particularly unpleasant formula. However, if we factor out the denominator from the square root we arrive at,

$$L = \int_{\alpha}^{\beta} \frac{1}{\left|\frac{dx}{dt}\right|}\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\ \frac{dx}{dt}dt$$

Now, making use of our assumption that the curve is being traced out from left to right we can drop the absolute value bars on the derivative which will allow us to cancel the two derivatives that are outside the square root and this gives,

Arc Length for Parametric Equations

$$L = \int_{\alpha}^{\beta} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt$$

Notice that we could have used the second formula for ds above if we had assumed instead that

$$\frac{dy}{dt} \geq 0 \qquad \text{for } \alpha \leq t \leq \beta$$

If we had gone this route in the derivation we would have gotten the same formula.

Let's take a look at an example.

Example 1

Determine the length of the parametric curve given by the following parametric equations.

$$x = 3\sin(t) \qquad y = 3\cos(t) \qquad 0 \leq t \leq 2\pi$$

Solution

We know that this is a circle of radius 3 centered at the origin from our prior discussion about graphing parametric curves. We also know from this discussion that it will be traced out exactly once in this range.

So, we can use the formula we derived above. We'll first need the following,

$$\frac{dx}{dt} = 3\cos(t) \qquad \frac{dy}{dt} = -3\sin(t)$$

The length is then,

$$\begin{aligned} L &= \int_0^{2\pi} \sqrt{9\sin^2(t) + 9\cos^2(t)}\, dt \\ &= \int_0^{2\pi} 3\sqrt{\sin^2(t) + \cos^2(t)}\, dt \\ &= 3\int_0^{2\pi} dt \\ &= 6\pi \end{aligned}$$

Since this is a circle we could have just used the fact that the length of the circle is just the circumference of the circle. This is a nice way, in this case, to verify our result.

Let's take a look at one possible consequence if a curve is traced out more than once and we try to find the length of the curve without taking this into account.

Example 2

Use the arc length formula for the following parametric equations.

$$x = 3\sin(3t) \qquad y = 3\cos(3t) \qquad 0 \le t \le 2\pi$$

Solution

Notice that this is the identical circle that we had in the previous example and so the length is still 6 p. However, for the range given we know it will trace out the curve three times instead once as required for the formula. Despite that restriction let's use the formula anyway and see what happens.

In this case the derivatives are,

$$\frac{dx}{dt} = 9\cos(3t) \qquad \frac{dy}{dt} = -9\sin(3t)$$

and the length formula gives,

$$\begin{aligned} L &= \int_0^{2\pi} \sqrt{81\sin^2(3t) + 81\cos^2(3t)}\,dt \\ &= \int_0^{2\pi} 9\,dt \\ &= 18\pi \end{aligned}$$

The answer we got form the arc length formula in this example was 3 times the actual length. Recalling that we also determined that this circle would trace out three times in the range given, the answer should make some sense.

If we had wanted to determine the length of the circle for this set of parametric equations we would need to determine a range of t for which this circle is traced out exactly once. This is, $0 \le t \le \frac{2\pi}{3}$. Using this range of t we get the following for the length.

$$\begin{aligned} L &= \int_0^{\frac{2\pi}{3}} \sqrt{81\sin^2(3t) + 81\cos^2(3t)}\,dt \\ &= \int_0^{\frac{2\pi}{3}} 9\,dt \\ &= 6\pi \end{aligned}$$

which is the correct answer.

Be careful to not make the assumption that this is always what will happen if the curve is traced out more than once. Just because the curve traces out n times does not mean that the arc length formula will give us n times the actual length of the curve!

Before moving on to the next section let's notice that we can put the arc length formula derived in this section into the same form that we had when we first looked at arc length. The only difference is that we will add in a definition for ds when we have parametric equations.

The arc length formula can be summarized as,

$$L = \int ds$$

where,

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \qquad \text{if } y = f(x),\ a \le x \le b$$

$$ds = \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy \qquad \text{if } x = h(y),\ c \le y \le d$$

$$ds = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt \qquad \text{if } x = f(t), y = g(t),\ \alpha \le t \le \beta$$

9.5 Surface Area with Parametric Equations

In this final section of looking at calculus applications with parametric equations we will take a look at determining the surface area of a region obtained by rotating a parametric curve about the x or y-axis.

We will rotate the parametric curve given by,

$$x = f(t) \qquad y = g(t) \qquad \alpha \le t \le \beta$$

about the x or y-axis. We are going to assume that the curve is traced out exactly once as t increases from α to β. At this point there actually isn't all that much to do. We know that the surface area can be found by using one of the following two formulas depending on the axis of rotation (recall the Surface Area section of the Applications of Integrals chapter).

$$S = \int 2\pi y\, ds \qquad \text{rotation about } x-\text{axis}$$
$$S = \int 2\pi x\, ds \qquad \text{rotation about } y-\text{axis}$$

All that we need is a formula for ds to use and from the previous section we have,

$$ds = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt \qquad \text{if } x = f(t), y = g(t),\ \alpha \le t \le \beta$$

which is exactly what we need.

We will need to be careful with the x or y that is in the original surface area formula. Back when we first looked at surface area we saw that sometimes we had to substitute for the variable in the integral and at other times we didn't. This was dependent upon the ds that we used. In this case however, we will always have to substitute for the variable. The ds that we use for parametric equations introduces a dt into the integral and that means that everything needs to be in terms of t. Therefore, we will need to substitute the appropriate parametric equation for x or y depending on the axis of rotation.

Let's take a quick look at an example.

Example 1

Determine the surface area of the solid obtained by rotating the following parametric curve about the x-axis.

$$x = \cos^3(\theta) \qquad y = \sin^3(\theta) \qquad 0 \le \theta \le \frac{\pi}{2}$$

Solution

We'll first need the derivatives of the parametric equations.

$$\frac{dx}{d\theta} = -3\cos^2(\theta)\sin(\theta) \qquad\qquad \frac{dy}{d\theta} = 3\sin^2(\theta)\cos(\theta)$$

Before plugging into the surface area formula let's get the ds out of the way.

$$\begin{aligned} ds &= \sqrt{9\cos^4(\theta)\sin^2(\theta) + 9\sin^4(\theta)\cos^2(\theta)}\, d\theta \\ &= 3\left|\cos(\theta)\sin(\theta)\right|\sqrt{\cos^2(\theta) + \sin^2(\theta)}\, d\theta \\ &= 3\cos(\theta)\sin(\theta)\, d\theta \end{aligned}$$

Notice that we could drop the absolute value bars since both sine and cosine are positive in this range of q given.

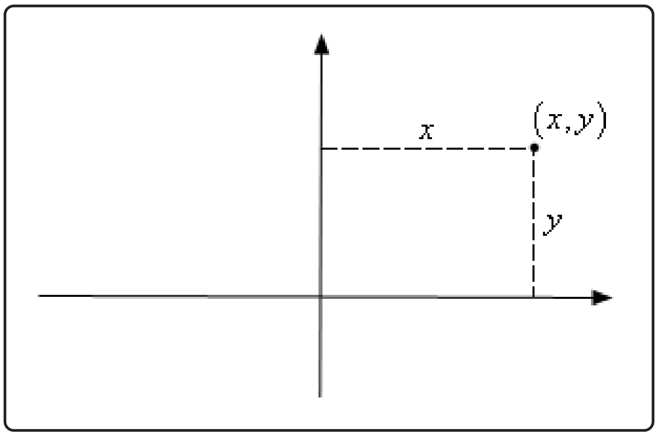
Now let's get the surface area and don't forget to also plug in for the y.

$$\begin{aligned} S &= \int 2\pi y\, ds \\ &= 2\pi \int_0^{\frac{\pi}{2}} \sin^3(\theta)\big(3\cos(\theta)\sin(\theta)\big)\, d\theta \\ &= 6\pi \int_0^{\frac{\pi}{2}} \sin^4(\theta)\cos(\theta)\, d\theta \qquad\qquad u = \sin(\theta) \\ &= 6\pi \int_0^1 u^4\, du \\ &= \frac{6\pi}{5} \end{aligned}$$

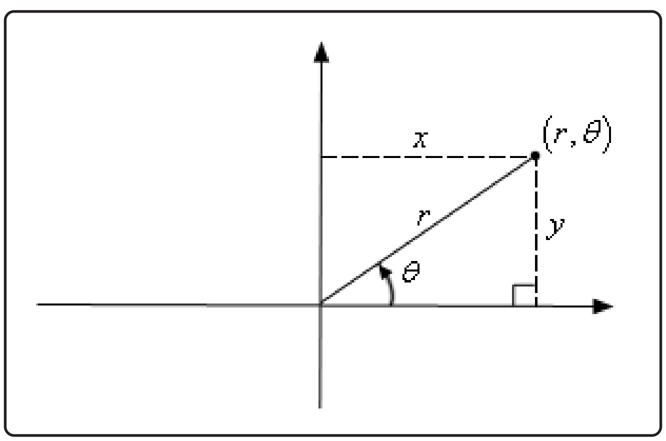
9.6 Polar Coordinates

Up to this point we've dealt exclusively with the Cartesian (or Rectangular, or *x-y*) coordinate system. However, as we will see, this is not always the easiest coordinate system to work in. So, in this section we will start looking at the polar coordinate system.

Coordinate systems are really nothing more than a way to define a point in space. For instance in the Cartesian coordinate system at point is given the coordinates (x, y) and we use this to define the point by starting at the origin and then moving x units horizontally followed by y units vertically. This is shown in the sketch below.

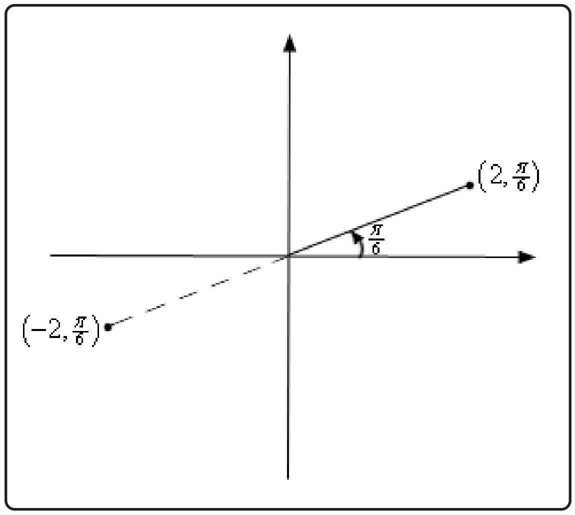


This is not, however, the only way to define a point in two dimensional space. Instead of moving vertically and horizontally from the origin to get to the point we could instead go straight out of the origin until we hit the point and then determine the angle this line makes with the positive x-axis. We could then use the distance of the point from the origin and the amount we needed to rotate from the positive x-axis as the coordinates of the point. This is shown in the sketch below.



Coordinates in this form are called **polar coordinates**.

The above discussion may lead one to think that r must be a positive number. However, we also allow r to be negative. Below is a sketch of the two points $\left(2, \frac{\pi}{6}\right)$ and $\left(-2, \frac{\pi}{6}\right)$.

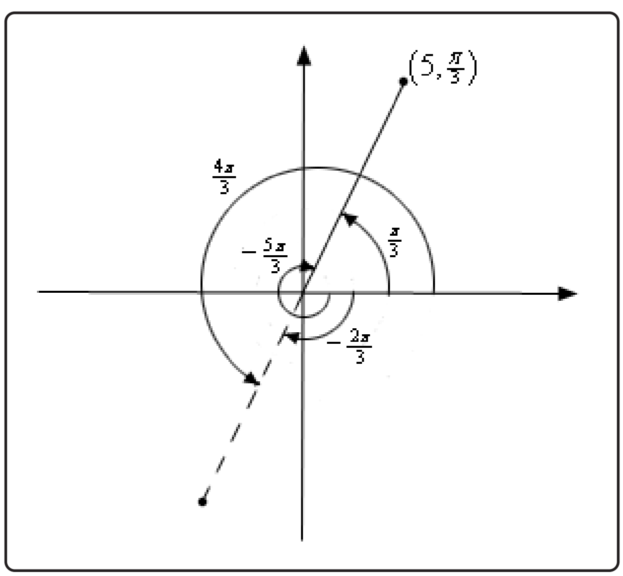


From this sketch we can see that if r is positive the point will be in the same quadrant as θ. On the other hand if r is negative the point will end up in the quadrant exactly opposite θ. Notice as well that the coordinates $\left(-2, \frac{\pi}{6}\right)$ describe the same point as the coordinates $\left(2, \frac{7\pi}{6}\right)$ do. The coordinates $\left(2, \frac{7\pi}{6}\right)$ tells us to rotate an angle of $\frac{7\pi}{6}$ from the positive x-axis, this would put us on the dashed line in the sketch above, and then move out a distance of 2.

This leads to an important difference between Cartesian coordinates and polar coordinates. In Cartesian coordinates there is exactly one set of coordinates for any given point. With polar coordinates this isn't true. In polar coordinates there is literally an infinite number of coordinates for a given point. For instance, the following four points are all coordinates for the same point.

$$\left(5, \frac{\pi}{3}\right) = \left(5, -\frac{5\pi}{3}\right) = \left(-5, \frac{4\pi}{3}\right) = \left(-5, -\frac{2\pi}{3}\right)$$

Here is a sketch of the angles used in these four sets of coordinates.



In the second coordinate pair we rotated in a clock-wise direction to get to the point. We shouldn't forget about rotating in the clock-wise direction. Sometimes it's what we have to do.

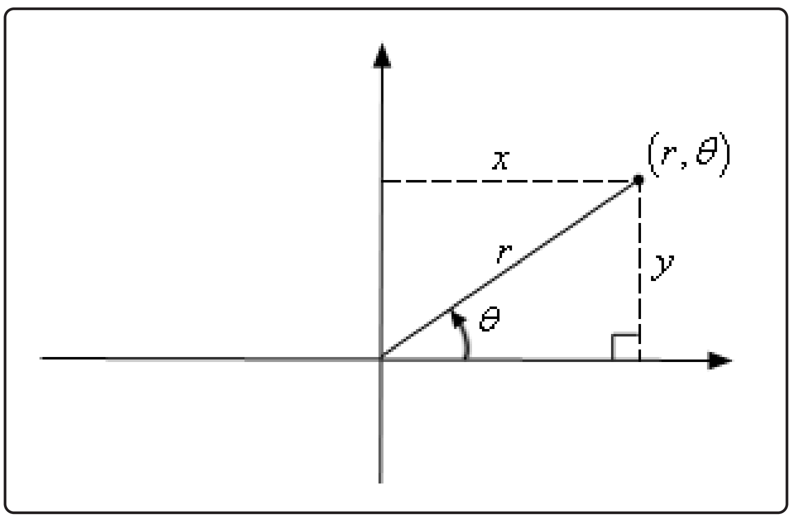
The last two coordinate pairs use the fact that if we end up in the opposite quadrant from the point we can use a negative r to get back to the point and of course there is both a counter clock-wise and a clock-wise rotation to get to the angle.

These four points only represent the coordinates of the point without rotating around the system more than once. If we allow the angle to make as many complete rotations about the axis system as we want then there are an infinite number of coordinates for the same point. In fact, the point (r, θ) can be represented by any of the following coordinate pairs.

$$(r, \theta + 2\pi n) \qquad (-r, \theta + (2n+1)\pi), \qquad \text{where } n \text{ is any integer.}$$

Next, we should talk about the origin of the coordinate system. In polar coordinates the origin is often called the **pole**. Because we aren't actually moving away from the origin/pole we know that $r = 0$. However, we can still rotate around the system by any angle we want and so the coordinates of the origin/pole are $(0, \theta)$.

Now that we've got a grasp on polar coordinates we need to think about converting between the two coordinate systems. Well start out with the following sketch reminding us how both coordinate systems work.



Note that we've got a right triangle above and with that we can get the following equations that will convert polar coordinates into Cartesian coordinates.

Polar to Cartesian Conversion Formulas

$$x = r\cos(\theta) \qquad y = r\sin(\theta)$$

Converting from Cartesian is almost as easy. Let's first notice the following.

$$\begin{aligned} x^2 + y^2 &= \left(r\cos(\theta)\right)^2 + \left(r\sin(\theta)\right)^2 \\ &= r^2\cos^2(\theta) + r^2\sin^2(\theta) \\ &= r^2\left(\cos^2(\theta) + \sin^2(\theta)\right) = r^2 \end{aligned}$$

This is a very useful formula that we should remember, however we are after an equation for r so let's take the square root of both sides. This gives,

$$r = \sqrt{x^2 + y^2}$$

Note that technically we should have a plus or minus in front of the root since we know that r can be either positive or negative. We will run with the convention of positive r here.

Getting an equation for θ is almost as simple. We'll start with,

$$\frac{y}{x} = \frac{r\sin(\theta)}{r\cos(\theta)} = \tan(\theta)$$

Taking the inverse tangent of both sides gives,

$$\theta = \tan^{-1}\left(\frac{y}{x}\right)$$

We will need to be careful with this because inverse tangents only return values in the range $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$. Recall that there is a second possible angle and that the second angle is given by $\theta + \pi$.

Summarizing then gives the following formulas for converting from Cartesian coordinates to polar coordinates.

Cartesian to Polar Conversion Formulas

$$r^2 = x^2 + y^2 \qquad\qquad r = \sqrt{x^2 + y^2}$$

$$\theta_1 = \tan^{-1}\left(\frac{y}{x}\right) \quad \text{OR} \quad \theta_2 = \theta_1 + \pi$$

Let's work a quick example.

Example 1

Convert each of the following points into the given coordinate system.

(a) Convert $\left(-4, \frac{2\pi}{3}\right)$ into Cartesian coordinates.

(b) Convert $(-1, -1)$ into polar coordinates.

Solution

(a) Convert $\left(-4, \frac{2\pi}{3}\right)$ into Cartesian coordinates.

This conversion is easy enough. All we need to do is plug the points into the formulas.

$$x = -4\cos\left(\frac{2\pi}{3}\right) = -4\left(-\frac{1}{2}\right) = 2$$
$$y = -4\sin\left(\frac{2\pi}{3}\right) = -4\left(\frac{\sqrt{3}}{2}\right) = -2\sqrt{3}$$

So, in Cartesian coordinates this point is $\left(2, -2\sqrt{3}\right)$.

(b) Convert $(-1, -1)$ into polar coordinates.

Let's first get r.

$$r = \sqrt{(-1)^2 + (-1)^2} = \sqrt{2}$$

Now, let's get θ.

$$\theta = \tan^{-1}\left(\frac{-1}{-1}\right) = \tan^{-1}(1) = \frac{\pi}{4}$$

This is not the correct angle however. This value of θ is in the first quadrant and the point we've been given is in the third quadrant. As noted above we can get the correct angle by adding p onto this. Therefore, the actual angle is,

$$\theta = \frac{\pi}{4} + \pi = \frac{5\pi}{4}$$

So, in polar coordinates the point is $\left(\sqrt{2}, \frac{5\pi}{4}\right)$. Note as well that we could have used the first θ that we got by using a negative r. In this case the point could also be written in polar coordinates as $\left(-\sqrt{2}, \frac{\pi}{4}\right)$.

We can also use the above formulas to convert equations from one coordinate system to the other.

Example 2

Convert each of the following into an equation in the given coordinate system.

(a) Convert $2x - 5x^3 = 1 + xy$ into polar coordinates.

(b) Convert $r = -8\cos(\theta)$ into Cartesian coordinates.

Solution

(a) Convert $2x - 5x^3 = 1 + xy$ into polar coordinates.

In this case there really isn't much to do other than plugging in the formulas for x and y (*i.e.* the Cartesian coordinates) in terms of r and θ (*i.e.* the polar coordinates).

$$2\left(r\cos(\theta)\right) - 5\left(r\cos(\theta)\right)^3 = 1 + \left(r\cos(\theta)\right)\left(r\sin(\theta)\right)$$
$$2r\cos(\theta) - 5r^3\cos^3(\theta) = 1 + r^2\cos(\theta)\sin(\theta)$$

(b) Convert $r = -8\cos(\theta)$ into Cartesian coordinates.

This one is a little trickier, but not by much. First notice that we could substitute straight for the r. However, there is no straight substitution for the cosine that will give us only Cartesian coordinates. If we had an r on the right along with the cosine then we could do a direct substitution. So, if an r on the right side would be convenient let's put one there, just don't forget to put one on the left side as well.

$$r^2 = -8r\cos(\theta)$$

We can now make some substitutions that will convert this into Cartesian coordinates.

$$x^2 + y^2 = -8x$$

Before moving on to the next subject let's do a little more work on the second part of the previous example.

The equation given in the second part is actually a fairly well known graph; it just isn't in a form that most people will quickly recognize. To identify it let's take the Cartesian coordinate equation and do a little rearranging.

$$x^2 + 8x + y^2 = 0$$

Now, complete the square on the x portion of the equation.

$$\begin{aligned} x^2 + 8x + 16 + y^2 &= 16 \\ (x+4)^2 + y^2 &= 16 \end{aligned}$$

So, this was a circle of radius 4 and center $(-4, 0)$.

This leads us into the final topic of this section.

Common Polar Coordinate Graphs

Let's identify a few of the more common graphs in polar coordinates. We'll also take a look at a couple of special polar graphs.

Lines

Some lines have fairly simple equations in polar coordinates.

1. $\theta = \beta$.

   We can see that this is a line by converting to Cartesian coordinates as follows

$$\begin{aligned} \theta &= \beta \\ \tan^{-1}\left(\frac{y}{x}\right) &= \beta \\ \frac{y}{x} &= \tan\beta \\ y &= (\tan\beta)\, x \end{aligned}$$

   This is a line that goes through the origin and makes an angle of β with the positive x-axis. Or, in other words it is a line through the origin with slope of $\tan\beta$.

2. $r\cos(\theta) = a$

   This is easy enough to convert to Cartesian coordinates to $x = a$. So, this is a vertical line.

3. $r\sin(\theta) = b$

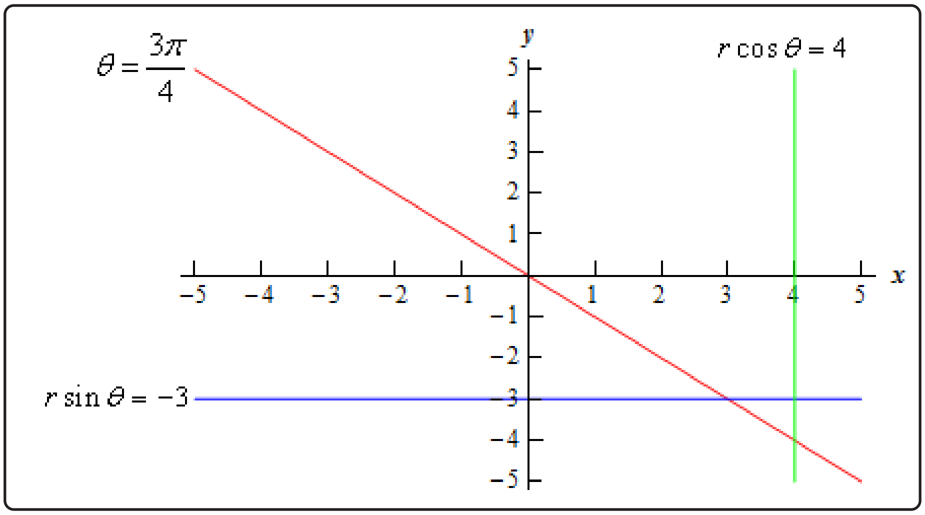
   Likewise, this converts to $y = b$ and so is a horizontal line.

Example 3

Graph $\theta = \frac{3\pi}{4}$, $r\cos(\theta) = 4$ and $r\sin(\theta) = -3$ on the same axis system.

Solution

There really isn't too much to this one other than doing the graph so here it is.



Circles

Let's take a look at the equations of circles in polar coordinates.

1. $r = a$.

   This equation is saying that no matter what angle we've got the distance from the origin must be a. If you think about it that is exactly the definition of a circle of radius a centered at the origin. So, this is a circle of radius a centered at the origin. This is also one of the reasons why we might want to work in polar coordinates. The equation of a circle centered at the origin has a very nice equation, unlike the corresponding equation in Cartesian coordinates.

2. $r = 2a\cos(\theta)$.

   We looked at a specific example of one of these when we were converting equations to Cartesian coordinates.

   This is a circle of radius $|a|$ and center $(a, 0)$. Note that a might be negative (as it was in our example above) and so the absolute value bars are required on the radius. They should not be used however on the center.

3. $r = 2b\sin(\theta)$.

   This is similar to the previous one. It is a circle of radius $|b|$ and center $(0, b)$.

4. $r = 2a\cos(\theta) + 2b\sin(\theta)$.

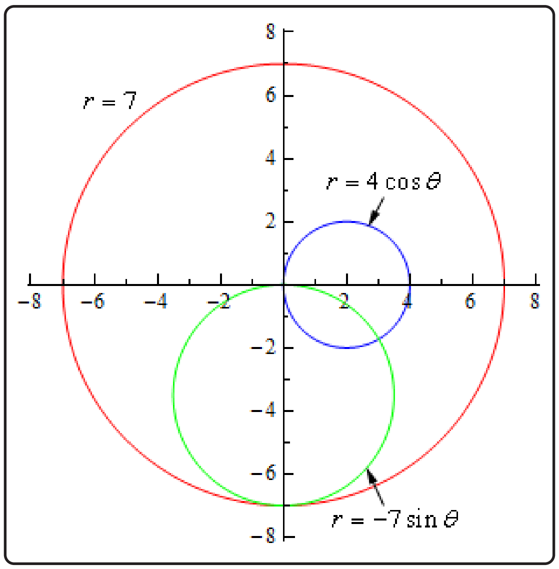
   This is a combination of the previous two and by completing the square twice it can be shown that this is a circle of radius $\sqrt{a^2 + b^2}$ and center (a, b). In other words, this is the general equation of a circle that isn't centered at the origin.

Example 4

Graph $r = 7$, $r = 4\cos(\theta)$, and $r = -7\sin(\theta)$ on the same axis system.

Solution

The first one is a circle of radius 7 centered at the origin. The second is a circle of radius 2 centered at $(2, 0)$. The third is a circle of radius $\frac{7}{2}$ centered at $\left(0, -\frac{7}{2}\right)$. Here is the graph of the three equations.



Note that it takes a range of $0 \leq \theta \leq 2\pi$ for a complete graph of $r = a$ and it only takes a range of $0 \leq \theta \leq \pi$ to graph the other circles given here. You can verify this with a quick table of values if you'd like to.

Cardioids and Limacons

These can be broken up into the following three cases.

1. Cardioids : $r = a \pm a\cos(\theta)$ and $r = a \pm a\sin(\theta)$.
   These have a graph that is vaguely heart shaped and always contain the origin.
2. Limacons with an inner loop : $r = a \pm b\cos(\theta)$ and $r = a \pm b\sin(\theta)$ with $a < b$.

   These will have an inner loop and will always contain the origin.
3. Limacons without an inner loop : $r = a \pm b\cos(\theta)$ and $r = a \pm b\sin(\theta)$ with $a > b$.

   These do not have an inner loop and do not contain the origin.

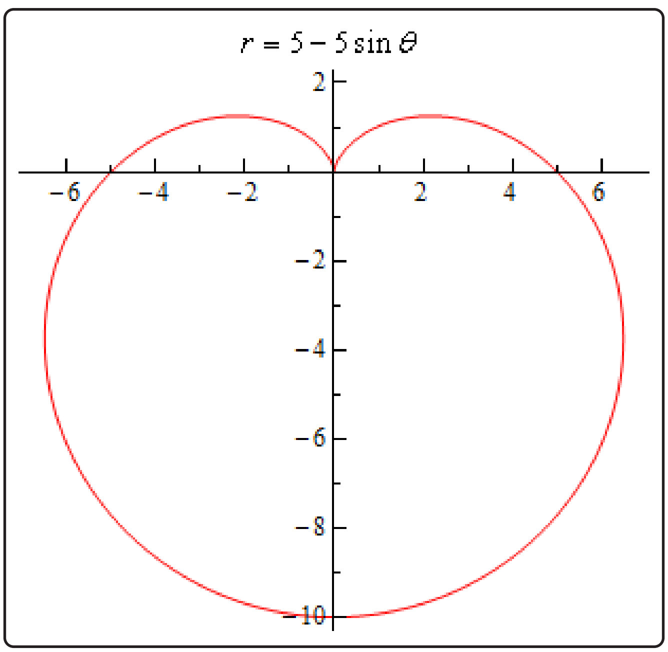
Example 5

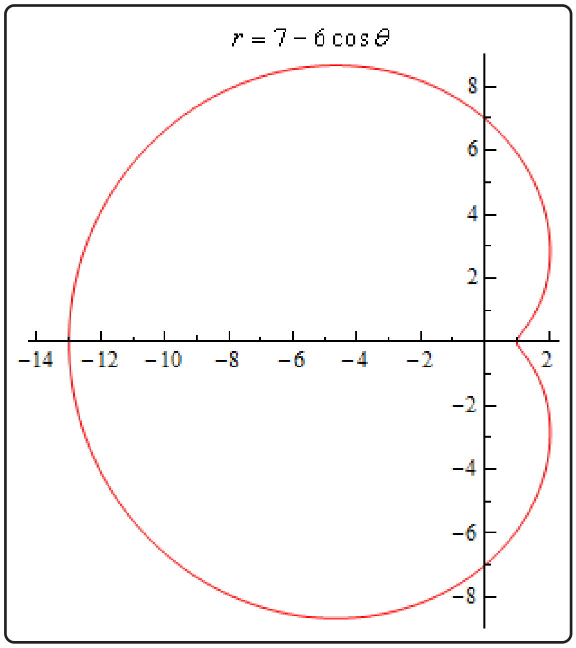
Graph $r = 5 - 5\sin(\theta)$, $r = 7 - 6\cos(\theta)$, and $r = 2 + 4\cos(\theta)$.

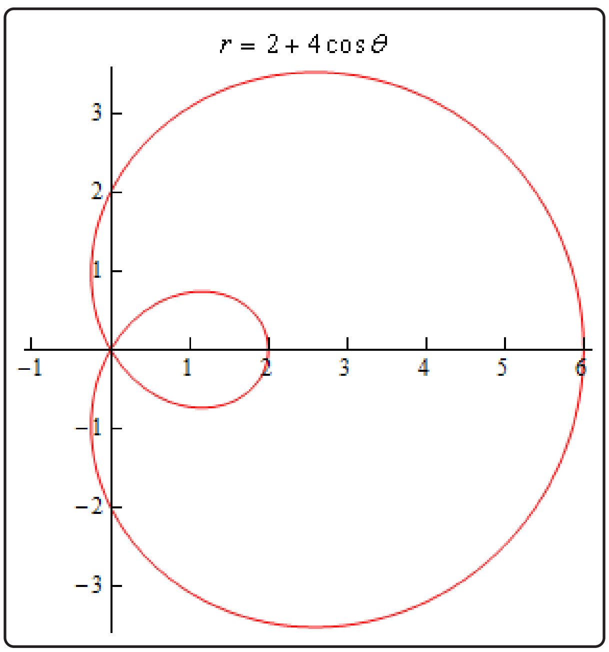
Solution

These will all graph out once in the range $0 \le \theta \le 2\pi$. Here is a table of values for each followed by graphs of each.

θ	$r = 5 - 5\sin(\theta)$	$r = 7 - 6\cos(\theta)$	$r = 2 + 4\cos(\theta)$
0	5	1	6
$\frac{\pi}{2}$	0	7	2
π	5	13	-2
$\frac{3\pi}{2}$	10	7	2
2π	5	1	6







There is one final thing that we need to do in this section. In the third graph in the previous example we had an inner loop. We will, on occasion, need to know the value of θ for which the graph will pass through the origin. To find these all we need to do is set the equation equal to zero and solve as follows,

$$0 = 2 + 4\cos(\theta) \qquad \Rightarrow \qquad \cos(\theta) = -\frac{1}{2} \qquad \Rightarrow \qquad \theta = \frac{2\pi}{3}, \frac{4\pi}{3}$$

9.7 Tangents with Polar Coordinates

We now need to discuss some calculus topics in terms of polar coordinates.

We will start with finding tangent lines to polar curves. In this case we are going to assume that the equation is in the form $r = f(\theta)$. With the equation in this form we can actually use the equation for the derivative $\frac{dy}{dx}$ we derived when we looked at tangent lines with parametric equations. To do this however requires us to come up with a set of parametric equations to represent the curve. This is actually pretty easy to do.

From our work in the previous section we have the following set of conversion equations for going from polar coordinates to Cartesian coordinates.

$$x = r\cos(\theta) \qquad\qquad y = r\sin(\theta)$$

Now, we'll use the fact that we're assuming that the equation is in the form $r = f(\theta)$. Substituting this into these equations gives the following set of parametric equations (with θ as the parameter) for the curve.

$$x = f(\theta)\cos(\theta) \qquad\qquad y = f(\theta)\sin(\theta)$$

Now, we will need the following derivatives.

$$\begin{aligned}\frac{dx}{d\theta} &= f'(\theta)\cos(\theta) - f(\theta)\sin(\theta) \\ &= \frac{dr}{d\theta}\cos(\theta) - r\sin(\theta)\end{aligned} \qquad\qquad \begin{aligned}\frac{dy}{d\theta} &= f'(\theta)\sin(\theta) + f(\theta)\cos(\theta) \\ &= \frac{dr}{d\theta}\sin(\theta) + r\cos(\theta)\end{aligned}$$

The derivative $\frac{dy}{dx}$ is then,

Derivative with Polar Coordinates, $\frac{dy}{dx}$

$$\frac{dy}{dx} = \frac{\frac{dr}{d\theta}\sin(\theta) + r\cos(\theta)}{\frac{dr}{d\theta}\cos(\theta) - r\sin(\theta)}$$

Note that rather than trying to remember this formula it would probably be easier to remember how we derived it and just remember the formula for parametric equations.

Let's work a quick example with this.

Example 1

Determine the equation of the tangent line to $r = 3 + 8\sin(\theta)$ at $\theta = \dfrac{\pi}{6}$.

Solution

We'll first need the following derivative.

$$\frac{dr}{d\theta} = 8\cos(\theta)$$

The formula for the derivative $\frac{dy}{dx}$ becomes,

$$\frac{dy}{dx} = \frac{8\cos(\theta)\sin(\theta) + (3 + 8\sin(\theta))\cos(\theta)}{8\cos^2(\theta) - (3 + 8\sin(\theta))\sin(\theta)} = \frac{16\cos(\theta)\sin(\theta) + 3\cos(\theta)}{8\cos^2(\theta) - 3\sin(\theta) - 8\sin^2(\theta)}$$

The slope of the tangent line is,

$$m = \left.\frac{dy}{dx}\right|_{\theta=\frac{\pi}{6}} = \frac{4\sqrt{3} + \frac{3\sqrt{3}}{2}}{4 - \frac{3}{2}} = \frac{11\sqrt{3}}{5}$$

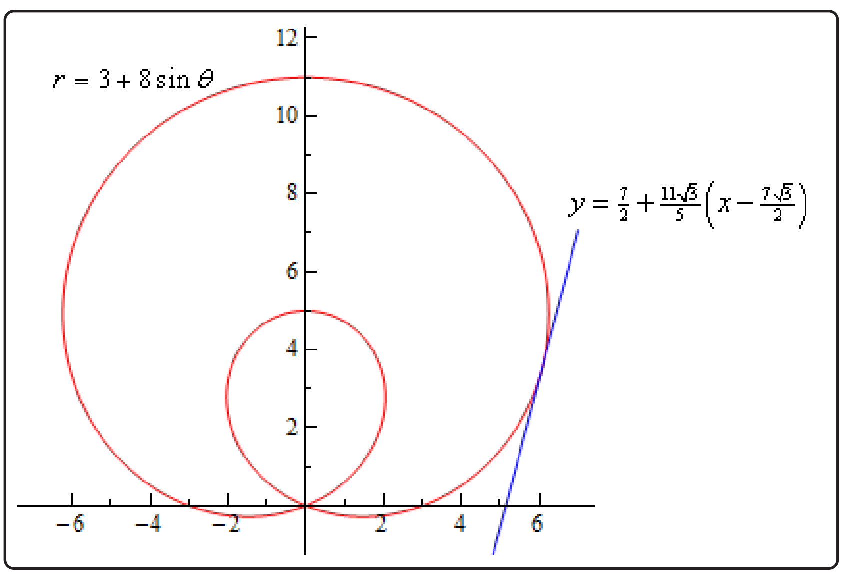
Now, at $\theta = \frac{\pi}{6}$ we have $r = 7$. We'll need to get the corresponding x-y coordinates so we can get the tangent line.

$$x = 7\cos\left(\frac{\pi}{6}\right) = \frac{7\sqrt{3}}{2} \qquad y = 7\sin\left(\frac{\pi}{6}\right) = \frac{7}{2}$$

The tangent line is then,

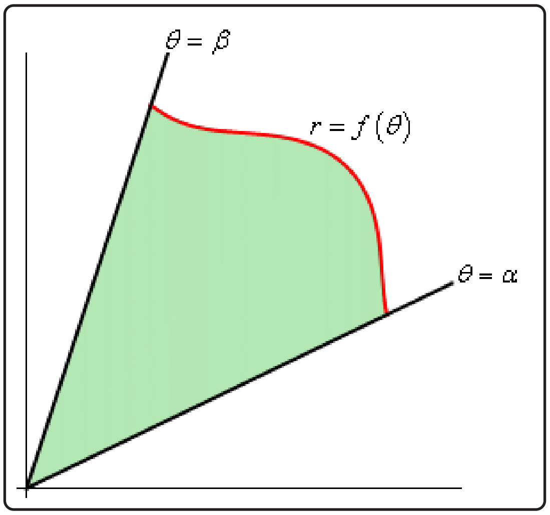
$$y = \frac{7}{2} + \frac{11\sqrt{3}}{5}\left(x - \frac{7\sqrt{3}}{2}\right)$$

For the sake of completeness here is a graph of the curve and the tangent line.



9.8 Area with Polar Coordinates

In this section we are going to look at areas enclosed by polar curves. Note as well that we said "enclosed by" instead of "under" as we typically have in these problems. These problems work a little differently in polar coordinates. Here is a sketch of what the area that we'll be finding in this section looks like.



We'll be looking for the shaded area in the sketch above. The formula for finding this area is,

Area Enclosed by Curve

$$A = \int_{\alpha}^{\beta} \frac{1}{2} r^2 \, d\theta$$

Notice that we use r in the integral instead of $f(\theta)$ so make sure and substitute accordingly when doing the integral.

Let's take a look at an example.

Example 1

Determine the area of the inner loop of $r = 2 + 4\cos(\theta)$.

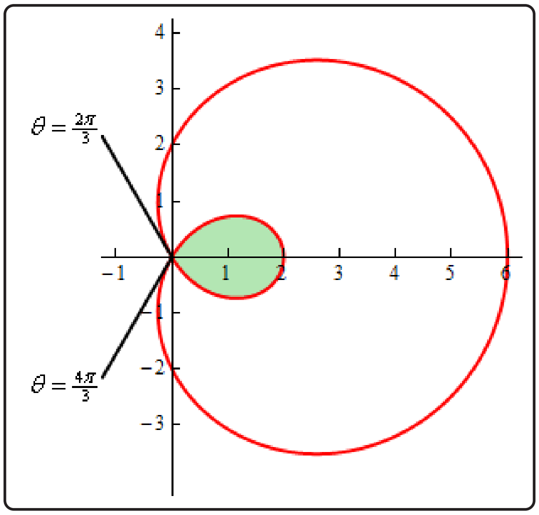
Solution

We graphed this function back when we first started looking at polar coordinates. For this problem we'll also need to know the values of θ where the curve goes through the origin.

We can get these by setting the equation equal to zero and solving.

$$0 = 2 + 4\cos(\theta)$$
$$\cos(\theta) = -\frac{1}{2} \qquad \Rightarrow \qquad \theta = \frac{2\pi}{3}, \frac{4\pi}{3}$$

Here is the sketch of this curve with the inner loop shaded in.



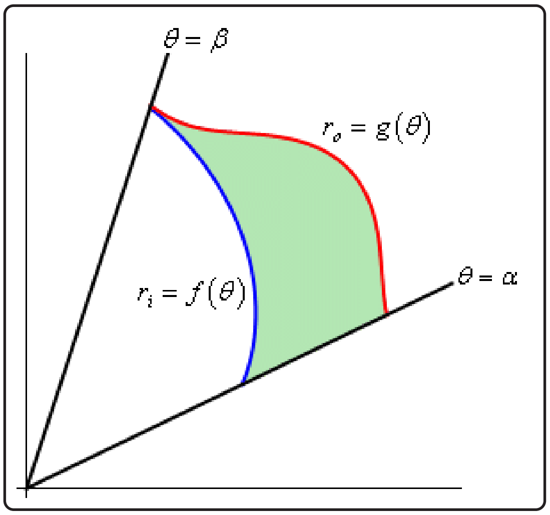
Can you see why we needed to know the values of θ where the curve goes through the origin? These points define where the inner loop starts and ends and hence are also the limits of integration in the formula.

So, the area is then,

$$\begin{aligned} A &= \int_{\frac{2\pi}{3}}^{\frac{4\pi}{3}} \frac{1}{2}\left(2 + 4\cos(\theta)\right)^2 d\theta \\ &= \int_{\frac{2\pi}{3}}^{\frac{4\pi}{3}} \frac{1}{2}\left(4 + 16\cos(\theta) + 16\cos^2(\theta)\right) d\theta \\ &= \int_{\frac{2\pi}{3}}^{\frac{4\pi}{3}} 2 + 8\cos(\theta) + 4\left(1 + \cos(2\theta)\right) d\theta \\ &= \int_{\frac{2\pi}{3}}^{\frac{4\pi}{3}} 6 + 8\cos(\theta) + 4\cos(2\theta)\, d\theta \\ &= \left(6\theta + 8\sin(\theta) + 2\sin(2\theta)\right)\Big|_{\frac{2\pi}{3}}^{\frac{4\pi}{3}} \\ &= 4\pi - 6\sqrt{3} = 2.174 \end{aligned}$$

You did follow the work done in this integral didn't you? You'll run into quite a few integrals of trig functions in this section so if you need to you should go back to the Integrals Involving Trig Functions sections and do a quick review.

So, that's how we determine areas that are enclosed by a single curve, but what about situations like the following sketch where we want to find the area between two curves.



In this case we can use the above formula to find the area enclosed by both and then the actual area is the difference between the two. The formula for this is,

Area Between Curves

$$A = \int_{\alpha}^{\beta} \frac{1}{2}\left(r_o^2 - r_i^2\right)\,d\theta$$

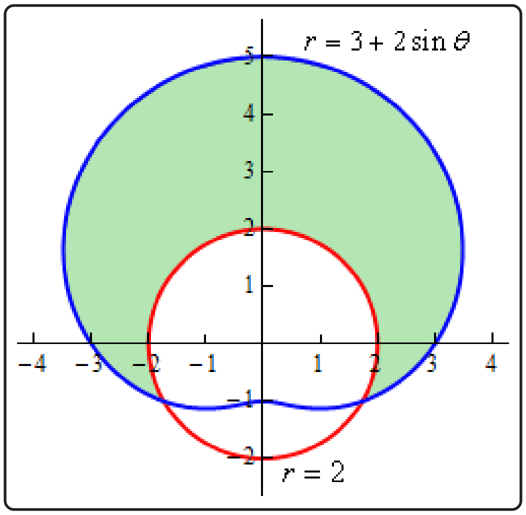
Let's take a look at an example of this.

Example 2

Determine the area that lies inside $r = 3 + 2\sin(\theta)$ and outside $r = 2$.

Solution

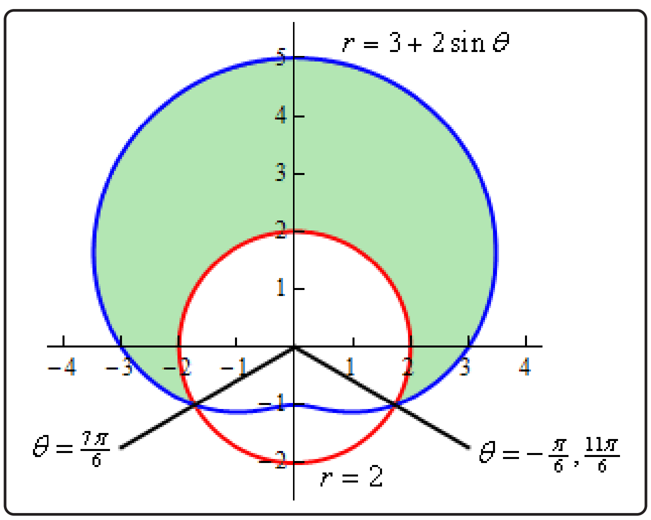
Here is a sketch of the region that we are after.



To determine this area, we'll need to know the values of θ for which the two curves intersect. We can determine these points by setting the two equations and solving.

$$\begin{aligned} 3 + 2\sin(\theta) &= 2 \\ \sin(\theta) &= -\frac{1}{2} \qquad \Rightarrow \qquad \theta = \frac{7\pi}{6}, \frac{11\pi}{6} \end{aligned}$$

Here is a sketch of the figure with these angles added.



Note as well here that we also acknowledged that another representation for the angle $\frac{11\pi}{6}$ is $-\frac{\pi}{6}$. This is important for this problem. In order to use the formula above the area must be enclosed as we increase from the smaller to larger angle. So, if we use $\frac{7\pi}{6}$ to $\frac{11\pi}{6}$ we will not enclose the shaded area, instead we will enclose the bottom most of the three regions. However, if we use the angles $-\frac{\pi}{6}$ to $\frac{7\pi}{6}$ we will enclose the area that we're after.

So, the area is then,

$$\begin{aligned} A &= \int_{-\frac{\pi}{6}}^{\frac{7\pi}{6}} \frac{1}{2}\left((3+2\sin(\theta))^2 - (2)^2\right) d\theta \\ &= \int_{-\frac{\pi}{6}}^{\frac{7\pi}{6}} \frac{1}{2}\left(5+12\sin(\theta)+4\sin^2(\theta)\right) d\theta \\ &= \int_{-\frac{\pi}{6}}^{\frac{7\pi}{6}} \frac{1}{2}\left(7+12\sin(\theta)-2\cos(2\theta)\right) d\theta \\ &= \frac{1}{2}\left(7\theta - 12\cos(\theta) - \sin(2\theta)\right)\Bigg|_{-\frac{\pi}{6}}^{\frac{7\pi}{6}} = \frac{11\sqrt{3}}{2} + \frac{14\pi}{3} = 24.187 \end{aligned}$$

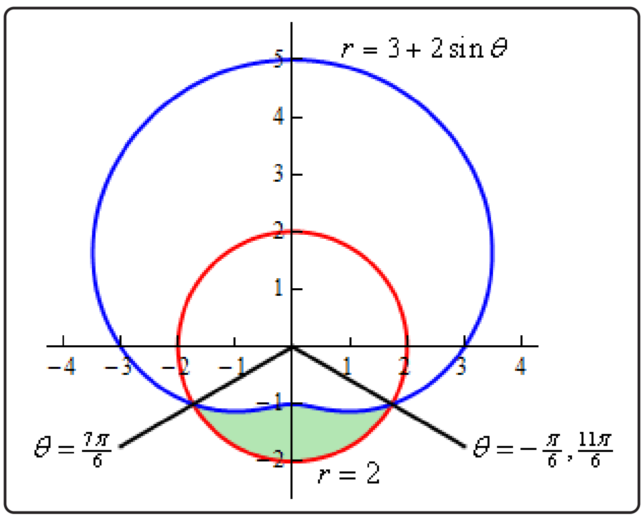
Let's work a slight modification of the previous example.

Example 3

Determine the area of the region outside $r = 3 + 2\sin(\theta)$ and inside $r = 2$.

Solution

This time we're looking for the following region.



So, this is the region that we get by using the limits $\frac{7\pi}{6}$ to $\frac{11\pi}{6}$. The area for this region is,

$$\begin{aligned} A &= \int_{\frac{7\pi}{6}}^{\frac{11\pi}{6}} \frac{1}{2}\left((2)^2 - (3 + 2\sin(\theta))^2\right) d\theta \\ &= \int_{\frac{7\pi}{6}}^{\frac{11\pi}{6}} \frac{1}{2}\left(-5 - 12\sin(\theta) - 4\sin^2(\theta)\right) d\theta \\ &= \int_{\frac{7\pi}{6}}^{\frac{11\pi}{6}} \frac{1}{2}\left(-7 - 12\sin(\theta) + 2\cos(2\theta)\right) d\theta \\ &= \frac{1}{2}\left(-7\theta + 12\cos(\theta) + \sin(2\theta)\right)\Big|_{\frac{7\pi}{6}}^{\frac{11\pi}{6}} = \frac{11\sqrt{3}}{2} - \frac{7\pi}{3} = 2.196 \end{aligned}$$

Notice that for this area the "outer" and "inner" function were opposite!

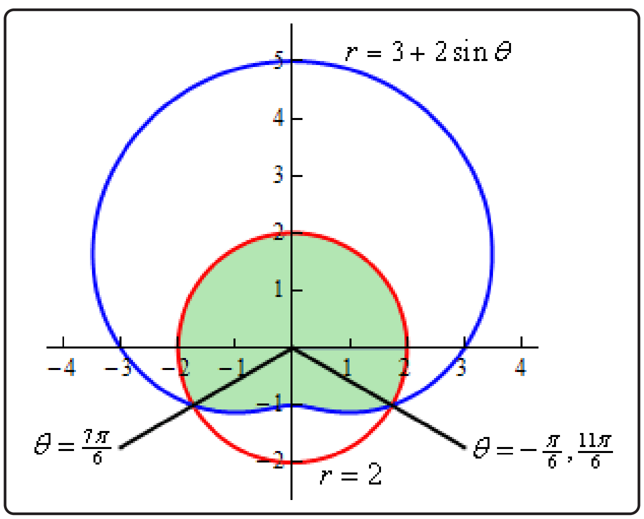
Let's do one final modification of this example.

Example 4

Determine the area that is inside both $r = 3 + 2\sin(\theta)$ and $r = 2$.

Solution

Here is the sketch for this example.



We are not going to be able to do this problem in the same fashion that we did the previous two. There is no set of limits that will allow us to enclose this area as we increase from one to the other. Remember that as we increase θ the area we're after must be enclosed.

However, the only two ranges for θ that we can work with enclose the area from the previous two examples and not this region.

In this case however, that is not a major problem. There are two ways to do get the area in this problem. We'll take a look at both of them.

Solution 1

In this case let's notice that the circle is divided up into two portions and we're after the upper portion. Also notice that we found the area of the lower portion in Example 3. Therefore, the area is,

$$\begin{aligned}\text{Area} &= \text{Area of Circle} - \text{Area from Example 3}\\ &= \pi(2)^2 - 2.196\\ &= 10.370\end{aligned}$$

Solution 2

In this case we do pretty much the same thing except this time we'll think of the area as the other portion of the limacon than the portion that we were dealing with in Example 2. We'll also need to actually compute the area of the limacon in this case.

So, the area using this approach is then,

$$\begin{aligned}\text{Area} &= \text{Area of Limacon} - \text{Area from Example 2}\\ &= \int_0^{2\pi} \frac{1}{2}\Big(3 + 2\sin(\theta)\Big)^2\, d\theta - 24.187\\ &= \int_0^{2\pi} \frac{1}{2}\Big(9 + 12\sin(\theta) + 4\sin^2(\theta)\Big)\, d\theta - 24.187\\ &= \int_0^{2\pi} \frac{1}{2}\Big(11 + 12\sin(\theta) - 2\cos(2\theta)\Big)\, d\theta - 24.187\\ &= \frac{1}{2}\Big(11\theta - 12\cos(\theta) - \sin(2\theta)\Big)\Big|_0^{2\pi} - 24.187\\ &= 11\pi - 24.187\\ &= 10.370\end{aligned}$$

A slightly longer approach, but sometimes we are forced to take this longer approach.

As this last example has shown we will not be able to get all areas in polar coordinates straight from an integral.

9.9 Arc Length with Polar Coordinates

We now need to move into the Calculus II applications of integrals and how we do them in terms of polar coordinates. In this section we'll look at the arc length of the curve given by,

$$r = f(\theta) \qquad \alpha \le \theta \le \beta$$

where we also assume that the curve is traced out exactly once. Just as we did with the tangent lines in polar coordinates we'll first write the curve in terms of a set of parametric equations,

$$\begin{aligned} x &= r\cos(\theta) & y &= r\sin(\theta) \\ &= f(\theta)\cos(\theta) & &= f(\theta)\sin(\theta) \end{aligned}$$

and we can now use the parametric formula for finding the arc length.

We'll need the following derivatives for these computations.

$$\begin{aligned} \frac{dx}{d\theta} &= f'(\theta)\cos(\theta) - f(\theta)\sin(\theta) & \frac{dy}{d\theta} &= f'(\theta)\sin(\theta) + f(\theta)\cos(\theta) \\ &= \frac{dr}{d\theta}\cos(\theta) - r\sin(\theta) & &= \frac{dr}{d\theta}\sin(\theta) + r\cos(\theta) \end{aligned}$$

We'll need the following for our ds.

$$\begin{aligned} \left(\frac{dx}{d\theta}\right)^2 + \left(\frac{dy}{d\theta}\right)^2 &= \left(\frac{dr}{d\theta}\cos(\theta) - r\sin(\theta)\right)^2 + \left(\frac{dr}{d\theta}\sin(\theta) + r\cos(\theta)\right)^2 \\ &= \left(\frac{dr}{d\theta}\right)^2\cos^2(\theta) - 2r\frac{dr}{d\theta}\cos(\theta)\sin(\theta) + r^2\sin^2(\theta) \\ &\qquad + \left(\frac{dr}{d\theta}\right)^2\sin^2(\theta) + 2r\frac{dr}{d\theta}\cos(\theta)\sin(\theta) + r^2\cos^2(\theta) \\ &= \left(\frac{dr}{d\theta}\right)^2\left(\cos^2(\theta) + \sin^2(\theta)\right) + r^2\left(\cos^2(\theta) + \sin^2(\theta)\right) \\ &= r^2 + \left(\frac{dr}{d\theta}\right)^2 \end{aligned}$$

The arc length formula for polar coordinates is then,

Arc Length with Polar Coordinates

$$L = \int ds$$

where,

$$ds = \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\, d\theta$$

Let's work a quick example of this.

Example 1

Determine the length of $r = \theta$, $0 \le \theta \le 1$.

Solution

Okay, let's just jump straight into the formula since this is a fairly simple function.

$$L = \int_0^1 \sqrt{\theta^2 + 1}\, d\theta$$

We'll need to use a trig substitution here.

$$\theta = \tan(x) \qquad d\theta = \sec^2(x)\, dx$$

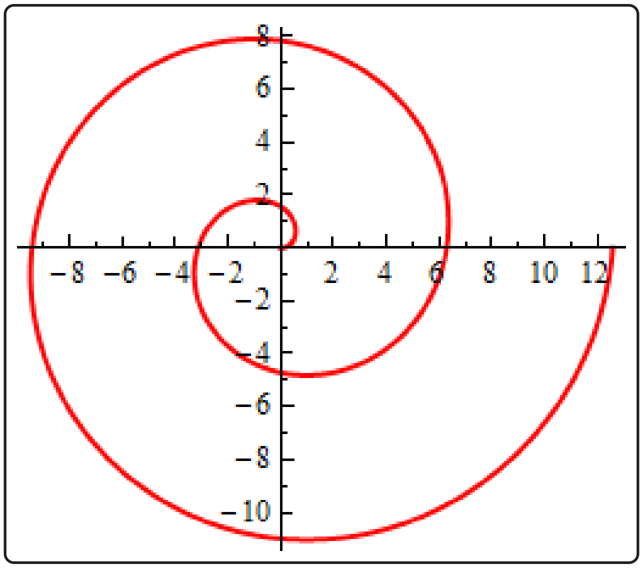
$$\begin{array}{lll} \theta = 0 & 0 = \tan(x) & x = 0 \\ \theta = 1 & 1 = \tan(x) & x = \dfrac{\pi}{4} \end{array}$$

$$\sqrt{\theta^2 + 1} = \sqrt{\tan^2(x) + 1} = \sqrt{\sec^2(x)} = |\sec(x)| = \sec(x)$$

The arc length is then,

$$\begin{aligned} L &= \int_0^1 \sqrt{\theta^2 + 1}\, d\theta \\ &= \int_0^{\frac{\pi}{4}} \sec^3(x)\, dx \\ &= \frac{1}{2}\Big(\sec(x)\tan(x) + \ln|\sec(x) + \tan(x)|\Big)\Big|_0^{\frac{\pi}{4}} \\ &= \frac{1}{2}\left(\sqrt{2} + \ln\left(1 + \sqrt{2}\right)\right) \end{aligned}$$

Just as an aside before we leave this chapter. The polar equation $r = \theta$ is the equation of a spiral. Here is a quick sketch of $r = \theta$ for $0 \le \theta \le 4\pi$.


8
6
4
2
−8
−6
−4
−2
2
4
6
8
10
12
−2
−4
−6
−8
−10

9.10 Surface Area with Polar Coordinates

We will be looking at surface area in polar coordinates in this section. Note however that all we're going to do is give the formulas for the surface area since most of these integrals tend to be fairly difficult.

We want to find the surface area of the region found by rotating,

$$r = f(\theta) \qquad \alpha \le \theta \le \beta$$

about the x or y-axis.

As we did in the tangent and arc length sections we'll write the curve in terms of a set of parametric equations.

$$\begin{aligned} x &= r\cos(\theta) & \qquad y &= r\sin(\theta) \\ &= f(\theta)\cos(\theta) & &= f(\theta)\sin(\theta) \end{aligned}$$

If we now use the parametric formula for finding the surface area we'll get,

Surface Area with Polar Coordinates

$$S = \int 2\pi y\, ds \qquad \text{rotation about } x - \text{axis}$$

$$S = \int 2\pi x\, ds \qquad \text{rotation about } y - \text{axis}$$

where,

$$ds = \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\, d\theta \qquad r = f(\theta), \quad \alpha \le \theta \le \beta$$

Note that because we will pick up a $d\theta$ from the ds we'll need to substitute one of the parametric equations in for x or y depending on the axis of rotation. This will often mean that the integrals will be somewhat unpleasant and so we will not be doing an example in this section.

9.11 Arc Length and Surface Area Revisited

We won't be working any examples in this section. This section is here solely for the purpose of summarizing up all the arc length and surface area problems.

Over the course of the last two chapters the topic of arc length and surface area has arisen many times and each time we got a new formula out of the mix. Students often get a little overwhelmed with all the formulas.

However, there really aren't as many formulas as it might seem at first glance. There is exactly one arc length formula and exactly two surface area formulas. These are,

Arc Length and Surface Area Formulas

$$L = \int ds$$

$$S = \int 2\pi y\, ds \qquad \text{rotation about } x - \text{axis}$$

$$S = \int 2\pi x\, ds \qquad \text{rotation about } y - \text{axis}$$

The problems arise because we have quite a few ds's that we can use. Again, students often have trouble deciding which one to use. The examples/problems usually suggest the correct one to use however. Here is a complete listing of all the ds's that we've seen and when they are used.

Various Formulas for ds

$$ds = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx \qquad \text{if } y = f\left(x\right),\ a \le x \le b$$

$$ds = \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy \qquad \text{if } x = h\left(y\right),\ c \le y \le d$$

$$ds = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt \qquad \text{if } x = f\left(t\right), y = g\left(t\right),\ \alpha \le t \le \beta$$

$$ds = \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\, d\theta \qquad \text{if } r = f\left(\theta\right),\ \alpha \le \theta \le \beta$$

Depending on the form of the function we can quickly tell which ds to use.

There is only one other thing to worry about in terms of the surface area formula. The ds will introduce a new differential to the integral. Before integrating make sure all the variables are in terms of this new differential. For example, if we have parametric equations we'll use the third ds and then we'll need to make sure and substitute for the x or y depending on which axis we rotate about to get everything in terms of t.

Likewise, if we have a function in the form $x = h(y)$ then we'll use the second ds and if the rotation is about the y-axis we'll need to substitute for the x in the integral. On the other hand, if we rotate about the x-axis we won't need to do a substitution for the y.

Keep these rules in mind and you'll always be able to determine which formula to use and how to correctly do the integral.

10 Series and Sequences

Once again, as with the last chapter, we are going to be looking at a completely different topic. The only material from previous chapters that will be needed here will be the ability to compute limits at infinity (we'll do a fair amount of these), compute the occasional derivative and integral. The integrals will, generally, be fairly simple and needing u substitutions every once in a while although we will see the occasional integral requiring integration by parts or partial fractions. So, basically, the material in this chapter doesn't rely all that much on previous material.

Series is one of those topics that many students don't find all that useful. To be honest, many students will never see series outside of their calculus class. However, series do play an important role in the field of ordinary differential equations and without series large portions of the field of partial differential equations would not be possible.

In other words, series is an important topic even if you won't ever see any of the applications. Most of the applications are beyond the scope of most Calculus courses and tend to occur in classes that many students don't take. So, as you go through this material keep in mind that these do have applications even if we won't really be covering many of them in this class.

The first topic we'll be looking at in this chapter is that of a sequence. We'll define just what we mean by a sequence and look at some basic topics and concepts that we'll need to work with them.

The other topic will be that of (infinite) series. In fact, we will spend the vast majority of this chapter deal with series. We can't, however, fully discuss series without understanding sequences and hence the reason for discussing sequences first. We will define just what an infinite series is and what it means for a series to converge or diverge. The majority of this chapter will then be spent discussing a variety of methods for testing whether or not a series will converge or diverge.

We'll close out the chapter with a discussion of power series and Taylor series as well as a couple of quick applications of series that we can easily discuss without needing any extra knowledge (as is needed for most applications of series).

10.1 Sequences

Let's start off this section with a discussion of just what a sequence is. A sequence is nothing more than a list of numbers written in a specific order. The list may or may not have an infinite number of terms in them although we will be dealing exclusively with infinite sequences in this class. General sequence terms are denoted as follows,

$$
\begin{aligned}
&a_1 - \text{first term}\\
&a_2 - \text{second term}\\
&\qquad\vdots\\
&a_n - n^{th}\text{ term}\\
&a_{n+1} - (n+1)^{\text{st}}\text{ term}\\
&\qquad\vdots
\end{aligned}
$$

Because we will be dealing with infinite sequences each term in the sequence will be followed by another term as noted above. In the notation above we need to be very careful with the subscripts. The subscript of $n+1$ denotes the next term in the sequence and NOT one plus the n^{th} term! In other words,

$$a_{n+1} \neq a_n + 1$$

so be very careful when writing subscripts to make sure that the "$+1$" doesn't migrate out of the subscript! This is an easy mistake to make when you first start dealing with this kind of thing.

There is a variety of ways of denoting a sequence. Each of the following are equivalent ways of denoting a sequence.

$$\{a_1, a_2, \ldots, a_n, a_{n+1}, \ldots\} \qquad \{a_n\} \qquad \{a_n\}_{n=1}^{\infty}$$

In the second and third notations above a_n is usually given by a formula.

A couple of notes are now in order about these notations. First, note the difference between the second and third notations above. If the starting point is not important or is implied in some way by the problem it is often not written down as we did in the third notation. Next, we used a starting point of $n = 1$ in the third notation only so we could write one down. There is absolutely no reason to believe that a sequence will start at $n = 1$. A sequence will start where ever it needs to start.

Let's take a look at a couple of sequences.

Example 1

Write down the first few terms of each of the following sequences.

(a) $\left\{\dfrac{n+1}{n^2}\right\}_{n=1}^{\infty}$

(b) $\left\{\dfrac{(-1)^{n+1}}{2^n}\right\}_{n=0}^{\infty}$

(c) $\{b_n\}_{n=1}^{\infty}$, where $b_n = n^{th}$ digit of π

Solution

(a) $\left\{\dfrac{n+1}{n^2}\right\}_{n=1}^{\infty}$

To get the first few sequence terms here all we need to do is plug in values of n into the formula given and we'll get the sequence terms.

$$\left\{\frac{n+1}{n^2}\right\}_{n=1}^{\infty} = \left\{\underbrace{2}_{n=1}, \underbrace{\frac{3}{4}}_{n=2}, \underbrace{\frac{4}{9}}_{n=3}, \underbrace{\frac{5}{16}}_{n=4}, \underbrace{\frac{6}{25}}_{n=5}, \ldots\right\}$$

Note the inclusion of the "..." at the end! This is an important piece of notation as it is the only thing that tells us that the sequence continues on and doesn't terminate at the last term.

(b) $\left\{\dfrac{(-1)^{n+1}}{2^n}\right\}_{n=0}^{\infty}$

This one is similar to the first one. The main difference is that this sequence doesn't start at $n = 1$.

$$\left\{\frac{(-1)^{n+1}}{2^n}\right\}_{n=0}^{\infty} = \left\{-1, \frac{1}{2}, -\frac{1}{4}, \frac{1}{8}, -\frac{1}{16}, \ldots\right\}$$

Note that the terms in this sequence alternate in signs. Sequences of this kind are sometimes called alternating sequences.

(c) $\{b_n\}_{n=1}^{\infty}$, where $b_n = n^{th}$ digit of π

This sequence is different from the first two in the sense that it doesn't have a specific formula for each term. However, it does tell us what each term should be. Each term should be the n^{th} digit of π. So we know that $\pi = 3.14159265359\ldots$

The sequence is then,

$$\{3, 1, 4, 1, 5, 9, 2, 6, 5, 3, 5, \ldots\}$$

In the first two parts of the previous example note that we were really treating the formulas as functions that can only have integers plugged into them. Or,

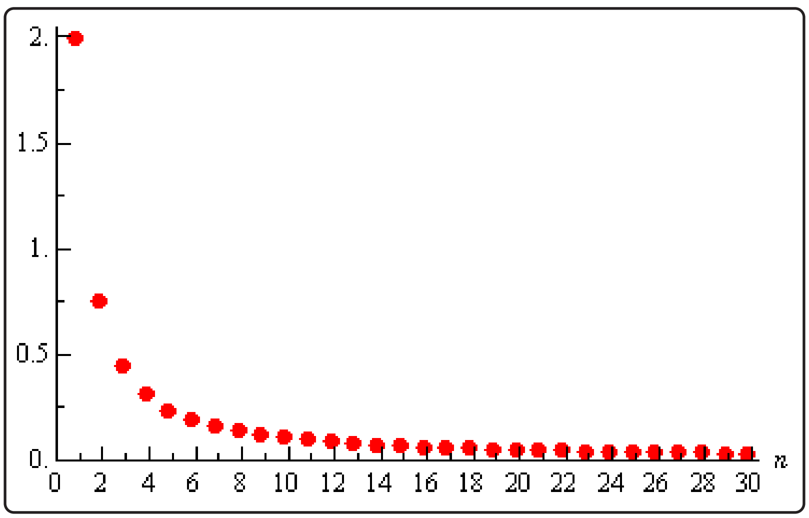
$$f(n) = \frac{n+1}{n^2} \qquad\qquad g(n) = \frac{(-1)^{n+1}}{2^n}$$

This is an important idea in the study of sequences (and series). Treating the sequence terms as function evaluations will allow us to do many things with sequences that we couldn't do otherwise. Before delving further into this idea however we need to get a couple more ideas out of the way.

First, we want to think about "graphing" a sequence. To graph the sequence $\{a_n\}$ we plot the points (n, a_n) as n ranges over all possible values on a graph. For instance, let's graph the sequence $\left\{\frac{n+1}{n^2}\right\}_{n=1}^{\infty}$. The first few points on the graph are,

$$(1,2), \quad \left(2, \frac{3}{4}\right), \quad \left(3, \frac{4}{9}\right), \quad \left(4, \frac{5}{16}\right), \quad \left(5, \frac{6}{25}\right), \quad \ldots$$

The graph, for the first 30 terms of the sequence, is then,



This graph leads us to an important idea about sequences. Notice that as n increases the sequence terms in our sequence, in this case, get closer and closer to zero. We then say that zero is the **limit** (or sometimes the **limiting value**) of the sequence and write,

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} \frac{n+1}{n^2} = 0$$

This notation should look familiar to you. It is the same notation we used when we talked about the limit of a function. In fact, if you recall, we said earlier that we could think of sequences as functions in some way and so this notation shouldn't be too surprising.

Using the ideas that we developed for limits of functions we can write down the following *working definition* for limits of sequences.

Working Definition of Limit

1. We say that
$$\lim_{n\to\infty} a_n = L$$
if we can make a_n as close to L as we want for all sufficiently large n. In other words, the value of the a_n's approach L as n approaches infinity.

2. We say that
$$\lim_{n\to\infty} a_n = \infty$$
if we can make a_n as large as we want for all sufficiently large n. Again, in other words, the value of the a_n's get larger and larger without bound as n approaches infinity.

3. We say that
$$\lim_{n\to\infty} a_n = -\infty$$
if we can make a_n as large and negative as we want for all sufficiently large n. Again, in other words, the value of the a_n's are negative and get larger and larger without bound as n approaches infinity.

The working definitions of the various sequence limits are nice in that they help us to visualize what the limit actually is. Just like with limits of functions however, there is also a precise definition for each of these limits. Let's give those before proceeding

Precise Definition of Limit

1. We say that $\lim\limits_{n\to\infty} a_n = L$ if for every number $\varepsilon > 0$ there is an integer N such that
$$|a_n - L| < \varepsilon \qquad \text{whenever} \qquad n > N$$

2. We say that $\lim\limits_{n\to\infty} a_n = \infty$ if for every number $M > 0$ there is an integer N such that
$$a_n > M \qquad \text{whenever} \qquad n > N$$

3. We say that $\lim\limits_{n\to\infty} a_n = -\infty$ if for every number $M < 0$ there is an integer N such that
$$a_n < M \qquad \text{whenever} \qquad n > N$$

We won't be using the precise definition often, but it will show up occasionally.

Note that both definitions tell us that in order for a limit to exist and have a finite value all the sequence terms must be getting closer and closer to that finite value as n increases.

Now that we have the definitions of the limit of sequences out of the way we have a bit of terminology that we need to look at. If $\lim\limits_{n\to\infty} a_n$ exists and is finite we say that the sequence is **convergent**. If $\lim\limits_{n\to\infty} a_n$ doesn't exist or is infinite we say the sequence **diverges**. Note that sometimes we will say the sequence **diverges to** ∞ if $\lim\limits_{n\to\infty} a_n = \infty$ and if $\lim\limits_{n\to\infty} a_n = -\infty$ we will sometimes say that the sequence **diverges to** $-\infty$.

Get used to the terms "convergent" and "divergent" as we'll be seeing them quite a bit throughout this chapter.

So just how do we find the limits of sequences? Most limits of most sequences can be found using one of the following theorems.

Theorem 1

Given the sequence $\{a_n\}$ if we have a function $f(x)$ such that $f(n) = a_n$ and $\lim\limits_{x\to\infty} f(x) = L$ then $\lim\limits_{n\to\infty} a_n = L$

This theorem is basically telling us that we take the limits of sequences much like we take the limit of functions. In fact, in most cases we'll not even really use this theorem by explicitly writing down a function. We will more often just treat the limit as if it were a limit of a function and take the limit as we always did back in Calculus I when we were taking the limits of functions.

So, now that we know that taking the limit of a sequence is nearly identical to taking the limit of a function we also know that all the properties from the limits of functions will also hold.

Properties

If $\{a_n\}$ and $\{b_n\}$ are both convergent sequences then,

1. $\lim\limits_{n\to\infty} (a_n \pm b_n) = \lim\limits_{n\to\infty} a_n \pm \lim\limits_{n\to\infty} b_n$
2. $\lim\limits_{n\to\infty} c a_n = c \lim\limits_{n\to\infty} a_n$
3. $\lim\limits_{n\to\infty} (a_n\, b_n) = \left(\lim\limits_{n\to\infty} a_n\right)\left(\lim\limits_{n\to\infty} b_n\right)$
4. $\lim\limits_{n\to\infty} \dfrac{a_n}{b_n} = \dfrac{\lim\limits_{n\to\infty} a_n}{\lim\limits_{n\to\infty} b_n}$, provided $\lim\limits_{n\to\infty} b_n \neq 0$
5. $\lim\limits_{n\to\infty} a_n^p = \left[\lim\limits_{n\to\infty} a_n\right]^p$ provided $a_n \geq 0$

These properties can be proved using Theorem 1 above and the function limit properties we saw in Calculus I or we can prove them directly using the precise definition of a limit using nearly identical proofs of the function limit properties.

Next, just as we had a Squeeze Theorem for function limits we also have one for sequences and it is pretty much identical to the function limit version.

Squeeze Theorem for Sequences

If $a_n \le c_n \le b_n$ for all $n > N$ for some N and $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} b_n = L$ then $\lim\limits_{n\to\infty} c_n = L$.

Note that in this theorem the "for all $n > N$ for some N" is really just telling us that we need to have $a_n \le c_n \le b_n$ for all sufficiently large n, but if it isn't true for the first few n that won't invalidate the theorem.

As we'll see not all sequences can be written as functions that we can actually take the limit of. This will be especially true for sequences that alternate in signs. While we can always write these sequence terms as a function we simply don't know how to take the limit of a function like that. The following theorem will help with some of these sequences.

Theorem 2

If $\lim\limits_{n\to\infty} |a_n| = 0$ then $\lim\limits_{n\to\infty} a_n = 0$.

Note that in order for this theorem to hold the limit MUST be zero and it won't work for a sequence whose limit is not zero. This theorem is easy enough to prove so let's do that.

Proof of Theorem 2

The main thing to this proof is to note that,

$$-|a_n| \le a_n \le |a_n|$$

Then note that,

$$\lim_{n\to\infty} (-|a_n|) = -\lim_{n\to\infty} |a_n| = 0$$

We then have $\lim_{n\to\infty} (-|a_n|) = \lim_{n\to\infty} |a_n| = 0$ and so by the Squeeze Theorem we must also have,

$$\lim_{n\to\infty} a_n = 0$$

The next theorem is a useful theorem giving the convergence/divergence and value (for when it's convergent) of a sequence that arises on occasion.

Theorem 3

The sequence $\{r^n\}_{n=0}^{\infty}$ converges if $-1 < r \le 1$ and diverges for all other values of r. Also,

$$\lim_{n\to\infty} r^n = \begin{cases} 0 & \text{if } -1 < r < 1 \\ 1 & \text{if } r = 1 \end{cases}$$

Here is a quick (well not so quick, but definitely simple) partial proof of this theorem.

Partial Proof of Theorem 3

We'll do this by a series of cases although the last case will not be completely proven.

Case 1 : $r > 1$

We know from Calculus I that $\lim_{x\to\infty} r^x = \infty$ if $r > 1$ and so by Theorem 1 above we also know that $\lim_{n\to\infty} r^n = \infty$ and so the sequence diverges if $r > 1$.

Case 2 : $r = 1$

In this case we have,

$$\lim_{n\to\infty} r^n = \lim_{n\to\infty} 1^n = \lim_{n\to\infty} 1 = 1$$

So, the sequence converges for $r = 1$ and in this case its limit is 1.

Case 3 : $0 < r < 1$

We know from Calculus I that $\lim_{x\to\infty} r^x = 0$ if $0 < r < 1$ and so by Theorem 1 above we also know that $\lim_{n\to\infty} r^n = 0$ and so the sequence converges if $0 < r < 1$ and in this case its limit is zero.

Case 4 : $r = 0$

In this case we have,

$$\lim_{n\to\infty} r^n = \lim_{n\to\infty} 0^n = \lim_{n\to\infty} 0 = 0$$

So, the sequence converges for $r = 0$ and in this case its limit is zero.

Case 5 : $-1 < r < 0$

First let's note that if $-1 < r < 0$ then $0 < |r| < 1$ then by Case 3 above we have,

$$\lim_{n\to\infty} |r^n| = \lim_{n\to\infty} |r|^n = 0$$

Theorem 2 above now tells us that we must also have, $\lim_{n\to\infty} r^n = 0$ and so if $-1 < r < 0$ the sequence converges and has a limit of 0.

Case 6 : $r = -1$

In this case the sequence is,

$$\{r^n\}_{n=0}^{\infty} = \{(-1)^n\}_{n=0}^{\infty} = \{1, -1, 1, -1, 1, -1, 1, -1, \ldots\}_{n=0}^{\infty}$$

and hopefully it is clear that $\lim_{n\to\infty} (-1)^n$ doesn't exist. Recall that in order of this limit to exist the terms must be approaching a single value as n increases. In this case however the terms just alternate between 1 and -1 and so the limit does not exist.

So, the sequence diverges for $r = -1$.

Case 7: $r < -1$

In this case we're not going to go through a complete proof. Let's just see what happens if we let $r = -2$ for instance. If we do that the sequence becomes,

$$\{r^n\}_{n=0}^{\infty} = \{(-2)^n\}_{n=0}^{\infty} = \{1, -2, 4, -8, 16, -32, \ldots\}_{n=0}^{\infty}$$

So, if $r = -2$ we get a sequence of terms whose values alternate in sign and get larger and larger and so $\lim_{n\to\infty} (-2)^n$ doesn't exist. It does not settle down to a single value as n increases nor do the terms ALL approach infinity. So, the sequence diverges for $r = -2$.

We could do something similar for any value of r such that $r < -1$ and so the sequence diverges for $r < -1$.

Let's take a look at a couple of examples of limits of sequences.

Example 2

Determine if the following sequences converge or diverge. If the sequence converges determine its limit.

(a) $\left\{ \dfrac{3n^2 - 1}{10n + 5n^2} \right\}_{n=2}^{\infty}$

(b) $\left\{ \dfrac{\mathbf{e}^{2n}}{n} \right\}_{n=1}^{\infty}$

(c) $\left\{ \dfrac{(-1)^n}{n} \right\}_{n=1}^{\infty}$

(d) $\left\{ (-1)^n \right\}_{n=0}^{\infty}$

Solution

(a) $\left\{ \dfrac{3n^2 - 1}{10n + 5n^2} \right\}_{n=2}^{\infty}$

In this case all we need to do is recall the method that was developed in Calculus I to deal with the limits of rational functions. See the Limits At Infinity, Part I section of the Calculus I notes for a review of this if you need to.

To do a limit in this form all we need to do is factor from the numerator and denominator the largest power of n, cancel and then take the limit.

$$\lim_{n \to \infty} \frac{3n^2 - 1}{10n + 5n^2} = \lim_{n \to \infty} \frac{n^2 \left(3 - \frac{1}{n^2}\right)}{n^2 \left(\frac{10}{n} + 5\right)} = \lim_{n \to \infty} \frac{3 - \frac{1}{n^2}}{\frac{10}{n} + 5} = \frac{3}{5}$$

So, the sequence converges and its limit is $\frac{3}{5}$.

(b) $\left\{ \dfrac{\mathbf{e}^{2n}}{n} \right\}_{n=1}^{\infty}$

We will need to be careful with this one. We will need to use L'Hospital's Rule on this sequence. The problem is that L'Hospital's Rule only works on functions and not on sequences. Normally this would be a problem, but we've got Theorem 1 from above to help us out. Let's define

$$f(x) = \frac{\mathbf{e}^{2x}}{x}$$

and note that,

$$f(n) = \frac{\mathbf{e}^{2n}}{n}$$

Theorem 1 says that all we need to do is take the limit of the function.

$$\lim_{n\to\infty} \frac{\mathbf{e}^{2n}}{n} = \lim_{x\to\infty} \frac{\mathbf{e}^{2x}}{x} = \lim_{x\to\infty} \frac{2\mathbf{e}^{2x}}{1} = \infty$$

So, the sequence in this part diverges (to ∞).

More often than not we just do L'Hospital's Rule on the sequence terms without first converting to x's since the work will be identical regardless of whether we use x or n. However, we really should remember that technically we can't do the derivatives while dealing with sequence terms.

(c) $\left\{ \frac{(-1)^n}{n} \right\}_{n=1}^{\infty}$

We will also need to be careful with this sequence. We might be tempted to just say that the limit of the sequence terms is zero (and we'd be correct). However, technically we can't take the limit of sequences whose terms alternate in sign, because we don't know how to do limits of functions that exhibit that same behavior. Also, we want to be very careful to not rely too much on intuition with these problems. As we will see in the next section, and in later sections, our intuition can lead us astray in these problems if we aren't careful.

So, let's work this one by the book. We will need to use Theorem 2 on this problem. To this we'll first need to compute,

$$\lim_{n\to\infty} \left| \frac{(-1)^n}{n} \right| = \lim_{n\to\infty} \frac{1}{n} = 0$$

Therefore, since the limit of the sequence terms with absolute value bars on them goes to zero we know by Theorem 2 that,

$$\lim_{n\to\infty} \frac{(-1)^n}{n} = 0$$

which also means that the sequence converges to a value of zero.

(d) $\left\{ (-1)^n \right\}_{n=0}^{\infty}$

For this theorem note that all we need to do is realize that this is the sequence in Theorem 3 above using $r = -1$. So, by Theorem 3 this sequence diverges.

We now need to give a warning about misusing Theorem 2. Theorem 2 only works if the limit is zero. If the limit of the absolute value of the sequence terms is not zero then the theorem will not hold. The last part of the previous example is a good example of this (and in fact this warning is the whole reason that part is there). Notice that

$$\lim_{n\to\infty} |(-1)^n| = \lim_{n\to\infty} 1 = 1$$

and yet, $\lim_{n\to\infty} (-1)^n$ doesn't even exist let alone equal 1. So, be careful using this Theorem 2. You must always remember that it only works if the limit is zero.

Before moving onto the next section we need to give one more theorem that we'll need for a proof down the road.

Theorem 4

For the sequence $\{a_n\}$ if both $\lim_{n\to\infty} a_{2n} = L$ and $\lim_{n\to\infty} a_{2n+1} = L$ then $\{a_n\}$ is convergent and $\lim_{n\to\infty} a_n = L$.

Proof of Theorem 4

Let $\varepsilon > 0$.

Then since $\lim_{n\to\infty} a_{2n} = L$ there is an $N_1 > 0$ such that if $n > N_1$ we know that,

$$|a_{2n} - L| < \varepsilon$$

Likewise, because $\lim_{n\to\infty} a_{2n+1} = L$ there is an $N_2 > 0$ such that if $n > N_2$ we know that,

$$|a_{2n+1} - L| < \varepsilon$$

Now, let $N = \max\{2N_1, 2N_2 + 1\}$ and let $n > N$. Then either $a_n = a_{2k}$ for some $k > N_1$ or $a_n = a_{2k+1}$ for some $k > N_2$ and so in either case we have that,

$$|a_n - L| < \varepsilon$$

Therefore, $\lim_{n\to\infty} a_n = L$ and so $\{a_n\}$ is convergent.

10.2 More on Sequences

In the previous section we introduced the concept of a sequence and talked about limits of sequences and the idea of convergence and divergence for a sequence. In this section we want to take a quick look at some ideas involving sequences.

Let's start off with some terminology and definitions.

Definition

Given any sequence $\{a_n\}$ we have the following.

1. We call the sequence **increasing** if $a_n < a_{n+1}$ for every n.
2. We call the sequence **decreasing** if $a_n > a_{n+1}$ for every n.
3. If $\{a_n\}$ is an increasing sequence or $\{a_n\}$ is a decreasing sequence we call it **monotonic**.
4. If there exists a number m such that $m \leq a_n$ for every n we say the sequence is **bounded below**. The number m is sometimes called a **lower bound** for the sequence.
5. If there exists a number M such that $a_n \leq M$ for every n we say the sequence is **bounded above**. The number M is sometimes called an **upper bound** for the sequence.
6. If the sequence is both bounded below and bounded above we call the sequence **bounded**.

Note that in order for a sequence to be increasing or decreasing it must be increasing/decreasing for every n. In other words, a sequence that increases for three terms and then decreases for the rest of the terms is NOT a decreasing sequence! Also note that a monotonic sequence must always increase or it must always decrease.

Before moving on we should make a quick point about the bounds for a sequence that is bounded above and/or below. We'll make the point about lower bounds, but we could just as easily make it about upper bounds.

A sequence is bounded below if we can find any number m such that $m \leq a_n$ for every n. Note however that if we find one number m to use for a lower bound then any number smaller than m will also be a lower bound. Also, just because we find one lower bound that doesn't mean there won't be a "better" lower bound for the sequence than the one we found. In other words, there are an infinite number of lower bounds for a sequence that is bounded below, some will be better than others. In my class all that I'm after will be a lower bound. I don't necessarily need the best lower bound, just a number that will be a lower bound for the sequence.

Let's take a look at a couple of examples.

Example 1

Determine if the following sequences are monotonic and/or bounded.

(a) $\left\{-n^2\right\}_{n=0}^{\infty}$

(b) $\left\{(-1)^{n+1}\right\}_{n=1}^{\infty}$

(c) $\left\{\dfrac{2}{n^2}\right\}_{n=5}^{\infty}$

Solution

(a) $\left\{-n^2\right\}_{n=0}^{\infty}$

This sequence is a decreasing sequence (and hence monotonic) because,

$$-n^2 > -(n+1)^2 \qquad \text{for every } n$$

Also, since the sequence terms will be either zero or negative this sequence is bounded above. We can use any positive number or zero as the bound, M, however, it's standard to choose the smallest possible bound if we can and it's a nice number. So, we'll choose $M = 0$ since,

$$-n^2 \le 0 \qquad \text{for every } n$$

This sequence is not bounded below however since we can always get below any potential bound by taking n large enough. Therefore, while the sequence is bounded above it is not bounded.

As a side note we can also note that this sequence diverges (to $-\infty$ if we want to be specific).

(b) $\left\{(-1)^{n+1}\right\}_{n=1}^{\infty}$

The sequence terms in this sequence alternate between 1 and -1 and so the sequence is neither an increasing sequence or a decreasing sequence. Since the sequence is neither an increasing nor decreasing sequence it is not a monotonic sequence.

The sequence is bounded however since it is bounded above by 1 and bounded below by -1.

Again, we can note that this sequence is also divergent.

(c) $\left\{ \dfrac{2}{n^2} \right\}_{n=5}^{\infty}$

This sequence is a decreasing sequence (and hence monotonic) since,

$$\frac{2}{n^2} > \frac{2}{(n+1)^2}$$

The terms in this sequence are all positive and so it is bounded below by zero. Also, since the sequence is a decreasing sequence the first sequence term will be the largest and so we can see that the sequence will also be bounded above by $\frac{2}{25}$. Therefore, this sequence is bounded.

We can also take a quick limit and note that this sequence converges and its limit is zero.

Now, let's work a couple more examples that are designed to make sure that we don't get too used to relying on our intuition with these problems. As we noted in the previous section our intuition can often lead us astray with some of the concepts we'll be looking at in this chapter.

Example 2

Determine if the following sequences are monotonic and/or bounded.

(a) $\left\{ \dfrac{n}{n+1} \right\}_{n=1}^{\infty}$

(b) $\left\{ \dfrac{n^3}{n^4 + 10000} \right\}_{n=0}^{\infty}$

Solution

(a) $\left\{ \dfrac{n}{n+1} \right\}_{n=1}^{\infty}$

We'll start with the bounded part of this example first and then come back and deal with the increasing/decreasing question since that is where students often make mistakes with this type of sequence.

First, n is positive and so the sequence terms are all positive. The sequence is therefore bounded below by zero. Likewise, each sequence term is the quotient of a number divided by a larger number and so is guaranteed to be less than one. The sequence is then bounded above by one. So, this sequence is bounded.

Now let's think about the monotonic question. First, students will often make the mistake of assuming that because the denominator is larger the quotient must be decreasing. This will not always be the case and in this case we would be wrong. This sequence is increasing as we'll see.

To determine the increasing/decreasing nature of this sequence we will need to resort to Calculus I techniques. First consider the following function and its derivative.

$$f(x) = \frac{x}{x+1} \qquad\qquad f'(x) = \frac{1}{(x+1)^2}$$

We can see that the first derivative is always positive and so from Calculus I we know that the function must then be an increasing function. So, how does this help us? Notice that,

$$f(n) = \frac{n}{n+1} = a_n$$

Therefore because $n < n+1$ and $f(x)$ is increasing we can also say that,

$$a_n = \frac{n}{n+1} = f(n) < f(n+1) = \frac{n+1}{n+1+1} = a_{n+1} \qquad \Rightarrow \qquad a_n < a_{n+1}$$

In other words, the sequence must be increasing.

Note that now that we know the sequence is an increasing sequence we can get a better lower bound for the sequence. Since the sequence is increasing the first term in the sequence must be the smallest term and so since we are starting at $n = 1$ we could also use a lower bound of $\frac{1}{2}$ for this sequence. It is important to remember that any number that is always less than or equal to all the sequence terms can be a lower bound. Some are better than others however.

A quick limit will also tell us that this sequence converges with a limit of 1.

Before moving on to the next part there is a natural question that many students will have at this point. Why did we use Calculus to determine the increasing/decreasing nature of the sequence when we could have just plugged in a couple of n's and quickly determined the same thing?

The answer to this question is the next part of this example!

(b) $\left\{ \dfrac{n^3}{n^4 + 10000} \right\}_{n=0}^{\infty}$

This is a messy looking sequence, but it needs to be in order to make the point of this part.

First, notice that, as with the previous part, the sequence terms are all positive and will all be less than one (since the numerator is guaranteed to be less than the denominator) and so the sequence is bounded.

Now, let's move on to the increasing/decreasing question. As with the last problem, many students will look at the exponents in the numerator and denominator and determine based on that that sequence terms must decrease.

This however, isn't a decreasing sequence. Let's take a look at the first few terms to see this.

$$a_1 = \frac{1}{10001} \approx 0.00009999 \qquad a_2 = \frac{1}{1252} \approx 0.0007987$$
$$a_3 = \frac{27}{10081} \approx 0.005678 \qquad a_4 = \frac{4}{641} \approx 0.006240$$
$$a_5 = \frac{1}{85} \approx 0.011756 \qquad a_6 = \frac{27}{1412} \approx 0.019122$$
$$a_7 = \frac{343}{12401} \approx 0.02766 \qquad a_8 = \frac{32}{881} \approx 0.03632$$
$$a_9 = \frac{729}{16561} \approx 0.04402 \qquad a_{10} = \frac{1}{20} = 0.05$$

The first 10 terms of this sequence are all increasing and so clearly the sequence can't be a decreasing sequence. Recall that a sequence can only be decreasing if ALL the terms are decreasing.

Now, we can't make another common mistake and assume that because the first few terms increase then whole sequence must also increase. If we did that we would also be mistaken as this is also not an increasing sequence.

This sequence is neither decreasing or increasing. The only sure way to see this is to do the Calculus I approach to increasing/decreasing functions.

In this case we'll need the following function and its derivative.

$$f(x) = \frac{x^3}{x^4 + 10000} \qquad f'(x) = \frac{-x^2\left(x^4 - 30000\right)}{\left(x^4 + 10000\right)^2}$$

This function will have the following three critical points,

$$x = 0, \;\; x = \sqrt[4]{30000} \approx 13.1607, \qquad x = -\sqrt[4]{30000} \approx -13.1607$$

Why critical points? Remember these are the only places where the derivative *may* change sign! Our sequence starts at $n = 0$ and so we can ignore the third one since it lies outside the values of n that we're considering. By plugging in some test values of x we can quickly determine that the derivative is positive for $0 < x < \sqrt[4]{30000} \approx 13.16$ and so the function is increasing in this range. Likewise, we can see that the derivative is negative for $x > \sqrt[4]{30000} \approx 13.16$ and so the function will be decreasing in this range.

So, our sequence will be increasing for $0 \le n \le 13$ and decreasing for $n \ge 13$. Therefore, the function is not monotonic.

Finally, note that this sequence will also converge and has a limit of zero.

So, as the last example has shown we need to be careful in making assumptions about sequences. Our intuition will often not be sufficient to get the correct answer and we can NEVER make assumptions about a sequence based on the value of the first few terms. As the last part has shown there are sequences which will increase or decrease for a few terms and then change direction after that.

Note as well that we said “first few terms” here, but it is completely possible for a sequence to decrease for the first 10,000 terms and then start increasing for the remaining terms. In other words, there is no “magical” value of n for which all we have to do is check up to that point and then we'll know what the whole sequence will do.

The only time that we'll be able to avoid using Calculus I techniques to determine the increasing/decreasing nature of a sequence is in sequences like part (c) of Example 1. In this case increasing n only changed (in fact increased) the denominator and so we were able to determine the behavior of the sequence based on that.

In Example 2 however, increasing n increased both the denominator and the numerator. In cases like this there is no way to determine which increase will “win out” and cause the sequence terms to increase or decrease and so we need to resort to Calculus I techniques to answer the question.

We'll close out this section with a nice theorem that we'll use in some of the proofs later in this chapter.

Theorem

If $\{a_n\}$ is bounded and monotonic then $\{a_n\}$ is convergent.

Be careful to not misuse this theorem. It does not say that if a sequence is not bounded and/or not monotonic that it is divergent. Example 2b is a good case in point. The sequence in that example was not monotonic but it does converge.

Note as well that we can make several variants of this theorem. If $\{a_n\}$ is bounded above and increasing then it converges and likewise if $\{a_n\}$ is bounded below and decreasing then it converges.

10.3 Series - Basics

In this section we will introduce the topic that we will be discussing for the rest of this chapter. That topic is infinite series. So just what is an infinite series? Well, let's start with a sequence $\{a_n\}_{n=1}^{\infty}$ (note the $n = 1$ is for convenience, it can be anything) and define the following,

$$\begin{aligned}
s_1 &= a_1 \\
s_2 &= a_1 + a_2 \\
s_3 &= a_1 + a_2 + a_3 \\
s_4 &= a_1 + a_2 + a_3 + a_4 \\
&\ \vdots \\
s_n &= a_1 + a_2 + a_3 + a_4 + \cdots + a_n = \sum_{i=1}^{n} a_i
\end{aligned}$$

The s_n are called **partial sums** and notice that they will form a sequence, $\{s_n\}_{n=1}^{\infty}$. Also recall that the Σ is used to represent this summation and called a variety of names. The most common names are : **series notation**, **summation notation**, and **sigma notation**.

You should have seen this notation, at least briefly, back when you saw the definition of a definite integral in Calculus I. If you need a quick refresher on summation notation see the review of summation notation in the Calculus I notes.

Now back to series. We want to take a look at the limit of the sequence of partial sums, $\{s_n\}_{n=1}^{\infty}$. To make the notation go a little easier we'll define,

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} \sum_{i=1}^{n} a_i = \sum_{i=1}^{\infty} a_i$$

We will call $\sum\limits_{i=1}^{\infty} a_i$ an **infinite series** and note that the series "starts" at $i = 1$ because that is where our original sequence, $\{a_n\}_{n=1}^{\infty}$, started. Had our original sequence started at 2 then our infinite series would also have started at 2. The infinite series will start at the same value that the sequence of terms (as opposed to the sequence of partial sums) starts.

It is important to note that $\sum\limits_{i=1}^{\infty} a_i$ is really nothing more than a convenient notation for $\lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} a_i$ so we do not need to keep writing the limit down. We do, however, always need to remind ourselves that we really do have a limit there!

If the sequence of partial sums, $\{s_n\}_{n=1}^{\infty}$, is convergent and its limit is finite then we also call the infinite series, $\sum\limits_{i=1}^{\infty} a_i$ **convergent** and if the sequence of partial sums is divergent then the infinite series is also called **divergent**.

Note that sometimes it is convenient to write the infinite series as,

$$\sum_{i=1}^{\infty} a_i = a_1 + a_2 + a_3 + \cdots + a_n + \cdots$$

We do have to be careful with this however. This implies that an infinite series is just an infinite sum of terms and as we'll see in the next section this is not really true for many series.

In the next section we're going to be discussing in greater detail the value of an infinite series, provided it has one of course, as well as the ideas of convergence and divergence.

This section is going to be devoted mostly to notational issues as well as making sure we can do some basic manipulations with infinite series so we are ready for them when we need to be able to deal with them in later sections.

First, we should note that in most of this chapter we will refer to infinite series as simply series. If we ever need to work with both infinite and finite series we'll be more careful with terminology, but in most sections we'll be dealing exclusively with infinite series and so we'll just call them series.

Now, in $\sum\limits_{i=1}^{\infty} a_i$ the i is called the **index of summation** or just **index** for short and note that the letter we use to represent the index does not matter. So for example the following series are all the same. The only difference is the letter we've used for the index.

$$\sum_{i=0}^{\infty} \frac{3}{i^2+1} = \sum_{k=0}^{\infty} \frac{3}{k^2+1} = \sum_{n=0}^{\infty} \frac{3}{n^2+1} \quad etc.$$

It is important to again note that the index will start at whatever value the sequence of series terms starts at and this can literally be anything. So far we've used $n = 0$ and $n = 1$ but the index could have started anywhere. In fact, we will usually use $\sum a_n$ to represent an infinite series in which the starting point for the index is not important. When we drop the initial value of the index we'll also drop the infinity from the top so don't forget that it is still technically there.

We will be dropping the initial value of the index in quite a few facts and theorems that we'll be seeing throughout this chapter. In these facts/theorems the starting point of the series will not affect the result and so to simplify the notation and to avoid giving the impression that the starting point is important we will drop the index from the notation. Do not forget however, that there is a starting point and that this will be an infinite series.

Note however, that if we do put an initial value of the index on a series in a fact/theorem it is there because it really does need to be there.

Now that some of the notational issues are out of the way we need to start thinking about various ways that we can manipulate series.

We'll start this off with basic arithmetic with infinite series as we'll need to be able to do that on occasion. We have the following properties.

Properties

If $\sum a_n$ and $\sum b_n$ are both convergent series then,

1. $\sum ca_n$, where c is any number, is also convergent and

$$\sum ca_n = c\sum a_n$$

2. $\sum_{n=k}^{\infty} a_n \pm \sum_{n=k}^{\infty} b_n$ is also convergent and,

$$\sum_{n=k}^{\infty} a_n \pm \sum_{n=k}^{\infty} b_n = \sum_{n=k}^{\infty} (a_n \pm b_n)$$

The first property is simply telling us that we can always factor a multiplicative constant out of an infinite series and again recall that if we don't put in an initial value of the index that the series can start at any value. Also recall that in these cases we won't put an infinity at the top either.

The second property says that if we add/subtract series all we really need to do is add/subtract the series terms. Note as well that in order to add/subtract series we need to make sure that both have the same initial value of the index and the new series will also start at this value.

Before we move on to a different topic let's discuss multiplication of series briefly. We'll start both series at $n = 0$ for a later formula and then note that,

$$\left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right) \neq \sum_{n=0}^{\infty} (a_n b_n)$$

To convince yourself that this isn't true consider the following product of two finite sums.

$$(2 + x)\left(3 - 5x + x^2\right) = 6 - 7x - 3x^2 + x^3$$

Yeah, it was just the multiplication of two polynomials. Each is a finite sum and so it makes the point. In doing the multiplication we didn't just multiply the constant terms, then the x terms, *etc*. Instead we had to distribute the 2 through the second polynomial, then distribute the x through the second polynomial and finally combine like terms.

Multiplying infinite series (even though we said we can't think of an infinite series as an infinite sum) needs to be done in the same manner. With multiplication we're really asking us to do the following,

$$\left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right) = (a_0 + a_1 + a_2 + a_3 + \cdots)(b_0 + b_1 + b_2 + b_3 + \cdots)$$

To do this multiplication we would have to distribute the a_0 through the second term, distribute the a_1 through, *etc* then combine like terms. This is pretty much impossible since both series have an infinite set of terms in them, however the following formula can be used to determine the product of two series.

$$\left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right) = \sum_{n=0}^{\infty} c_n \quad \text{where} \quad c_n = \sum_{i=0}^{n} a_i b_{n-i}$$

We also can't say a lot about the convergence of the product. Even if both of the original series are convergent it is possible for the product to be divergent. The reality is that multiplication of series is a somewhat difficult process and in general is avoided if possible. We will take a brief look at it towards the end of the chapter when we've got more work under our belt and we run across a situation where it might actually be what we want to do. Until then, don't worry about multiplying series.

The next topic that we need to discuss in this section is that of **index shift**. To be honest this is not a topic that we'll see all that often in this course. In fact, we'll use it once in the next section and then not use it again in all likelihood. Despite the fact that we won't use it much in this course doesn't mean however that it isn't used often in other classes where you might run across series. So, we will cover it briefly here so that you can say you've seen it.

The basic idea behind index shifts is to start a series at a different value for whatever the reason (and yes, there are legitimate reasons for doing that).

Consider the following series,

$$\sum_{n=2}^{\infty} \frac{n+5}{2^n}$$

Suppose that for some reason we wanted to start this series at $n = 0$, but we didn't want to change the value of the series. This means that we can't just change the $n = 2$ to $n = 0$ as this would add in two new terms to the series and thus change its value.

Performing an index shift is a fairly simple process to do. We'll start by defining a new index, say i, as follows,

$$i = n - 2$$

Now, when $n = 2$, we will get $i = 0$. Notice as well that if $n = \infty$ then $i = \infty - 2 = \infty$, so only the lower limit will change here. Next, we can solve this for n to get,

$$n = i + 2$$

We can now completely rewrite the series in terms of the index i instead of the index n simply by plugging in our equation for n in terms of i.

$$\sum_{n=2}^{\infty} \frac{n+5}{2^n} = \sum_{i=0}^{\infty} \frac{(i+2)+5}{2^{i+2}} = \sum_{i=0}^{\infty} \frac{i+7}{2^{i+2}}$$

To finish the problem out we'll recall that the letter we used for the index doesn't matter and so we'll change the final i back into an n to get,

$$\sum_{n=2}^{\infty} \frac{n+5}{2^n} = \sum_{n=0}^{\infty} \frac{n+7}{2^{n+2}}$$

To convince yourselves that these really are the same summation let's write out the first couple of terms for each of them,

$$\sum_{n=2}^{\infty} \frac{n+5}{2^n} = \frac{7}{2^2} + \frac{8}{2^3} + \frac{9}{2^4} + \frac{10}{2^5} + \cdots$$
$$\sum_{n=0}^{\infty} \frac{n+7}{2^{n+2}} = \frac{7}{2^2} + \frac{8}{2^3} + \frac{9}{2^4} + \frac{10}{2^5} + \cdots$$

So, sure enough the two series do have exactly the same terms.

There is actually an easier way to do an index shift. The method given above is the technically correct way of doing an index shift. However, notice in the above example we decreased the initial value of the index by 2 and all the n's in the series terms increased by 2 as well.

This will always work in this manner. If we decrease the initial value of the index by a set amount then all the other n's in the series term will increase by the same amount. Likewise, if we increase the initial value of the index by a set amount, then all the n's in the series term will decrease by the same amount.

Let's do a couple of examples using this shorthand method for doing index shifts.

Example 1

Perform the following index shifts.

(a) Write $\displaystyle\sum_{n=1}^{\infty} ar^{n-1}$ as a series that starts at $n = 0$.

(b) Write $\displaystyle\sum_{n=1}^{\infty} \frac{n^2}{1-3^{n+1}}$ as a series that starts at $n = 3$.

Solution

(a) Write $\displaystyle\sum_{n=1}^{\infty} ar^{n-1}$ as a series that starts at $n = 0$.

In this case we need to decrease the initial value by 1 and so the n's (okay the single n) in the term must increase by 1 as well.

$$\sum_{n=1}^{\infty} ar^{n-1} = \sum_{n=0}^{\infty} ar^{(n+1)-1} = \sum_{n=0}^{\infty} ar^{n}$$

(b) Write $\displaystyle\sum_{n=1}^{\infty} \frac{n^2}{1-3^{n+1}}$ as a series that starts at $n = 3$.

For this problem we want to increase the initial value by 2 and so all the n's in the series term must decrease by 2.

$$\sum_{n=1}^{\infty} \frac{n^2}{1-3^{n+1}} = \sum_{n=3}^{\infty} \frac{(n-2)^2}{1-3^{(n-2)+1}} = \sum_{n=3}^{\infty} \frac{(n-2)^2}{1-3^{n-1}}$$

The final topic in this section is again a topic that we'll not be seeing all that often in this class, although we will be seeing it more often than the index shifts. This final topic is really more about alternate ways to write series when the situation requires it.

Let's start with the following series and note that the $n = 1$ starting point is only for convenience since we need to start the series somewhere.

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + a_4 + a_5 + \cdots$$

Notice that if we ignore the first term the remaining terms will also be a series that will start at $n = 2$ instead of $n = 1$ So, we can rewrite the original series as follows,

$$\sum_{n=1}^{\infty} a_n = a_1 + \sum_{n=2}^{\infty} a_n$$

In this example we say that we've stripped out the first term.

We could have stripped out more terms if we wanted to. In the following series we've stripped out the first two terms and the first four terms respectively.

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + \sum_{n=3}^{\infty} a_n$$
$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + a_4 + \sum_{n=5}^{\infty} a_n$$

Being able to strip out terms will, on occasion, simplify our work or allow us to reuse a prior result so it's an important idea to remember.

Notice that in the second example above we could have also denoted the four terms that we stripped out as a finite series as follows,

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + a_4 + \sum_{n=5}^{\infty} a_n = \sum_{n=1}^{4} a_n + \sum_{n=5}^{\infty} a_n$$

This is a convenient notation when we are stripping out a large number of terms or if we need to strip out an undetermined number of terms. In general, we can write a series as follows,

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{N} a_n + \sum_{n=N+1}^{\infty} a_n$$

We'll leave this section with an important warning about terminology. Don't get sequences and series confused! A sequence is a list of numbers written in a specific order while an infinite series is a limit of a sequence of finite series and hence, if it exists will be a single value.

So, once again, a sequence is a list of numbers while a series is a single number, provided it makes sense to even compute the series. Students will often confuse the two and try to use facts pertaining to one on the other. However, since they are different beasts this just won't work. There will be problems where we are using both sequences and series so we'll always have to remember that they are different.

10.4 Convergence & Divergence of Series

In the previous section we spent some time getting familiar with series and we briefly defined convergence and divergence. Before worrying about convergence and divergence of a series we wanted to make sure that we've started to get comfortable with the notation involved in series and some of the various manipulations of series that we will, on occasion, need to be able to do.

As noted in the previous section most of what we were doing there won't be done much in this chapter. So, it is now time to start talking about the convergence and divergence of a series as this will be a topic that we'll be dealing with to one extent or another in almost all of the remaining sections of this chapter.

So, let's recap just what an infinite series is and what it means for a series to be convergent or divergent. We'll start with a sequence $\{a_n\}_{n=1}^{\infty}$ and again note that we're starting the sequence at $n = 1$ only for the sake of convenience and it can, in fact, be anything.

Next, we define the partial sums of the series as,

$$\begin{aligned} s_1 &= a_1 \\ s_2 &= a_1 + a_2 \\ s_3 &= a_1 + a_2 + a_3 \\ s_4 &= a_1 + a_2 + a_3 + a_4 \\ &\vdots \\ s_n &= a_1 + a_2 + a_3 + a_4 + \cdots + a_n = \sum_{i=1}^{n} a_i \end{aligned}$$

and these form a new sequence, $\{s_n\}_{n=1}^{\infty}$.

An infinite series, or just series here since almost every series that we'll be looking at will be an infinite series, is then the limit of the partial sums. Or,

$$\sum_{i=1}^{\infty} a_i = \lim_{n\to\infty} s_n$$

It is important to remember that $\sum\limits_{i=1}^{\infty} a_i$ is really nothing more than a convenient notation for $\lim\limits_{n\to\infty} \sum\limits_{i=1}^{n} a_i$ so we do not need to keep writing the limit down. We do, however, always need to remind ourselves that we really do have a limit there!

If the sequence of partial sums is a convergent sequence (*i.e.* its limit exists and is finite) then the series is also called **convergent** and in this case if $\lim\limits_{n\to\infty} s_n = s$ then, $\sum\limits_{i=1}^{\infty} a_i = s$. Likewise, if the sequence of partial sums is a divergent sequence (*i.e.* its limit doesn't exist or is plus or minus infinity) then the series is also called **divergent**.

Let's take a look at some series and see if we can determine if they are convergent or divergent and see if we can determine the value of any convergent series we find.

Example 1

Determine if the following series is convergent or divergent. If it converges determine the value of the series.

$$\sum_{n=1}^{\infty} n$$

Solution

To determine if the series is convergent we first need to get our hands on a formula for the general term in the sequence of partial sums.

$$s_n = \sum_{i=1}^{n} i$$

This is a known series and its value can be shown to be,

$$s_n = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

Don't worry if you didn't know this formula (we'd be surprised if anyone knew it...) as you won't be required to know it in my course.

So, to determine if the series is convergent we will first need to see if the sequence of partial sums,

$$\left\{\frac{n(n+1)}{2}\right\}_{n=1}^{\infty}$$

is convergent or divergent. That's not terribly difficult in this case. The limit of the sequence terms is,

$$\lim_{n\to\infty} \frac{n(n+1)}{2} = \infty$$

Therefore, the sequence of partial sums diverges to ∞ and so the series also diverges.

So, as we saw in this example we had to know a fairly obscure formula in order to determine the convergence of this series. In general finding a formula for the general term in the sequence of partial sums is a very difficult process. In fact after the next section we'll not be doing much with the partial sums of series due to the extreme difficulty faced in finding the general formula. This also means that we'll not be doing much work with the value of series since in order to get the value we'll also need to know the general formula for the partial sums.

We will continue with a few more examples however, since this is technically how we determine convergence and the value of a series. Also, the remaining examples we'll be looking at in this section will lead us to a very important fact about the convergence of series.

So, let's take a look at a couple more examples.

Example 2

Determine if the following series converges or diverges. If it converges determine the value of the series.

$$\sum_{n=2}^{\infty} \frac{1}{n^2-1}$$

Solution

This is actually one of the few series in which we are able to determine a formula for the general term in the sequence of partial fractions. However, in this section we are more interested in the general idea of convergence and divergence and so we'll put off discussing the process for finding the formula until the next section.

The general formula for the partial sums is,

$$s_n = \sum_{i=2}^{n} \frac{1}{i^2-1} = \frac{3}{4} - \frac{1}{2n} - \frac{1}{2(n+1)}$$

and in this case we have,

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} \left(\frac{3}{4} - \frac{1}{2n} - \frac{1}{2(n+1)} \right) = \frac{3}{4}$$

The sequence of partial sums converges and so the series converges also and its value is,

$$\sum_{n=2}^{\infty} \frac{1}{n^2-1} = \frac{3}{4}$$

Example 3

Determine if the following series converges or diverges. If it converges determine the value of the series.

$$\sum_{n=0}^{\infty} (-1)^n$$

Solution

In this case we really don't need a general formula for the partial sums to determine the convergence of this series. Let's just write down the first few partial sums.

$$\begin{aligned}
s_0 &= 1 \\
s_1 &= 1 - 1 = 0 \\
s_2 &= 1 - 1 + 1 = 1 \\
s_3 &= 1 - 1 + 1 - 1 = 0 \\
&etc.
\end{aligned}$$

So, it looks like the sequence of partial sums is,

$$\{s_n\}_{n=0}^{\infty} = \{1, 0, 1, 0, 1, 0, 1, 0, 1, \ldots\}$$

and this sequence diverges since $\lim\limits_{n\to\infty} s_n$ doesn't exist. Therefore, the series also diverges.

Example 4

Determine if the following series converges or diverges. If it converges determine the value of the series.

$$\sum_{n=1}^{\infty} \frac{1}{3^{n-1}}$$

Solution

Here is the general formula for the partial sums for this series.

$$s_n = \sum_{i=1}^{n} \frac{1}{3^{i-1}} = \frac{3}{2}\left(1 - \frac{1}{3^n}\right)$$

Again, do not worry about knowing this formula. This is not something that you'll ever be asked to know in my class.

In this case the limit of the sequence of partial sums is,

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} \frac{3}{2}\left(1 - \frac{1}{3^n}\right) = \frac{3}{2}$$

The sequence of partial sums is convergent and so the series will also be convergent. The value of the series is,

$$\sum_{n=1}^{\infty} \frac{1}{3^{n-1}} = \frac{3}{2}$$

As we already noted, do not get excited about determining the general formula for the sequence of partial sums. There is only going to be one type of series where you will need to determine this formula and the process in that case isn't too bad. In fact, you already know how to do most of the work in the process as you'll see in the next section.

So, we've determined the convergence of four series now. Two of the series converged and two diverged. Let's go back and examine the series terms for each of these. For each of the series let's take the limit as n goes to infinity of the series terms (not the partial sums!!).

$\lim_{n\to\infty} n = \infty$	this series diverged
$\lim_{n\to\infty} \frac{1}{n^2-1} = 0$	this series converged
$\lim_{n\to\infty} (-1)^n$ doesn't exist	this series diverged
$\lim_{n\to\infty} \frac{1}{3^{n-1}} = 0$	this series converged

Notice that for the two series that converged the series term itself was zero in the limit. This will always be true for convergent series and leads to the following theorem.

Theorem

If $\sum a_n$ converges then $\lim_{n\to\infty} a_n = 0$.

Proof

First let's suppose that the series starts at $n = 1$. If it doesn't then we can modify things as appropriate below. Then the partial sums are,

$$s_{n-1} = \sum_{i=1}^{n-1} a_i = a_1 + a_2 + a_3 + a_4 + \cdots + a_{n-1} \qquad s_n = \sum_{i=1}^{n} a_i = a_1 + a_2 + a_3 + a_4 + \cdots + a_{n-1} + a_n$$

Next, we can use these two partial sums to write,

$$a_n = s_n - s_{n-1}$$

Now because we know that $\sum a_n$ is convergent we also know that the sequence $\{s_n\}_{n=1}^{\infty}$ is also convergent and that $\lim\limits_{n\to\infty} s_n = s$ for some finite value s. However, since $n-1\to\infty$ as $n\to\infty$ we also have $\lim\limits_{n\to\infty} s_{n-1} = s$.

We now have,

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} (s_n - s_{n-1}) = \lim_{n\to\infty} s_n - \lim_{n\to\infty} s_{n-1} = s - s = 0$$

Be careful to not misuse this theorem! This theorem gives us a requirement for convergence but not a guarantee of convergence. In other words, the converse is NOT true. If $\lim\limits_{n\to\infty} a_n = 0$ the series may actually diverge! Consider the following two series.

$$\sum_{n=1}^{\infty} \frac{1}{n} \qquad\qquad \sum_{n=1}^{\infty} \frac{1}{n^2}$$

In both cases the series terms are zero in the limit as n goes to infinity, yet only the second series converges. The first series diverges. It will be a couple of sections before we can prove this, so at this point please believe this and know that you'll be able to prove the convergence of these two series in a couple of sections.

Again, as noted above, all this theorem does is give us a requirement for a series to converge. In order for a series to converge the series terms must go to zero in the limit. If the series terms do not go to zero in the limit then there is no way the series can converge since this would violate the theorem.

This leads us to the first of many tests for the convergence/divergence of a series that we'll be seeing in this chapter.

Divergence Test

If $\lim\limits_{n\to\infty} a_n \neq 0$ then $\sum a_n$ will diverge.

Again, do NOT misuse this test. This test only says that a series is guaranteed to diverge if the series terms don't go to zero in the limit. If the series terms do happen to go to zero the series may or may not converge! Again, recall the following two series,

$$\sum_{n=1}^{\infty} \frac{1}{n} \qquad \text{diverges}$$

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \qquad \text{converges}$$

One of the more common mistakes that students make when they first get into series is to assume that if $\lim\limits_{n\to\infty} a_n = 0$ then $\sum a_n$ will converge. There is just no way to guarantee this so be careful!

Let's take a quick look at an example of how this test can be used.

Example 5

Determine if the following series is convergent or divergent.

$$\sum_{n=0}^{\infty} \frac{4n^2 - n^3}{10 + 2n^3}$$

Solution

With almost every series we'll be looking at in this chapter the first thing that we should do is take a look at the series terms and see if they go to zero or not. If it's clear that the terms don't go to zero use the Divergence Test and be done with the problem.

That's what we'll do here.

$$\lim_{n\to\infty} \frac{4n^2 - n^3}{10 + 2n^3} = -\frac{1}{2} \neq 0$$

The limit of the series terms isn't zero and so by the Divergence Test the series diverges.

The divergence test is the first test of many tests that we will be looking at over the course of the next several sections. You will need to keep track of all these tests, the conditions under which they can be used and their conclusions all in one place so you can quickly refer back to them as you need to.

Next we should briefly revisit arithmetic of series and convergence/divergence. As we saw in the previous section if $\sum a_n$ and $\sum b_n$ are both convergent series then so are $\sum c a_n$ and $\sum\limits_{n=k}^{\infty} (a_n \pm b_n)$. Furthermore, these series will have the following sums or values.

$$\sum c a_n = c \sum a_n \qquad\qquad \sum_{n=k}^{\infty} (a_n \pm b_n) = \sum_{n=k}^{\infty} a_n \pm \sum_{n=k}^{\infty} b_n$$

We'll see an example of this in the next section after we get a few more examples under our belt. At this point just remember that a sum of convergent series is convergent and multiplying a convergent series by a number will not change its convergence.

We need to be a little careful with these facts when it comes to divergent series. In the first case if $\sum a_n$ is divergent then $\sum c a_n$ will also be divergent (provided c isn't zero of course) since multiplying a series that is infinite in value or doesn't have a value by a finite value (*i.e.* c) won't change the fact that the series has an infinite or no value. However, it is possible to have both $\sum a_n$ and $\sum b_n$ be divergent series and yet have $\sum\limits_{n=k}^{\infty} (a_n \pm b_n)$ be a convergent series.

Now, since the main topic of this section is the convergence of a series we should mention a stronger type of convergence. A series $\sum a_n$ is said to **converge absolutely** if $\sum |a_n|$ also converges. Absolute convergence is *stronger* than convergence in the sense that a series that is absolutely convergent will also be convergent, but a series that is convergent may or may not be absolutely convergent.

In fact if $\sum a_n$ converges and $\sum |a_n|$ diverges the series $\sum a_n$ is called **conditionally convergent**.

At this point we don't really have the tools at hand to properly investigate this topic in detail nor do we have the tools in hand to determine if a series is absolutely convergent or not. So we'll not say anything more about this subject for a while. When we finally have the tools in hand to discuss this topic in more detail we will revisit it. Until then don't worry about it. The idea is mentioned here only because we were already discussing convergence in this section and it ties into the last topic that we want to discuss in this section.

In the previous section after we'd introduced the idea of an infinite series we commented on the fact that we shouldn't think of an infinite series as an infinite sum despite the fact that the notation we use for infinite series seems to imply that it is an infinite sum. It's now time to briefly discuss this.

First, we need to introduce the idea of a **rearrangement**. A rearrangement of a series is exactly what it might sound like, it is the same series with the terms rearranged into a different order.

For example, consider the following infinite series.

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 + \cdots$$

A rearrangement of this series is,

$$\sum_{n=1}^{\infty} a_n = a_2 + a_1 + a_3 + a_{14} + a_5 + a_9 + a_4 + \cdots$$

The issue we need to discuss here is that for some series each of these arrangements of terms can have different values despite the fact that they are using exactly the same terms.

Here is an example of this. It can be shown that,

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8} + \cdots = \ln 2 \tag{10.1}$$

Since this series converges we know that if we multiply it by a constant c its value will also be multiplied by c. So, let's multiply this by $\frac{1}{2}$ to get,

$$\frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \frac{1}{10} - \frac{1}{12} + \frac{1}{14} - \frac{1}{16} + \cdots = \frac{1}{2}\ln 2 \tag{10.2}$$

Now, let's add in a zero between each term as follows.

$$0 + \frac{1}{2} + 0 - \frac{1}{4} + 0 + \frac{1}{6} + 0 - \frac{1}{8} + 0 + \frac{1}{10} + 0 - \frac{1}{12} + 0 + \cdots = \frac{1}{2}\ln 2 \tag{10.3}$$

Note that this won't change the value of the series because the partial sums for this series will be the partial sums for the Equation 10.2 except that each term will be repeated. Repeating terms in a series will not affect its limit however and so both Equation 10.2 and Equation 10.3 will be the same.

We know that if two series converge we can add them by adding term by term and so add Equation 10.1 and Equation 10.3 to get,

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \cdots = \frac{3}{2}\ln 2 \tag{10.4}$$

Now, notice that the terms of Equation 10.4 are simply the terms of Equation 10.1 rearranged so that each negative term comes after two positive terms. The values however are definitely different despite the fact that the terms are the same.

Note as well that this is not one of those "tricks" that you see occasionally where you get a contradictory result because of a hard to spot math/logic error. This is a very real result and we've not made any logic mistakes/errors.

Here is a nice set of facts that govern this idea of when a rearrangement will lead to a different value of a series.

Facts

Given the series $\sum a_n$,

1. If $\sum a_n$ is absolutely convergent and its value is s then any rearrangement of $\sum a_n$ will also have a value of s.
2. If $\sum a_n$ is conditionally convergent and r is any real number then there is a rearrangement of $\sum a_n$ whose value will be r.

Again, we do not have the tools in hand yet to determine if a series is absolutely convergent and so don't worry about this at this point. This is here just to make sure that you understand that we have to be very careful in thinking of an infinite series as an infinite sum. There are times when we can (*i.e.* the series is absolutely convergent) and there are times when we can't (*i.e.* the series is conditionally convergent).

As a final note, the fact above tells us that the series,

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$$

must be conditionally convergent since two rearrangements gave two separate values of this series. Eventually it will be very simple to show that this series is conditionally convergent.

10.5 Special Series

In this section we are going to take a brief look at three special series. Actually, special may not be the correct term. All three have been named which makes them special in some way, however the main reason that we're going to look at two of them in this section is that they are the only types of series that we'll be looking at for which we will be able to get actual values for the series. The third type is divergent and so won't have a value to worry about.

In general, determining the value of a series is very difficult and outside of these two kinds of series that we'll look at in this section we will not be determining the value of series in this chapter.

So, let's get started.

Geometric Series

A geometric series is any series that can be written in the form,

$$\sum_{n=1}^{\infty} ar^{n-1}$$

or, with an index shift the geometric series will often be written as,

$$\sum_{n=0}^{\infty} ar^{n}$$

These are identical series and will have identical values, provided they converge of course.

If we start with the first form it can be shown that the partial sums are,

$$s_n = \frac{a\left(1-r^n\right)}{1-r} = \frac{a}{1-r} - \frac{ar^n}{1-r}$$

The series will converge provided the partial sums form a convergent sequence, so let's take the limit of the partial sums.

$$\begin{aligned}\lim_{n\to\infty} s_n &= \lim_{n\to\infty}\left(\frac{a}{1-r} - \frac{ar^n}{1-r}\right)\\ &= \lim_{n\to\infty}\frac{a}{1-r} - \lim_{n\to\infty}\frac{ar^n}{1-r}\\ &= \frac{a}{1-r} - \frac{a}{1-r}\lim_{n\to\infty} r^n\end{aligned}$$

Now, from Theorem 3 from the Sequences section we know that the limit above will exist and be finite provided $-1 < r \le 1$. However, note that we can't let $r = 1$ since this will give division by zero. Therefore, this will exist and be finite provided $-1 < r < 1$ and in this case the limit is zero and so we get,

$$\lim_{n\to\infty} s_n = \frac{a}{1-r}$$

Therefore, a geometric series will converge if $-1 < r < 1$, which is usually written $|r| < 1$, its value is,

$$\sum_{n=1}^{\infty} ar^{n-1} = \sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$$

Note that in using this formula we'll need to make sure that we are in the correct form. In other words, if the series starts at $n = 0$ then the exponent on the r must be n. Likewise, if the series starts at $n = 1$ then the exponent on the r must be $n - 1$.

Example 1

Determine if the following series converge or diverge. If they converge give the value of the series.

(a) $\displaystyle\sum_{n=1}^{\infty} 9^{-n+2}4^{n+1}$

(b) $\displaystyle\sum_{n=0}^{\infty} \frac{(-4)^{3n}}{5^{n-1}}$

Solution

(a) $\displaystyle\sum_{n=1}^{\infty} 9^{-n+2}4^{n+1}$

This series doesn't really look like a geometric series. However, notice that both parts of the series term are numbers raised to a power. This means that it can be put into the form of a geometric series. We will just need to decide which form is the correct form. Since the series starts at $n = 1$ we will want the exponents on the numbers to be $n - 1$.

It will be fairly easy to get this into the correct form. Let's first rewrite things slightly. One of the n's in the exponent has a negative in front of it and that can't be there in the geometric form. So, let's first get rid of that.

$$\sum_{n=1}^{\infty} 9^{-n+2}4^{n+1} = \sum_{n=1}^{\infty} 9^{-(n-2)}4^{n+1} = \sum_{n=1}^{\infty} \frac{4^{n+1}}{9^{n-2}}$$

Now let's get the correct exponent on each of the numbers. This can be done using simple exponent properties.

$$\sum_{n=1}^{\infty} 9^{-n+2}4^{n+1} = \sum_{n=1}^{\infty} \frac{4^{n+1}}{9^{n-2}} = \sum_{n=1}^{\infty} \frac{4^{n-1}4^2}{9^{n-1}9^{-1}}$$

Now, rewrite the term a little.

$$\sum_{n=1}^{\infty} 9^{-n+2}4^{n+1} = \sum_{n=1}^{\infty} 16\,(9)\,\frac{4^{n-1}}{9^{n-1}} = \sum_{n=1}^{\infty} 144\left(\frac{4}{9}\right)^{n-1}$$

So, this is a geometric series with $a = 144$ and $r = \frac{4}{9} < 1$. Therefore, since $|r| < 1$ we know the series will converge and its value will be,

$$\sum_{n=1}^{\infty} 9^{-n+2}4^{n+1} = \frac{144}{1-\frac{4}{9}} = \frac{9}{5}\,(144) = \frac{1296}{5}$$

(b) $\displaystyle\sum_{n=0}^{\infty} \frac{(-4)^{3n}}{5^{n-1}}$

Again, this doesn't look like a geometric series, but it can be put into the correct form. In this case the series starts at $n = 0$ so we'll need the exponents to be n on the terms. Note that this means we're going to need to rewrite the exponent on the numerator a little

$$\sum_{n=0}^{\infty} \frac{(-4)^{3n}}{5^{n-1}} = \sum_{n=0}^{\infty} \frac{\left((-4)^3\right)^n}{5^n 5^{-1}} = \sum_{n=0}^{\infty} 5\frac{(-64)^n}{5^n} = \sum_{n=0}^{\infty} 5\left(\frac{-64}{5}\right)^n$$

So, we've got it into the correct form and we can see that $a = 5$ and $r = -\frac{64}{5}$. Also note that $|r| \geq 1$ and so this series diverges.

Back in the Series - The Basics section we talked about stripping out terms from a series, but didn't really provide any examples of how this idea could be used in practice. We can now do some examples.

Example 2

Use the results from the previous example to determine the value of the following series.

(a) $\displaystyle\sum_{n=0}^{\infty} 9^{-n+2}4^{n+1}$

(b) $\displaystyle\sum_{n=3}^{\infty} 9^{-n+2}4^{n+1}$

Solution

(a) $\displaystyle\sum_{n=0}^{\infty} 9^{-n+2} 4^{n+1}$

In this case we could just acknowledge that this is a geometric series that starts at $n = 0$ and so we could put it into the correct form and be done with it. However, this does provide us with a nice example of how to use the idea of stripping out terms to our advantage.

Let's notice that if we strip out the first term from this series we arrive at,

$$\sum_{n=0}^{\infty} 9^{-n+2} 4^{n+1} = 9^2 4^1 + \sum_{n=1}^{\infty} 9^{-n+2} 4^{n+1} = 324 + \sum_{n=1}^{\infty} 9^{-n+2} 4^{n+1}$$

From the previous example we know the value of the new series that arises here and so the value of the series in this example is,

$$\sum_{n=0}^{\infty} 9^{-n+2} 4^{n+1} = 324 + \frac{1296}{5} = \frac{2916}{5}$$

(b) $\displaystyle\sum_{n=3}^{\infty} 9^{-n+2} 4^{n+1}$

In this case we can't strip out terms from the given series to arrive at the series used in the previous example. However, we can start with the series used in the previous example and strip terms out of it to get the series in this example. So, let's do that. We will strip out the first two terms from the series we looked at in the previous example.

$$\sum_{n=1}^{\infty} 9^{-n+2} 4^{n+1} = 9^1 4^2 + 9^0 4^3 + \sum_{n=3}^{\infty} 9^{-n+2} 4^{n+1} = 208 + \sum_{n=3}^{\infty} 9^{-n+2} 4^{n+1}$$

We can now use the value of the series from the previous example to get the value of this series.

$$\sum_{n=3}^{\infty} 9^{-n+2} 4^{n+1} = \sum_{n=1}^{\infty} 9^{-n+2} 4^{n+1} - 208 = \frac{1296}{5} - 208 = \frac{256}{5}$$

Notice that we didn't discuss the convergence of either of the series in the above example. Here's why. Consider the following series written in two separate ways (*i.e.* we stripped out a couple of terms from it).

$$\sum_{n=0}^{\infty} a_n = a_0 + a_1 + a_2 + \sum_{n=3}^{\infty} a_n$$

Let's suppose that we know $\sum_{n=3}^{\infty} a_n$ is a convergent series. This means that it's got a finite value and adding three finite terms onto this will not change that fact. So the value of $\sum_{n=0}^{\infty} a_n$ is also finite and so is convergent.

Likewise, suppose that $\sum_{n=0}^{\infty} a_n$ is convergent. In this case if we subtract three finite values from this value we will remain finite and arrive at the value of $\sum_{n=3}^{\infty} a_n$. This is now a finite value and so this series will also be convergent.

In other words, if we have two series and they differ only by the presence, or absence, of a finite number of finite terms they will either both be convergent or they will both be divergent. The difference of a few terms one way or the other will not change the convergence of a series. This is an important idea and we will use it several times in the following sections to simplify some of the tests that we'll be looking at.

Telescoping Series

It's now time to look at the second of the three series in this section. In this portion we are going to look at a series that is called a telescoping series. The name in this case comes from what happens with the partial sums and is best shown in an example.

Example 3

Determine if the following series converges or diverges. If it converges find its value.

$$\sum_{n=0}^{\infty} \frac{1}{n^2 + 3n + 2}$$

Solution

We first need the partial sums for this series.

$$s_n = \sum_{i=0}^{n} \frac{1}{i^2 + 3i + 2}$$

Now, let's notice that we can use partial fractions on the series term to get,

$$\frac{1}{i^2 + 3i + 2} = \frac{1}{(i+2)(i+1)} = \frac{1}{i+1} - \frac{1}{i+2}$$

We'll leave the details of the partial fractions to you. By now you should be fairly adept at this since we spent a fair amount of time doing partial fractions back in the Integration Techniques chapter. If you need a refresher you should go back and review that section.

So, what does this do for us? Well, let's start writing out the terms of the general partial sum for this series using the partial fraction form.

$$\begin{aligned} s_n &= \sum_{i=0}^{n} \left(\frac{1}{i+1} - \frac{1}{i+2} \right) \\ &= \left(\frac{1}{1} - \cancel{\frac{1}{2}} \right) + \left(\cancel{\frac{1}{2}} - \cancel{\frac{1}{3}} \right) + \left(\cancel{\frac{1}{3}} - \cancel{\frac{1}{4}} \right) + \cdots + \left(\cancel{\frac{1}{n}} - \cancel{\frac{1}{n+1}} \right) + \left(\cancel{\frac{1}{n+1}} - \frac{1}{n+2} \right) \\ &= 1 - \frac{1}{n+2} \end{aligned}$$

Notice that every term except the first and last term canceled out. This is the origin of the name **telescoping series**.

This also means that we can determine the convergence of this series by taking the limit of the partial sums.

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} \left(1 - \frac{1}{n+2} \right) = 1$$

The sequence of partial sums is convergent and so the series is convergent and has a value of

$$\sum_{n=0}^{\infty} \frac{1}{n^2+3n+2} = 1$$

In telescoping series be careful to not assume that successive terms will be the ones that cancel. Consider the following example.

Example 4

Determine if the following series converges or diverges. If it converges find its value.

$$\sum_{n=1}^{\infty} \frac{1}{n^2+4n+3}$$

Solution

As with the last example we'll leave the partial fractions details to you to verify. The partial sums are,

$$\begin{aligned}
s_n &= \sum_{i=1}^{n} \left(\frac{\frac{1}{2}}{i+1} - \frac{\frac{1}{2}}{i+3} \right) = \frac{1}{2} \sum_{i=1}^{n} \left(\frac{1}{i+1} - \frac{1}{i+3} \right) \\
&= \frac{1}{2} \left[\left(\frac{1}{2} - \cancel{\frac{1}{4}} \right) + \left(\frac{1}{3} - \cancel{\frac{1}{5}} \right) + \left(\cancel{\frac{1}{4}} - \cancel{\frac{1}{6}} \right) + \cdots + \left(\cancel{\frac{1}{n}} - \frac{1}{n+2} \right) + \left(\cancel{\frac{1}{n+1}} - \frac{1}{n+3} \right) \right] \\
&= \frac{1}{2} \left[\frac{1}{2} + \frac{1}{3} - \frac{1}{n+2} - \frac{1}{n+3} \right]
\end{aligned}$$

In this case instead of successive terms canceling a term will cancel with a term that is farther down the list. The end result this time is two initial and two final terms are left. Notice as well that in order to help with the work a little we factored the $\frac{1}{2}$ out of the series.

The limit of the partial sums is,

$$\lim_{n \to \infty} s_n = \lim_{n \to \infty} \frac{1}{2} \left(\frac{5}{6} - \frac{1}{n+2} - \frac{1}{n+3} \right) = \frac{5}{12}$$

So, this series is convergent (because the partial sums form a convergent sequence) and its value is,

$$\sum_{n=1}^{\infty} \frac{1}{n^2 + 4n + 3} = \frac{5}{12}$$

Note that it's not always obvious if a series is telescoping or not until you try to get the partial sums and then see if they are in fact telescoping. There is no test that will tell us that we've got a telescoping series right off the bat. Also note that just because you can do partial fractions on a series term does not mean that the series will be a telescoping series. The following series, for example, is not a telescoping series despite the fact that we can partial fraction the series terms.

$$\sum_{n=1}^{\infty} \frac{3 + 2n}{n^2 + 3n + 2} = \sum_{n=1}^{\infty} \left(\frac{1}{n+1} + \frac{1}{n+2} \right)$$

In order for a series to be a telescoping series we must get terms to cancel and all of these terms are positive and so none will cancel.

Next, we need to go back and address an issue that was first raised in the previous section. In that section we stated that the sum or difference of convergent series was also convergent and that the presence of a multiplicative constant would not affect the convergence of a series. Now that we have a few more series in hand let's work a quick example showing that.

Example 5

Determine the value of the following series.

$$\sum_{n=1}^{\infty}\left(\frac{4}{n^2+4n+3}-9^{-n+2}4^{n+1}\right)$$

Solution

To get the value of this series all we need to do is rewrite it and then use the previous results.

$$\begin{aligned}\sum_{n=1}^{\infty}\left(\frac{4}{n^2+4n+3}-9^{-n+2}4^{n+1}\right)&=\sum_{n=1}^{\infty}\frac{4}{n^2+4n+3}-\sum_{n=1}^{\infty}9^{-n+2}4^{n+1}\\&=4\sum_{n=1}^{\infty}\frac{1}{n^2+4n+3}-\sum_{n=1}^{\infty}9^{-n+2}4^{n+1}\\&=4\left(\frac{5}{12}\right)-\frac{1296}{5}\\&=-\frac{3863}{15}\end{aligned}$$

We didn't discuss the convergence of this series because it was the sum of two convergent series and that guaranteed that the original series would also be convergent.

Harmonic Series

This is the third and final series that we're going to look at in this section. Here is the harmonic series.

$$\sum_{n=1}^{\infty}\frac{1}{n}$$

You can read a little bit about why it is called a harmonic series (has to do with music) at the Wikipedia page for the harmonic series.

The harmonic series is divergent and we'll need to wait until the next section to show that. This series is here because it's got a name and so we wanted to put it here with the other two named series that we looked at in this section. We're also going to use the harmonic series to illustrate a couple of ideas about divergent series that we've already discussed for convergent series. We'll do that with the following example.

Example 6

Show that each of the following series are divergent.

(a) $\displaystyle\sum_{n=1}^{\infty} \frac{5}{n}$

(b) $\displaystyle\sum_{n=4}^{\infty} \frac{1}{n}$

Solution

(a) $\displaystyle\sum_{n=1}^{\infty} \frac{5}{n}$

To see that this series is divergent all we need to do is use the fact that we can factor a constant out of a series as follows,

$$\sum_{n=1}^{\infty} \frac{5}{n} = 5\sum_{n=1}^{\infty} \frac{1}{n}$$

Now, $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent and so five times this will still not be a finite number and so the series has to be divergent. In other words, if we multiply a divergent series by a constant it will still be divergent.

(b) $\displaystyle\sum_{n=4}^{\infty} \frac{1}{n}$

In this case we'll start with the harmonic series and strip out the first three terms.

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \sum_{n=4}^{\infty} \frac{1}{n} \qquad \Rightarrow \qquad \sum_{n=4}^{\infty} \frac{1}{n} = \left(\sum_{n=1}^{\infty} \frac{1}{n}\right) - \frac{11}{6}$$

In this case we are subtracting a finite number from a divergent series. This subtraction will not change the divergence of the series. We will either have infinity minus a finite number, which is still infinity, or a series with no value minus a finite number, which will still have no value.

Therefore, this series is divergent.

Just like with convergent series, adding/subtracting a finite number from a divergent series is not going to change the divergence of the series.

So, some general rules about the convergence/divergence of a series are now in order. Multiplying a series by a constant will not change the convergence/divergence of the series and adding or subtracting a constant from a series will not change the convergence/divergence of the series. These are nice ideas to keep in mind.

10.6 Integral Test

The last topic that we discussed in the previous section was the harmonic series. In that discussion we stated that the harmonic series was a divergent series. It is now time to prove that statement. This proof will also get us started on the way to our next test for convergence that we'll be looking at.

So, we will be trying to prove that the harmonic series,

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverges.

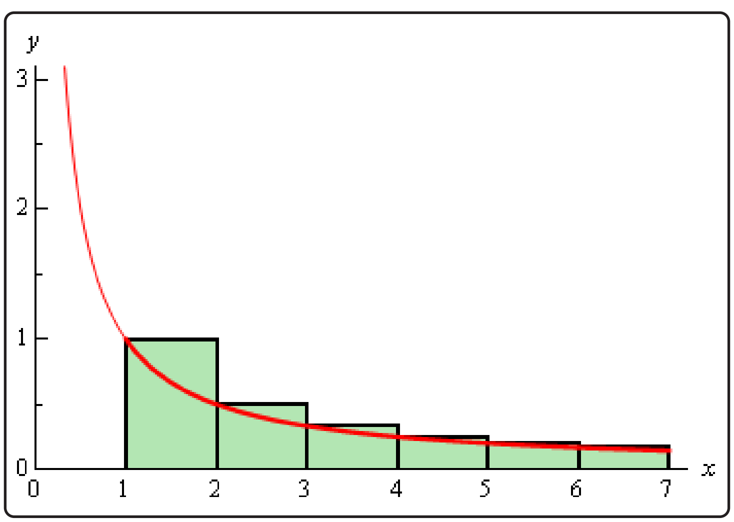
We'll start this off by looking at an apparently unrelated problem. Let's start off by asking what the area under $f(x) = \frac{1}{x}$ on the interval $[1, \infty)$. From the section on Improper Integrals we know that this is,

$$\int_{1}^{\infty} \frac{1}{x}\,dx = \infty$$

and so we called this integral divergent (yes, that's the same term we're using here with series....).

So, just how does that help us to prove that the harmonic series diverges? Well, recall that we can always estimate the area by breaking up the interval into segments and then sketching in rectangles and using the sum of the area all of the rectangles as an estimate of the actual area. Let's do that for this problem as well and see what we get.

We will break up the interval into subintervals of width 1 and we'll take the function value at the left endpoint as the height of the rectangle. The image below shows the first few rectangles for this area.



So, the area under the curve is approximately,

$$\begin{aligned} A &\approx \left(\frac{1}{1}\right)(1) + \left(\frac{1}{2}\right)(1) + \left(\frac{1}{3}\right)(1) + \left(\frac{1}{4}\right)(1) + \left(\frac{1}{5}\right)(1) + \cdots \\ &= \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \cdots \\ &= \sum_{n=1}^{\infty} \frac{1}{n} \end{aligned}$$

Now note a couple of things about this approximation. First, each of the rectangles overestimates the actual area and secondly the formula for the area is exactly the harmonic series!

Putting these two facts together gives the following,

$$A \approx \sum_{n=1}^{\infty} \frac{1}{n} > \int_1^{\infty} \frac{1}{x}\,dx = \infty$$

Notice that this tells us that we must have,

$$\sum_{n=1}^{\infty} \frac{1}{n} > \infty \qquad \Rightarrow \qquad \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

Since we can't really be larger than infinity the harmonic series must also be infinite in value. In other words, the harmonic series is in fact divergent.

So, we've managed to relate a series to an improper integral that we could compute and it turns out that the improper integral and the series have exactly the same convergence.

Let's see if this will also be true for a series that converges. When discussing the Divergence Test we made the claim that

$$\sum_{n=1}^{\infty} \frac{1}{n^2}$$

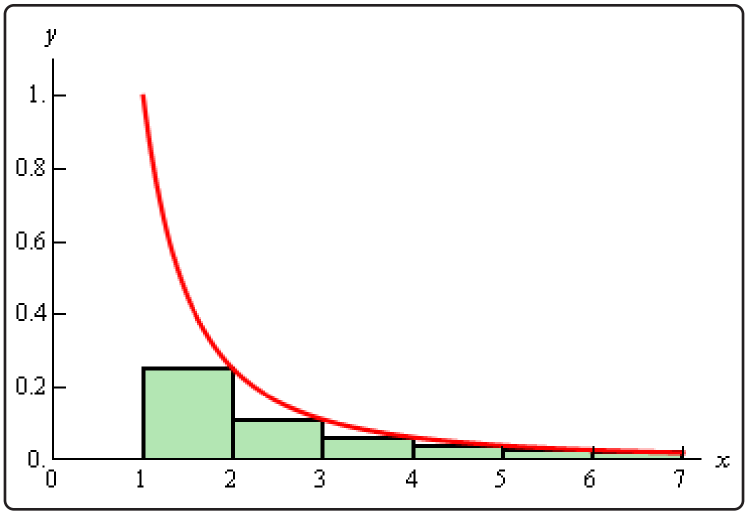
converges. Let's see if we can do something similar to the above process to prove this.

We will try to relate this to the area under $f(x) = \frac{1}{x^2}$ is on the interval $[1, \infty)$. Again, from the Improper Integral section we know that,

$$\int_1^{\infty} \frac{1}{x^2}\,dx = 1$$

and so this integral converges.

We will once again try to estimate the area under this curve. We will do this in an almost identical manner as the previous part with the exception that instead of using the left end points for the height of our rectangles we will use the right end points. Here is a sketch of this case,



In this case the area estimation is,

$$\begin{aligned} A &\approx \left(\frac{1}{2^2}\right)(1) + \left(\frac{1}{3^2}\right)(1) + \left(\frac{1}{4^2}\right)(1) + \left(\frac{1}{5^2}\right)(1) + \cdots \\ &= \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \cdots \end{aligned}$$

This time, unlike the first case, the area will be an underestimation of the actual area and the estimation is not quite the series that we are working with. Notice however that the only difference is that we're missing the first term. This means we can do the following,

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{1}{1^2} + \underbrace{\frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \cdots}_{\text{Area Estimation}} < 1 + \int_1^{\infty} \frac{1}{x^2}\,dx = 1 + 1 = 2$$

Or, putting all this together we see that,

$$\sum_{n=1}^{\infty} \frac{1}{n^2} < 2$$

With the harmonic series this was all that we needed to say that the series was divergent. With this series however, this isn't quite enough. For instance, $-\infty < 2$, and if the series did have a value of $-\infty$ then it would be divergent (when we want convergent). So, let's do a little more work.

First, let's notice that all the series terms are positive (that's important) and that the partial sums are,

$$s_n = \sum_{i=1}^{n} \frac{1}{i^2}$$

Because the terms are all positive we know that the partial sums must be an increasing sequence. In other words,

$$s_n = \sum_{i=1}^{n} \frac{1}{i^2} < \sum_{i=1}^{n+1} \frac{1}{i^2} = s_{n+1}$$

In s_{n+1} we are adding a single positive term onto s_n and so must get larger. Therefore, the partial sums form an increasing (and hence monotonic) sequence.

Also note that, since the terms are all positive, we can say,

$$s_n = \sum_{i=1}^{n} \frac{1}{i^2} < \sum_{n=1}^{\infty} \frac{1}{n^2} < 2 \qquad \Rightarrow \qquad s_n < 2$$

and so the sequence of partial sums is a bounded sequence.

In the second section on Sequences we gave a theorem that stated that a bounded and monotonic sequence was guaranteed to be convergent. This means that the sequence of partial sums is a convergent sequence. So, who cares right? Well recall that this means that the series must then also be convergent!

So, once again we were able to relate a series to an improper integral (that we could compute) and the series and the integral had the same convergence.

We went through a fair amount of work in both of these examples to determine the convergence of the two series. Luckily for us we don't need to do all this work every time. The ideas in these two examples can be summarized in the following test.

Integral Test

Suppose that $f(x)$ is a continuous, positive and decreasing function on the interval $[k, \infty)$ and that $f(n) = a_n$ then,

1. If $\displaystyle\int_k^{\infty} f(x)\, dx$ is convergent so is $\displaystyle\sum_{n=k}^{\infty} a_n$.
2. If $\displaystyle\int_k^{\infty} f(x)\, dx$ is divergent so is $\displaystyle\sum_{n=k}^{\infty} a_n$.

A formal proof of this test can be found at the end of this section.

There are a couple of things to note about the integral test. First, the lower limit on the improper integral must be the same value that starts the series.

Second, the function does not actually need to be decreasing and positive everywhere in the interval. All that's really required is that eventually the function is decreasing and positive. In other words, it is okay if the function (and hence series terms) increases or is negative for a while, but eventually the function (series terms) must decrease and be positive for all terms. To see why this is true let's suppose that the series terms increase and or are negative in the range $k \le n \le N$ and then decrease and are positive for $n \ge N+1$. In this case the series can be written as,

$$\sum_{n=k}^{\infty} a_n = \sum_{n=k}^{N} a_n + \sum_{n=N+1}^{\infty} a_n$$

Now, the first series is nothing more than a finite sum (no matter how large N is) of finite terms and so will be finite. So, the original series will be convergent/divergent only if the second infinite series on the right is convergent/divergent and the test can be done on the second series as it satisfies the conditions of the test.

A similar argument can be made using the improper integral as well.

The requirement in the test that the function/series be decreasing and positive everywhere in the range is required for the proof. In practice however, we only need to make sure that the function/series is eventually a decreasing and positive function/series. Also note that when computing the integral in the test we don't actually need to strip out the increasing/negative portion since the presence of a small range on which the function is increasing/negative will not change the integral from convergent to divergent or from divergent to convergent.

There is one more very important point that must be made about this test. This test does NOT give the value of a series. It will only give the convergence/divergence of the series. That's it. No value. We can use the above series as a perfect example of this. All that the test gave us was that,

$$\sum_{n=1}^{\infty} \frac{1}{n^2} < 2$$

So, we got an upper bound on the value of the series, but not an actual value for the series. In fact, from this point on we will not be asking for the value of a series we will only be asking whether a series converges or diverges. In a later section we look at estimating values of series, but even in that section still won't actually be getting values of series.

Just for the sake of completeness the value of this series is known.

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} = 1.644934... < 2$$

Let's work a couple of examples.

Example 1

Determine if the following series is convergent or divergent.

$$\sum_{n=2}^{\infty} \frac{1}{n\ln(n)}$$

Solution

In this case the function we'll use is,

$$f(x) = \frac{1}{x\ln(x)}$$

This function is clearly positive and if we make x larger the denominator will get larger and so the function is also decreasing. Therefore, all we need to do is determine the convergence of the following integral.

$$\begin{aligned}\int_2^{\infty} \frac{1}{x\ln(x)}\,dx &= \lim_{t\to\infty}\int_2^t \frac{1}{x\ln(x)}\,dx \qquad\qquad u = \ln(x)\\ &= \lim_{t\to\infty}\left(\ln(\ln(x))\right)\Big|_2^t\\ &= \lim_{t\to\infty}\left(\ln(\ln(t)) - \ln(\ln 2)\right)\\ &= \infty\end{aligned}$$

The integral is divergent and so the series is also divergent by the Integral Test.

Example 2

Determine if the following series is convergent or divergent.

$$\sum_{n=0}^{\infty} n\mathbf{e}^{-n^2}$$

Solution

The function that we'll use in this example is,

$$f(x) = x\mathbf{e}^{-x^2}$$

This function is always positive on the interval that we're looking at. Now we need to check that the function is decreasing. It is not clear that this function will always be decreasing on the interval given. We can use our Calculus I knowledge to help us however. The derivative of this function is,

$$f'(x) = \mathbf{e}^{-x^2}\left(1 - 2x^2\right)$$

This function has two critical points (which will tell us where the derivative changes sign) at $x = \pm\frac{1}{\sqrt{2}}$. Since we are starting at $n = 0$ we can ignore the negative critical point. Picking a couple of test points we can see that the function is increasing on the interval $\left[0, \frac{1}{\sqrt{2}}\right]$ and it is decreasing on $\left[\frac{1}{\sqrt{2}}, \infty\right)$. Therefore, eventually the function will be decreasing and that's all that's required for us to use the Integral Test.

$$\begin{aligned}\int_0^\infty x\mathbf{e}^{-x^2}\,dx &= \lim_{t\to\infty}\int_0^t x\mathbf{e}^{-x^2}\,dx \qquad\qquad u = -x^2\\ &= \lim_{t\to\infty}\left(-\frac{1}{2}\mathbf{e}^{-x^2}\right)\Bigg|_0^t\\ &= \lim_{t\to\infty}\left(\frac{1}{2} - \frac{1}{2}\mathbf{e}^{-t^2}\right) = \frac{1}{2}\end{aligned}$$

The integral is convergent and so the series must also be convergent by the Integral Test.

We can use the Integral Test to get the following fact/test for some series.

Fact (The p -series Test)

If $k > 0$ then $\displaystyle\sum_{n=k}^{\infty}\frac{1}{n^p}$ converges if $p > 1$ and diverges if $p \le 1$.

Sometimes the series in this fact are called p-series and so this fact is sometimes called the p-series test. This fact follows directly from the Integral Test and a similar fact we saw in the Improper Integral section. This fact says that the integral,

$$\int_k^\infty \frac{1}{x^p}\,dx$$

converges if $p > 1$ and diverges if $p \le 1$.

Using the p-series test makes it very easy to determine the convergence of some series.

Example 3

Determine if the following series are convergent or divergent.

(a) $\displaystyle\sum_{n=4}^{\infty} \frac{1}{n^7}$

(b) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$

Solution

(a) $\displaystyle\sum_{n=4}^{\infty} \frac{1}{n^7}$

In this case $p = 7 > 1$ and so by this fact the series is convergent.

(b) $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$

For this series $p = \frac{1}{2} \leq 1$ and so the series is divergent by the fact.

The last thing that we'll do in this section is give a quick proof of the Integral Test. We've essentially done the proof already at the beginning of the section when we were introducing the Integral Test, but let's go through it formally for a general function.

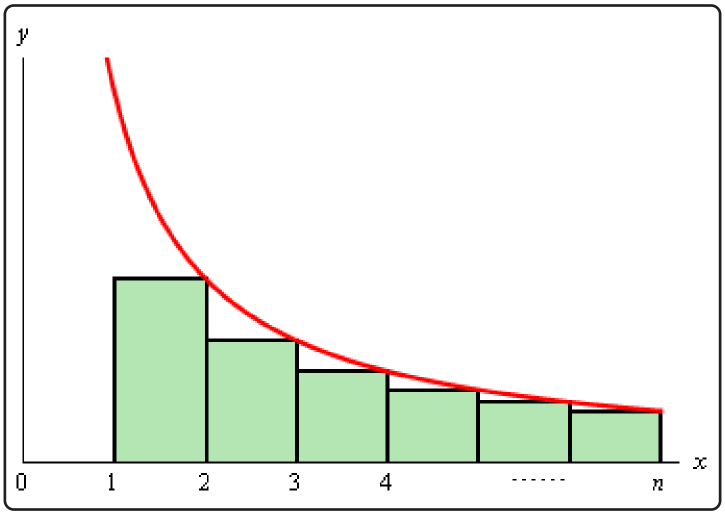
Proof of Integral Test

First, for the sake of the proof we'll be working with the series $\sum\limits_{n=1}^{\infty} a_n$. The original test statement was for a series that started at a general $n = k$ and while the proof can be done for that it will be easier if we assume that the series starts at $n = 1$.

Another way of dealing with the $n = k$ is we could do an index shift and start the series at $n = 1$ and then do the Integral Test. Either way proving the test for $n = 1$ will be sufficient.

Also note that while we allowed for the first few terms of the series to increase and/or be negative in working problems this proof does require that all the terms be decreasing and positive.

Let's start off and estimate the area under the curve on the interval $[1, n]$ and we'll underestimate the area by taking rectangles of width one and whose height is the right endpoint. This gives the following figure.



Now, note that,

$$f(2) = a_2 \qquad f(3) = a_3 \qquad \cdots \qquad f(n) = a_n$$

The approximate area is then,

$$A \approx (1)\,f(2) + (1)\,f(3) + \cdots + (1)\,f(n) = a_2 + a_3 + \cdots a_n$$

and we know that this underestimates the actual area so,

$$\sum_{i=2}^{n} a_i = a_2 + a_3 + \cdots a_n < \int_1^n f(x)\,dx$$

Now, let's suppose that $\int_1^\infty f(x)\,dx$ is convergent and so $\int_1^\infty f(x)\,dx$ must have a finite value. Also, because $f(x)$ is positive we know that,

$$\int_1^n f(x)\,dx < \int_1^\infty f(x)\,dx$$

This in turn means that,

$$\sum_{i=2}^{n} a_i < \int_1^n f(x)\,dx < \int_1^\infty f(x)\,dx$$

Our series starts at $n = 1$ so this isn't quite what we need. However, that's easy enough to deal with.

$$\sum_{i=1}^{n} a_i = a_1 + \sum_{i=2}^{n} a_i < a_1 + \int_1^\infty f(x)\,dx = M$$

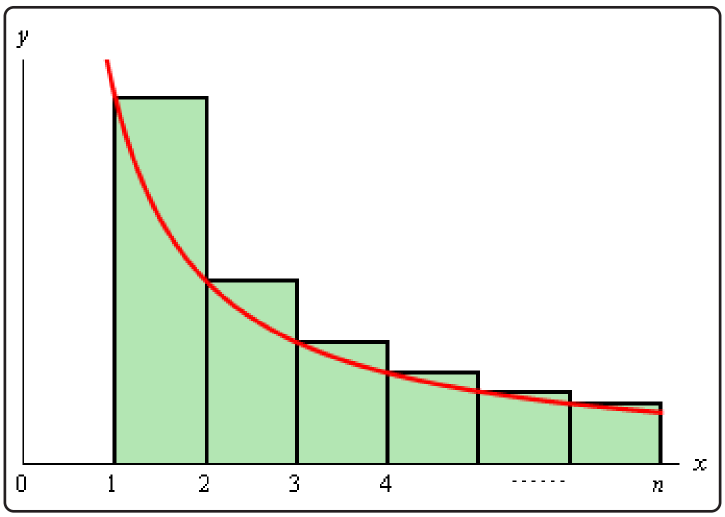
So, just what has this told us? Well we now know that the sequence of partial sums, $s_n = \sum_{i=1}^{n} a_i$ are bounded above by M.

Next, because the terms are positive we also know that,

$$s_n \le s_n + a_{n+1} = \sum_{i=1}^{n} a_i + a_{n+1} = \sum_{i=1}^{n+1} a_i = s_{n+1} \qquad \Rightarrow \qquad s_n \le s_{n+1}$$

and so the sequence $\{s_n\}_{n=1}^{\infty}$ is also an increasing sequence. So, we now know that the sequence of partial sums $\{s_n\}_{n=1}^{\infty}$ converges and hence our series $\sum_{n=1}^{\infty} a_n$ is convergent.

So, the first part of the test is proven. The second part is somewhat easier. This time let's overestimate the area under the curve by using the left endpoints of interval for the height of the rectangles as shown below.



In this case the area is approximately,

$$A \approx (1)\, f(1) + (1)\, f(2) + \cdots + (1)\, f(n-1) = a_1 + a_2 + \cdots a_{n-1}$$

Since we know this overestimates the area we also then know that,

$$s_{n-1} = \sum_{i=1}^{n-1} a_i = a_1 + a_2 + \cdots a_{n-1} > \int_1^{n-1} f(x)\, dx$$

Now, suppose that $\int_1^{\infty} f(x)\, dx$ is divergent. In this case this means that $\int_1^{n} f(x)\, dx \to \infty$ as $n \to \infty$ because $f(x) \ge 0$. However, because $n-1 \to \infty$ as $n \to \infty$ we also know that $\int_1^{n-1} f(x)\, dx \to \infty$.

Therefore, since $s_{n-1} > \int_1^{n-1} f(x)\,dx$ we know that as $n \to \infty$ we must have $s_{n-1} \to \infty$. This in turn tells us that $s_n \to \infty$ as $n \to \infty$.

So, we now know that the sequence of partial sums, $\{s_n\}_{n=1}^{\infty}$, is a divergent sequence and so $\sum\limits_{n=1}^{\infty} a_n$ is a divergent series.

It is important to note before leaving this section that in order to use the Integral Test the series terms MUST eventually be decreasing and positive. If they are not then the test doesn't work. Also remember that the test only determines the convergence of a series and does NOT give the value of the series.

10.7 Comparison Test/Limit Comparison Test

In the previous section we saw how to relate a series to an improper integral to determine the convergence of a series. While the integral test is a nice test, it does force us to do improper integrals which aren't always easy and, in some cases, may be impossible to determine the convergence of.

For instance, consider the following series.

$$\sum_{n=0}^{\infty} \frac{1}{3^n + n}$$

In order to use the Integral Test we would have to integrate

$$\int_0^{\infty} \frac{1}{3^x + x}\, dx$$

and we're not even sure if it's possible to do this integral. Nicely enough for us there is another test that we can use on this series that will be much easier to use.

First, let's note that the series terms are positive. As with the Integral Test that will be important in this section. Next let's note that we must have $x > 0$ since we are integrating on the interval $0 \le x < \infty$. Likewise, regardless of the value of x we will always have $3^x > 0$. So, if we drop the x from the denominator the denominator will get smaller and hence the whole fraction will get larger. So,

$$\frac{1}{3^n + n} < \frac{1}{3^n}$$

Now,

$$\sum_{n=0}^{\infty} \frac{1}{3^n}$$

is a geometric series and we know that since $|r| = \left|\frac{1}{3}\right| < 1$ the series will converge and its value will be,

$$\sum_{n=0}^{\infty} \frac{1}{3^n} = \frac{1}{1 - \frac{1}{3}} = \frac{3}{2}$$

Now, if we go back to our original series and write down the partial sums we get,

$$s_n = \sum_{i=0}^{n} \frac{1}{3^i + i}$$

Since all the terms are positive adding a new term will only make the number larger and so the sequence of partial sums must be an increasing sequence.

$$s_n = \sum_{i=0}^{n} \frac{1}{3^i + i} < \sum_{i=0}^{n+1} \frac{1}{3^i + i} = s_{n+1}$$

Then since,

$$\frac{1}{3^n + n} < \frac{1}{3^n}$$

and because the terms in these two sequences are positive we can also say that,

$$s_n = \sum_{i=0}^{n} \frac{1}{3^i + i} < \sum_{i=0}^{n} \frac{1}{3^i} < \sum_{i=0}^{\infty} \frac{1}{3^n} = \frac{3}{2} \qquad \Rightarrow \qquad s_n < \frac{3}{2}$$

Therefore, the sequence of partial sums is also a bounded sequence. Then from the second section on sequences we know that a monotonic and bounded sequence is also convergent.

So, the sequence of partial sums of our series is a convergent sequence. This means that the series itself,

$$\sum_{n=0}^{\infty} \frac{1}{3^n + n}$$

is also convergent.

So, what did we do here? We found a series whose terms were always larger than the original series terms and this new series was also convergent. Then since the original series terms were positive (very important) this meant that the original series was also convergent.

To show that a series (with only positive terms) was divergent we could go through a similar argument and find a new divergent series whose terms are always smaller than the original series. In this case the original series would have to take a value larger than the new series. However, since the new series is divergent its value will be infinite. This means that the original series must also be infinite and hence divergent.

We can summarize all this in the following test.

Comparison Test

Suppose that we have two series $\sum a_n$ and $\sum b_n$ with $a_n, b_n \geq 0$ for all n and $a_n \leq b_n$ for all n. Then,

1. If $\sum b_n$ is convergent then so is $\sum a_n$.
2. If $\sum a_n$ is divergent then so is $\sum b_n$.

In other words, we have two series of positive terms and the terms of one of the series is always larger than the terms of the other series. Then if the larger series is convergent the smaller series must also be convergent. Likewise, if the smaller series is divergent then the larger series must also be divergent. Note as well that in order to apply this test we need both series to start at the same place.

A formal proof of this test is at the end of this section.

Do not misuse this test. Just because the smaller of the two series converges does not say anything about the larger series. The larger series may still diverge. Likewise, just because we know that the larger of two series diverges we can't say that the smaller series will also diverge! Be very careful in using this test

Recall that we had a similar test for improper integrals back when we were looking at integration techniques. So, if you could use the comparison test for improper integrals you can use the comparison test for series as they are pretty much the same idea.

Note as well that the requirement that $a_n, b_n \geq 0$ and $a_n \leq b_n$ really only need to be true eventually. In other words, if a couple of the first terms are negative or $a_n \not\leq b_n$ for a couple of the first few terms we're okay. As long as we eventually reach a point where $a_n, b_n \geq 0$ and $a_n \leq b_n$ for all sufficiently large n the test will work.

To see why this is true let's suppose that the series start at $n = k$ and that the conditions of the test are only true for for $n \geq N + 1$ and for $k \leq n \leq N$ at least one of the conditions is not true. If we then look at $\sum a_n$ (the same thing could be done for $\sum b_n$) we get,

$$\sum_{n=k}^{\infty} a_n = \sum_{n=k}^{N} a_n + \sum_{n=N+1}^{\infty} a_n$$

The first series is nothing more than a finite sum (no matter how large N is) of finite terms and so will be finite. So, the original series will be convergent/divergent only if the second infinite series on the right is convergent/divergent and the test can be done on the second series as it satisfies the conditions of the test.

Let's take a look at some examples.

Example 1

Determine if the following series is convergent or divergent.

$$\sum_{n=1}^{\infty} \frac{n}{n^2 - \cos^2(n)}$$

Solution

Since the cosine term in the denominator doesn't get too large we can assume that the series terms will behave like,

$$\frac{n}{n^2} = \frac{1}{n}$$

which, as a series, will diverge. So, from this we can guess that the series will probably diverge and so we'll need to find a smaller series that will also diverge.

Recall that from the comparison test with improper integrals that we determined that we can make a fraction smaller by either making the numerator smaller or the denominator larger.

In this case the two terms in the denominator are both positive. So, if we drop the cosine term we will in fact be making the denominator larger since we will no longer be subtracting off a positive quantity. Therefore,

$$\frac{n}{n^2 - \cos^2(n)} > \frac{n}{n^2} = \frac{1}{n}$$

Then, since

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverges (it's harmonic or the p-series test) by the Comparison Test our original series must also diverge.

Example 2

Determine if the following series is convergent or divergent.

$$\sum_{n=1}^{\infty} \frac{\mathbf{e}^{-n}}{n + \cos^2(n)}$$

Solution

This example looks somewhat similar to the first one but we are going to have to be careful with it as there are some significant differences.

First, as with the first example the cosine term in the denominator will not get very large and so it won't affect the behavior of the terms in any meaningful way. Therefore, the temptation at this point is to focus in on the n in the denominator and think that because it is just an n the series will diverge.

That would be correct if we didn't have much going on in the numerator. In this example, however, we also have an exponential in the numerator that is going to zero very fast. In fact, it is going to zero so fast that it will, in all likelihood, force the series to converge.

So, let's guess that this series will converge and we'll need to find a larger series that will also converge.

First, because we are adding two positive numbers in the denominator we can drop the cosine term from the denominator. This will, in turn, make the denominator smaller and so the term will get larger or,

$$\frac{\mathbf{e}^{-n}}{n + \cos^2(n)} \le \frac{\mathbf{e}^{-n}}{n}$$

Next, we know that $n \geq 1$ and so if we replace the n in the denominator with its smallest possible value (i.e. 1) the term will again get larger. Doing this gives,

$$\frac{\mathbf{e}^{-n}}{n+\cos^2(n)} \leq \frac{\mathbf{e}^{-n}}{n} \leq \frac{\mathbf{e}^{-n}}{1} = \mathbf{e}^{-n}$$

We can't do much more, in a way that is useful anyway, to make this larger so let's see if we can determine if,

$$\sum_{n=1}^{\infty} \mathbf{e}^{-n}$$

converges or diverges.

We can notice that $f(x) = \mathbf{e}^{-x}$ is always positive and it is also decreasing (you can verify that correct?) and so we can use the Integral Test on this series. Doing this gives,

$$\int_1^{\infty} \mathbf{e}^{-x}\,dx = \lim_{t\to\infty}\int_1^t \mathbf{e}^{-x}\,dx = \lim_{t\to\infty}\left(-\mathbf{e}^{-x}\right)\Big|_1^t = \lim_{t\to\infty}\left(-\mathbf{e}^{-t}+\mathbf{e}^{-1}\right) = \mathbf{e}^{-1}$$

Okay, we now know that the integral is convergent and so the series $\sum\limits_{n=1}^{\infty} \mathbf{e}^{-n}$ must also be convergent.

Therefore, because $\sum\limits_{n=1}^{\infty} \mathbf{e}^{-n}$ is larger than the original series we know that the original series must also converge.

With each of the previous examples we saw that we can't always just focus in on the denominator when making a guess about the convergence of a series. Sometimes there is something going on in the numerator that will change the convergence of a series from what the denominator tells us should be happening.

We also saw in the previous example that, unlike most of the examples of the comparison test that we've done (or will do) both in this section and in the Comparison Test for Improper Integrals, that it won't always be the denominator that is driving the convergence or divergence. Sometimes it is the numerator that will determine if something will converge or diverge so do not get too locked into only looking at the denominator.

One of the more common mistakes is to just focus in on the denominator and make a guess based just on that. If we'd done that with both of the previous examples we would have guessed wrong so be careful.

Let's work another example of the comparison test before we move on to a different topic.

Example 3

Determine if the following series converges or diverges.

$$\sum_{n=1}^{\infty} \frac{n^2+2}{n^4+5}$$

Solution

In this case the "$+2$" and the "$+5$" don't really add anything to the series and so the series terms should behave pretty much like

$$\frac{n^2}{n^4} = \frac{1}{n^2}$$

which will converge as a series. Therefore, we can guess that the original series will converge and we will need to find a larger series which also converges.

This means that we'll either have to make the numerator larger or the denominator smaller. We can make the denominator smaller by dropping the "$+5$". Doing this gives,

$$\frac{n^2+2}{n^4+5} < \frac{n^2+2}{n^4}$$

At this point, notice that we can't drop the "$+2$" from the numerator since this would make the term smaller and that's not what we want. However, this is actually the furthest that we need to go. Let's take a look at the following series.

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{n^2+2}{n^4} &= \sum_{n=1}^{\infty} \frac{n^2}{n^4} + \sum_{n=1}^{\infty} \frac{2}{n^4} \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2} + \sum_{n=1}^{\infty} \frac{2}{n^4}\end{aligned}$$

As shown, we can write the series as a sum of two series and both of these series are convergent by the p-series test. Therefore, since each of these series are convergent we know that the sum,

$$\sum_{n=1}^{\infty} \frac{n^2+2}{n^4}$$

is also a convergent series. Recall that the sum of two convergent series will also be convergent.

Now, since the terms of this series are larger than the terms of the original series we know that the original series must also be convergent by the Comparison Test.

The comparison test is a nice test that allows us to do problems that either we couldn't have done with the integral test or at the best would have been very difficult to do with the integral test. That doesn't mean that it doesn't have problems of its own.

Consider the following series.

$$\sum_{n=0}^{\infty} \frac{1}{3^n - n}$$

This is not much different from the first series that we looked at. The original series converged because the 3^n gets very large very fast and will be significantly larger than the n. Therefore, the n doesn't really affect the convergence of the series in that case. The fact that we are now subtracting the n off instead of adding the n on really shouldn't change the convergence. We can say this because the 3^n gets very large very fast and the fact that we're subtracting n off won't really change the size of this term for all sufficiently large values of n.

So, we would expect this series to converge. However, the comparison test won't work with this series. To use the comparison test on this series we would need to find a larger series that we could easily determine the convergence of. In this case we can't do what we did with the original series. If we drop the n we will make the denominator larger (since the n was subtracted off) and so the fraction will get smaller and just like when we looked at the comparison test for improper integrals knowing that the smaller of two series converges does not mean that the larger of the two will also converge.

So, we will need something else to do help us determine the convergence of this series. The following variant of the comparison test will allow us to determine the convergence of this series.

Limit Comparison Test

Suppose that we have two series $\sum a_n$ and $\sum b_n$ with $a_n \geq 0, b_n > 0$ for all n. Define,

$$c = \lim_{n\to\infty} \frac{a_n}{b_n}$$

If c is positive (*i.e.* $c > 0$) and is finite (*i.e.* $c < \infty$) then either both series converge or both series diverge.

The proof of this test is at the end of this section.

Note that it doesn't really matter which series term is in the numerator for this test, we could just have easily defined c as,

$$c = \lim_{n\to\infty} \frac{b_n}{a_n}$$

and we would get the same results. To see why this is, consider the following two definitions.

$$c = \lim_{n\to\infty} \frac{a_n}{b_n} \qquad \bar{c} = \lim_{n\to\infty} \frac{b_n}{a_n}$$

Start with the first definition and rewrite it as follows, then take the limit.

$$c = \lim_{n\to\infty} \frac{a_n}{b_n} = \lim_{n\to\infty} \frac{1}{\frac{b_n}{a_n}} = \frac{1}{\lim\limits_{n\to\infty} \frac{b_n}{a_n}} = \frac{1}{\overline{c}}$$

In other words, if c is positive and finite then so is $\overline{c}$ and if $\overline{c}$ is positive and finite then so is c. Likewise if $\overline{c} = 0$ then $c = \infty$ and if $\overline{c} = \infty$ then $c = 0$. Both definitions will give the same results from the test so don't worry about which series terms should be in the numerator and which should be in the denominator. Choose this to make the limit easy to compute.

Also, this really is a comparison test in some ways. If c is positive and finite this is saying that both of the series terms will behave in generally the same fashion and so we can expect the series themselves to also behave in a similar fashion. If $c = 0$ or $c = \infty$ we can't say this and so the test fails to give any information.

The limit in this test will often be written as,

$$c = \lim_{n\to\infty} a_n \cdot \frac{1}{b_n}$$

since often both terms will be fractions and this will make the limit easier to deal with.

Let's see how this test works.

Example 4

Determine if the following series converges or diverges.

$$\sum_{n=0}^{\infty} \frac{1}{3^n - n}$$

Solution

To use the limit comparison test we need to find a second series that we can determine the convergence of easily and has what we assume is the same convergence as the given series. On top of that we will need to choose the new series in such a way as to give us an easy limit to compute for c.

We've already guessed that this series converges and since it's vaguely geometric let's use

$$\sum_{n=0}^{\infty} \frac{1}{3^n}$$

as the second series. We know that this series converges and there is a chance that since both series have the 3^n in it the limit won't be too bad.

Here's the limit.

$$c = \lim_{n\to\infty} \frac{1}{3^n} \frac{3^n - n}{1}$$
$$= \lim_{n\to\infty} 1 - \frac{n}{3^n}$$

Now, we'll need to use L'Hospital's Rule on the second term in order to actually evaluate this limit.

$$c = 1 - \lim_{n\to\infty} \frac{1}{3^n \ln(3)}$$
$$= 1$$

So, c is positive and finite so by the Comparison Test both series must converge since

$$\sum_{n=0}^{\infty} \frac{1}{3^n}$$

converges.

Example 5

Determine if the following series converges or diverges.

$$\sum_{n=2}^{\infty} \frac{4n^2 + n}{\sqrt[3]{n^7 + n^3}}$$

Solution

Fractions involving only polynomials or polynomials under radicals will behave in the same way as the largest power of n will behave in the limit. So, the terms in this series should behave as,

$$\frac{n^2}{\sqrt[3]{n^7}} = \frac{n^2}{n^{\frac{7}{3}}} = \frac{1}{n^{\frac{1}{3}}}$$

and as a series this will diverge by the p-series test. In fact, this would make a nice choice for our second series in the limit comparison test so let's use it.

$$\begin{aligned}\lim_{n\to\infty}\frac{4n^2+n}{\sqrt[3]{n^7+n^3}}\frac{n^{\frac{1}{3}}}{1} &= \lim_{n\to\infty}\frac{4n^{\frac{7}{3}}+n^{\frac{4}{3}}}{\sqrt[3]{n^7\left(1+\frac{1}{n^4}\right)}}\\ &= \lim_{n\to\infty}\frac{n^{\frac{7}{3}}\left(4+\frac{1}{n}\right)}{n^{\frac{7}{3}}\sqrt[3]{1+\frac{1}{n^4}}}\\ &= \frac{4}{\sqrt[3]{1}} = 4 = c\end{aligned}$$

So, c is positive and finite and so both limits will diverge since

$$\sum_{n=2}^{\infty}\frac{1}{n^{\frac{1}{3}}}$$

diverges.

Finally, to see why we need c to be positive and finite (*i.e.* $c \neq 0$ and $c \neq \infty$) consider the following two series.

$$\sum_{n=1}^{\infty}\frac{1}{n} \qquad \sum_{n=1}^{\infty}\frac{1}{n^2}$$

The first diverges and the second converges.

Now compute each of the following limits.

$$\lim_{n\to\infty}\frac{1}{n}\cdot\frac{n^2}{1} = \lim_{n\to\infty} n = \infty \qquad \lim_{n\to\infty}\frac{1}{n^2}\cdot\frac{n}{1} = \lim_{n\to\infty}\frac{1}{n} = 0$$

In the first case the limit from the limit comparison test yields $c = \infty$ and in the second case the limit yields $c = 0$. Clearly, both series do not have the same convergence.

Note however, that just because we get $c = 0$ or $c = \infty$ doesn't mean that the series will have the opposite convergence. To see this consider the series,

$$\sum_{n=1}^{\infty}\frac{1}{n^3} \qquad \sum_{n=1}^{\infty}\frac{1}{n^2}$$

Both of these series converge and here are the two possible limits that the limit comparison test uses.

$$\lim_{n\to\infty}\frac{1}{n^3}\cdot\frac{n^2}{1} = \lim_{n\to\infty}\frac{1}{n} = 0 \qquad \lim_{n\to\infty}\frac{1}{n^2}\cdot\frac{n^3}{1} = \lim_{n\to\infty} n = \infty$$

So, even though both series had the same convergence we got both $c = 0$ and $c = \infty$.

The point of all of this is to remind us that if we get $c = 0$ or $c = \infty$ from the limit comparison test we will know that we have chosen the second series incorrectly and we'll need to find a different choice in order to get any information about the convergence of the series.

We'll close out this section with proofs of the two tests.

Proof of Comparison Test

The test statement did not specify where each series should start. We only need to require that they start at the same place so to help with the proof we'll assume that the series start at $n = 1$. If the series don't start at $n = 1$ the proof can be redone in exactly the same manner or you could use an index shift to start the series at $n = 1$ and then this proof will apply.

We'll start off with the partial sums of each series.

$$s_n = \sum_{i=1}^{n} a_i \qquad\qquad t_n = \sum_{i=1}^{n} b_i$$

Let's notice a couple of nice facts about these two partial sums. First, because $a_n, b_n \geq 0$ we know that,

$$s_n \leq s_n + a_{n+1} = \sum_{i=1}^{n} a_i + a_{n+1} = \sum_{i=1}^{n+1} a_i = s_{n+1} \qquad \Rightarrow \qquad s_n \leq s_{n+1}$$
$$t_n \leq t_n + b_{n+1} = \sum_{i=1}^{n} b_i + b_{n+1} = \sum_{i=1}^{n+1} b_i = t_{n+1} \qquad \Rightarrow \qquad t_n \leq t_{n+1}$$

So, both partial sums form increasing sequences.

Also, because $a_n \leq b_n$ for all n we know that we must have $s_n \leq t_n$ for all n.

With these preliminary facts out of the way we can proceed with the proof of the test itself.

Let's start out by assuming that $\sum\limits_{n=1}^{\infty} b_n$ is a convergent series. Since $b_n \geq 0$ we know that,

$$t_n = \sum_{i=1}^{n} b_i \leq \sum_{i=1}^{\infty} b_i$$

However, we also have established that $s_n \leq t_n$ for all n and so for all n we also have,

$$s_n \leq \sum_{i=1}^{\infty} b_i$$

Finally, since $\sum\limits_{n=1}^{\infty} b_n$ is a convergent series it must have a finite value and so the partial sums, s_n are bounded above. Therefore, from the second section on sequences we know that a monotonic and bounded sequence is also convergent and so $\{s_n\}_{n=1}^{\infty}$ is a convergent sequence and so $\sum\limits_{n=1}^{\infty} a_n$ is convergent.

Next, let's assume that $\sum\limits_{n=1}^{\infty} a_n$ is divergent. Because $a_n \geq 0$ we then know that we must have $s_n \to \infty$ as $n \to \infty$. However, we also know that for all n we have $s_n \leq t_n$ and therefore we also know that $t_n \to \infty$ as $n \to \infty$.

So, $\{t_n\}_{n=1}^{\infty}$ is a divergent sequence and so $\sum\limits_{n=1}^{\infty} b_n$ is divergent.

Proof of Limit Comparison Test

Because $0 < c < \infty$ we can find two positive and finite numbers, m and M, such that $m < c < M$. Now, because $c = \lim\limits_{n\to\infty} \frac{a_n}{b_n}$ we know that for large enough n the quotient $\frac{a_n}{b_n}$ must be close to c and so there must be a positive integer N such that if $n > N$ we also have,

$$m < \frac{a_n}{b_n} < M$$

Multiplying through by b_n gives,

$$mb_n < a_n < Mb_n$$

provided $n > N$.

Now, if $\sum b_n$ diverges then so does $\sum mb_n$ and so since $mb_n < a_n$ for all sufficiently large n by the Comparison Test $\sum a_n$ also diverges.

Likewise, if $\sum b_n$ converges then so does $\sum Mb_n$ and since $a_n < Mb_n$ for all sufficiently large n by the Comparison Test $\sum a_n$ also converges.

10.8 Alternating Series Test

The last two tests that we looked at for series convergence have required that all the terms in the series be positive. Of course there are many series out there that have negative terms in them and so we now need to start looking at tests for these kinds of series.

The test that we are going to look into in this section will be a test for alternating series. An **alternating series** is any series, $\sum a_n$, for which the series terms can be written in one of the following two forms.

$$\begin{aligned} a_n &= (-1)^n b_n & b_n \geq 0 \\ a_n &= (-1)^{n+1} b_n & b_n \geq 0 \end{aligned}$$

There are many other ways to deal with the alternating sign, but they can all be written as one of the two forms above. For instance,

$$\begin{aligned} (-1)^{n+2} &= (-1)^n (-1)^2 = (-1)^n \\ (-1)^{n-1} &= (-1)^{n+1} (-1)^{-2} = (-1)^{n+1} \end{aligned}$$

There are of course many others, but they all follow the same basic pattern of reducing to one of the first two forms given. If you should happen to run into a different form than the first two, don't worry about converting it to one of those forms, just be aware that it can be and so the test from this section can be used.

Alternating Series Test

Suppose that we have a series $\sum a_n$ and either $a_n = (-1)^n b_n$ or $a_n = (-1)^{n+1} b_n$ where $b_n \geq 0$ for all n. Then if,

1. $\lim\limits_{n \to \infty} b_n = 0$ and,
2. $\{b_n\}$ is a decreasing sequence

the series $\sum a_n$ is convergent.

A proof of this test is at the end of the section.

There are a couple of things to note about this test. First, unlike the Integral Test and the Comparison/Limit Comparison Test, this test will only tell us when a series converges and not if a series will diverge.

Secondly, in the second condition all that we need to require is that the series terms, b_n will be eventually decreasing. It is possible for the first few terms of a series to increase and still have the test be valid. All that is required is that eventually we will have $b_n \geq b_{n+1}$ for all n after some point.

To see why this is consider the following series,

$$\sum_{n=1}^{\infty} (-1)^n b_n$$

Let's suppose that for $1 \le n \le N$ $\{b_n\}$ is not decreasing and that for $n \ge N+1$ $\{b_n\}$ is decreasing. The series can then be written as,

$$\sum_{n=1}^{\infty} (-1)^n b_n = \sum_{n=1}^{N} (-1)^n b_n + \sum_{n=N+1}^{\infty} (-1)^n b_n$$

The first series is a finite sum (no matter how large N is) of finite terms and so we can compute its value and it will be finite. The convergence of the series will depend solely on the convergence of the second (infinite) series. If the second series has a finite value then the sum of two finite values is also finite and so the original series will converge to a finite value. On the other hand, if the second series is divergent either because its value is infinite or it doesn't have a value then adding a finite number onto this will not change that fact and so the original series will be divergent.

The point of all this is that we don't need to require that the series terms be decreasing for all n. We only need to require that the series terms will eventually be decreasing since we can always strip out the first few terms that aren't actually decreasing and look only at the terms that are actually decreasing.

Note that, in practice, we don't actually strip out the terms that aren't decreasing. All we do is check that eventually the series terms are decreasing and then apply the test.

Let's work a couple of examples.

Example 1

Determine if the following series is convergent or divergent.

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$$

Solution

First, identify the b_n for the test.

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{1}{n} \qquad\qquad b_n = \frac{1}{n}$$

Now, all that we need to do is run through the two conditions in the test.

$$\lim_{n \to \infty} b_n = \lim_{n \to \infty} \frac{1}{n} = 0$$

$$b_n = \frac{1}{n} > \frac{1}{n+1} = b_{n+1}$$

Both conditions are met and so by the Alternating Series Test the series must converge.

The series from the previous example is sometimes called the **Alternating Harmonic Series**. Also, the $(-1)^{n+1}$ could be $(-1)^n$ or any other form of alternating sign and we'd still call it an Alternating Harmonic Series.

In the previous example it was easy to see that the series terms decreased since increasing n only increased the denominator for the term and hence made the term smaller. In general however, we will need to resort to Calculus I techniques to prove the series terms decrease. We'll see an example of this in a bit.

Example 2

Determine if the following series is convergent or divergent.

$$\sum_{n=1}^{\infty} \frac{(-1)^n n^2}{n^2+5}$$

Solution

First, identify the b_n for the test.

$$\sum_{n=1}^{\infty} \frac{(-1)^n n^2}{n^2+5} = \sum_{n=1}^{\infty} (-1)^n \frac{n^2}{n^2+5} \qquad \Rightarrow \qquad b_n = \frac{n^2}{n^2+5}$$

Let's check the conditions.

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} \frac{n^2}{n^2+5} = 1 \neq 0$$

So, the first condition isn't met and so there is no reason to check the second. Since this condition isn't met we'll need to use another test to check convergence. In these cases where the first condition isn't met it is usually best to use the divergence test.

So, the divergence test requires us to compute the following limit.

$$\lim_{n\to\infty} \frac{(-1)^n n^2}{n^2+5}$$

This limit can be somewhat tricky to evaluate. For a second let's consider the following,

$$\lim_{n\to\infty} \frac{(-1)^n n^2}{n^2+5} = \left(\lim_{n\to\infty} (-1)^n\right)\left(\lim_{n\to\infty} \frac{n^2}{n^2+5}\right)$$

Splitting this limit like this can't be done because this operation requires that both limits exist and while the second one does the first clearly does not. However, it does show us how we can at least convince ourselves that the overall limit does not exist (even if it won't be a direct proof of that fact).

So, let's start with,

$$\lim_{n\to\infty}\frac{(-1)^n n^2}{n^2+5}=\lim_{n\to\infty}\left[(-1)^n\frac{n^2}{n^2+5}\right]$$

Now, the second part of this clearly is going to 1 as $n\to\infty$ while the first part just alternates between 1 and -1. So, as $n\to\infty$ the terms are alternating between positive and negative values that are getting closer and closer to 1 and -1 respectively.

In order for limits to exist we know that the terms need to settle down to a single number and since these clearly don't this limit doesn't exist and so by the Divergence Test this series diverges.

Example 3

Determine if the following series is convergent or divergent.

$$\sum_{n=0}^{\infty}\frac{(-1)^{n-3}\sqrt{n}}{n+4}$$

Solution

Notice that in this case the exponent on the "-1" isn't n or $n+1$. That won't change how the test works however so we won't worry about that. In this case we have,

$$b_n=\frac{\sqrt{n}}{n+4}$$

so let's check the conditions.

The first is easy enough to check.

$$\lim_{n\to\infty}b_n=\lim_{n\to\infty}\frac{\sqrt{n}}{n+4}=0$$

The second condition requires some work however. It is not immediately clear that these terms will decrease. Increasing n to $n+1$ will increase both the numerator and the denominator. Increasing the numerator says the term should also increase while increasing the denominator says that the term should decrease. Since it's not clear which of these will win out we will need to resort to Calculus I techniques to show that the terms decrease.

Let's start with the following function and its derivative.

$$f(x) = \frac{\sqrt{x}}{x+4} \qquad\qquad f'(x) = \frac{4-x}{2\sqrt{x}(x+4)^2}$$

Now, there are two critical points for this function, $x = 0$, and $x = 4$. Note that $x = -4$ is not a critical point because the function is not defined at $x = -4$. The first is outside the bound of our series so we won't need to worry about that one. Using the test points,

$$f'(1) = \frac{3}{50} \qquad\qquad f'(5) = -\frac{\sqrt{5}}{810}$$

and so we can see that the function in increasing on $0 \leq x \leq 4$ and decreasing on $x \geq 4$. Therefore, since $f(n) = b_n$ we know as well that the b_n are also increasing on $0 \leq n \leq 4$ and decreasing on $n \geq 4$.

The b_n are then eventually decreasing and so the second condition is met.

Both conditions are met and so by the Alternating Series Test the series must be converging.

As the previous example has shown, we sometimes need to do a fair amount of work to show that the terms are decreasing. Do not just make the assumption that the terms will be decreasing and let it go at that.

Let's do one more example just to make a point.

Example 4

Determine if the following series is convergent or divergent.

$$\sum_{n=2}^{\infty} \frac{\cos(n\pi)}{\sqrt{n}}$$

Solution

The point of this problem is really just to acknowledge that it is in fact an alternating series. To see this we need to acknowledge that,

$$\cos(n\pi) = (-1)^n$$

If you aren't sure of this you can easily convince yourself that this is correct by plugging in a few values of n and checking.

So the series is really,

$$\sum_{n=2}^{\infty} \frac{\cos(n\pi)}{\sqrt{n}} = \sum_{n=2}^{\infty} \frac{(-1)^n}{\sqrt{n}} \qquad \Rightarrow \qquad b_n = \frac{1}{\sqrt{n}}$$

Checking the two condition gives,

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} \frac{1}{\sqrt{n}} = 0$$

$$b_n = \frac{1}{\sqrt{n}} > \frac{1}{\sqrt{n+1}} = b_{n+1}$$

The two conditions of the test are met and so by the Alternating Series Test the series is convergent.

It should be pointed out that the rewrite we did in previous example only works because n is an integer and because of the presence of the π. Without the π we couldn't do this and if n wasn't guaranteed to be an integer we couldn't do this.

Let's close this section out with a proof of the Alternating Series Test.

Proof of Alternating Series Test

Without loss of generality we can assume that the series starts at $n = 1$. If not we could modify the proof below to meet the new starting place or we could do an index shift to get the series to start at $n = 1$.

Also note that the assumption here is that we have $a_n = (-1)^{n+1} b_n$. To get the proof for $a_n = (-1)^n b_n$ we only need to make minor modifications of the proof and so will not give that proof.

Finally, in the examples all we really needed was for the b_n to be positive and decreasing eventually but for this proof to work we really do need them to be positive and decreasing for all n.

First, notice that because the terms of the sequence are decreasing for any two successive terms we can say,

$$b_n - b_{n+1} \geq 0$$

Now, let's take a look at the even partial sums.

$$\begin{aligned}
s_2 &= b_1 - b_2 \geq 0 \\
s_4 &= b_1 - b_2 + b_3 - b_4 = s_2 + b_3 - b_4 \geq s_2 && \text{because } b_3 - b_4 \geq 0 \\
s_6 &= s_4 + b_5 - b_6 \geq s_4 && \text{because } b_5 - b_6 \geq 0 \\
&\vdots \\
s_{2n} &= s_{2n-2} + b_{2n-1} - b_{2n} \geq s_{2n-2} && \text{because } b_{2n-1} - b_{2n} \geq 0
\end{aligned}$$

So, $\{s_{2n}\}$ is an increasing sequence.

Next, we can also write the general term as,

$$\begin{aligned}
s_{2n} &= b_1 - b_2 + b_3 - b_4 + b_5 + \cdots - b_{2n-2} + b_{2n-1} - b_{2n} \\
&= b_1 - (b_2 - b_3) - (b_4 - b_5) + \cdots - (b_{2n-2} - b_{2n-1}) - b_{2n}
\end{aligned}$$

Each of the quantities in parenthesis are positive and by assumption we know that b_{2n} is also positive. So, this tells us that $s_{2n} \leq b_1$ for all n.

We now know that $\{s_{2n}\}$ is an increasing sequence that is bounded above and so we know that it must also converge. So, let's assume that its limit is s or,

$$\lim_{n \to \infty} s_{2n} = s$$

Next, we can quickly determine the limit of the sequence of odd partial sums, $\{s_{2n+1}\}$, as follows,

$$\lim_{n \to \infty} s_{2n+1} = \lim_{n \to \infty} (s_{2n} + b_{2n+1}) = \lim_{n \to \infty} s_{2n} + \lim_{n \to \infty} b_{2n+1} = s + 0 = s$$

So, we now know that both $\{s_{2n}\}$ and $\{s_{2n+1}\}$ are convergent sequences and they both have the same limit and so we also know that $\{s_n\}$ is a convergent sequence with a limit of s. This in turn tells us that $\sum a_n$ is convergent.

10.9 Absolute Convergence

When we first talked about series convergence we briefly mentioned a stronger type of convergence but didn't do anything with it because we didn't have any tools at our disposal that we could use to work problems involving it. We now have some of those tools so it's now time to talk about absolute convergence in detail.

First, let's go back over the definition of absolute convergence.

Definition

A series $\sum a_n$ is called **absolutely convergent** if $\sum |a_n|$ is convergent. If $\sum a_n$ is convergent and $\sum |a_n|$ is divergent we call the series **conditionally convergent**.

We also have the following fact about absolute convergence.

Fact

If $\sum a_n$ is absolutely convergent then it is also convergent.

Proof

First notice that $|a_n|$ is either a_n or it is $-a_n$ depending on its sign. This means that we can then say,

$$0 \le a_n + |a_n| \le 2\,|a_n|$$

Now, since we are assuming that $\sum |a_n|$ is convergent then $\sum 2\,|a_n|$ is also convergent since we can just factor the 2 out of the series and 2 times a finite value will still be finite. This however allows us to use the Comparison Test to say that $\sum (a_n + |a_n|)$ is also a convergent series.

Finally, we can write,

$$\sum a_n = \sum (a_n + |a_n|) - \sum |a_n|$$

and so $\sum a_n$ is the difference of two convergent series and so is also convergent.

This fact is one of the ways in which absolute convergence is a "stronger" type of convergence. Series that are absolutely convergent are guaranteed to be convergent. However, series that are convergent may or may not be absolutely convergent.

Let's take a quick look at a couple of examples of absolute convergence.

Example 1

Determine if each of the following series are absolute convergent, conditionally convergent or divergent.

(a) $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$

(b) $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+2}}{n^2}$

(c) $\displaystyle\sum_{n=1}^{\infty} \frac{\sin(n)}{n^3}$

Solution

(a) $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$

This is the alternating harmonic series and we saw in the last section that it is a convergent series so we don't need to check that here. So, let's see if it is an absolutely convergent series. To do this we'll need to check the convergence of.

$$\sum_{n=1}^{\infty} \left| \frac{(-1)^n}{n} \right| = \sum_{n=1}^{\infty} \frac{1}{n}$$

This is the harmonic series and we know from the integral test section that it is divergent.

Therefore, this series is not absolutely convergent. It is however conditionally convergent since the series itself does converge.

(b) $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+2}}{n^2}$

In this case let's just check absolute convergence first since if it's absolutely convergent we won't need to bother checking convergence as we will get that for free.

$$\sum_{n=1}^{\infty} \left| \frac{(-1)^{n+2}}{n^2} \right| = \sum_{n=1}^{\infty} \frac{1}{n^2}$$

This series is convergent by the p-series test and so the series is absolute convergent. Note that this does say as well that it's a convergent series.

(c) $\displaystyle\sum_{n=1}^{\infty} \frac{\sin(n)}{n^3}$

In this part we need to be a little careful. First, this is NOT an alternating series and so we can't use any tools from that section.

What we'll do here is check for absolute convergence first again since that will also give convergence. This means that we need to check the convergence of the following series.

$$\sum_{n=1}^{\infty} \left| \frac{\sin(n)}{n^3} \right| = \sum_{n=1}^{\infty} \frac{\left|\sin(n)\right|}{n^3}$$

To do this we'll need to note that

$$-1 \leq \sin(n) \leq 1 \qquad \Rightarrow \qquad \left|\sin(n)\right| \leq 1$$

and so we have,

$$\frac{\left|\sin(n)\right|}{n^3} \leq \frac{1}{n^3}$$

Now we know that

$$\sum_{n=1}^{\infty} \frac{1}{n^3}$$

converges by the p-series test and so by the Comparison Test we also know that

$$\sum_{n=1}^{\infty} \frac{\left|\sin(n)\right|}{n^3}$$

converges.

Therefore, the original series is absolutely convergent (and hence convergent).

Let's close this section off by recapping a topic we saw earlier. When we first discussed the convergence of series in detail we noted that we can't think of series as an infinite sum because some series can have different sums if we rearrange their terms. In fact, we gave two rearrangements of an Alternating Harmonic series that gave two different values. We closed that section off with the following fact,

Facts

Given the series $\sum a_n$,

1. If $\sum a_n$ is absolutely convergent and its value is s then any rearrangement of $\sum a_n$ will also have a value of s.
2. If $\sum a_n$ is conditionally convergent and r is any real number then there is a rearrangement of $\sum a_n$ whose value will be r.

Now that we've got the tools under our belt to determine absolute and conditional convergence we can make a few more comments about this.

First, as we showed above in Example 1a an Alternating Harmonic is conditionally convergent and so no matter what value we chose there is some rearrangement of terms that will give that value. Note as well that this fact does not tell us what that rearrangement must be only that it does exist.

Next, we showed in Example 1b that,

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+2}}{n^2}$$

is absolutely convergent and so no matter how we rearrange the terms of this series we'll always get the same value. In fact, it can be shown that the value of this series is,

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+2}}{n^2} = -\frac{\pi^2}{12}$$

10.10 Ratio Test

In this section we are going to take a look at a test that we can use to see if a series is absolutely convergent or not. Recall that if a series is absolutely convergent then we will also know that it's convergent and so we will often use it to simply determine the convergence of a series.

Before proceeding with the test let's do a quick reminder of factorials. This test will be particularly useful for series that contain factorials (and we will see some in the applications) so let's make sure we can deal with them before we run into them in an example.

If n is an integer such that $n \geq 0$ then n factorial is defined as,

$$\begin{aligned} & n! = n\,(n-1)\,(n-2)\cdots(3)\,(2)\,(1) && \text{if } n \geq 1 \\ & 0! = 1 && \text{by definition} \end{aligned}$$

Let's compute a couple real quick.

$$\begin{aligned} 1! &= 1 \\ 2! &= 2\,(1) = 2 \\ 3! &= 3\,(2)\,(1) = 6 \\ 4! &= 4\,(3)\,(2)\,(1) = 24 \\ 5! &= 5\,(4)\,(3)\,(2)\,(1) = 120 \end{aligned}$$

In the last computation above, notice that we could rewrite the factorial in a couple of different ways. For instance,

$$\begin{aligned} 5! &= 5\,\underbrace{(4)\,(3)\,(2)\,(1)}_{4!} = 5 \cdot 4! \\ 5! &= 5\,(4)\,\underbrace{(3)\,(2)\,(1)}_{3!} = 5\,(4) \cdot 3! \end{aligned}$$

In general, we can always "strip out" terms from a factorial as follows.

$$\begin{aligned} n! &= n\,(n-1)\,(n-2)\cdots(n-k)\,(n-(k+1))\cdots(3)\,(2)\,(1) \\ &= n\,(n-1)\,(n-2)\cdots(n-k)\cdot(n-(k+1))! \\ &= n\,(n-1)\,(n-2)\cdots(n-k)\cdot(n-k-1)! \end{aligned}$$

We will need to do this on occasion so don't forget about it.

Also, when dealing with factorials we need to be very careful with parenthesis. For instance, $(2n)! \neq 2\ n!$ as we can see if we write each of the following factorials out.

$$\begin{aligned} (2n)! &= (2n)\,(2n-1)\,(2n-2)\cdots(3)\,(2)\,(1) \\ 2\ n! &= 2\big[(n)\,(n-1)\,(n-2)\cdots(3)\,(2)\,(1)\big] \end{aligned}$$

Again, we will run across factorials with parenthesis so don't drop them. This is often one of the more common mistakes that students make when they first run across factorials.

Okay, we are now ready for the test.

Ratio Test

Suppose we have the series $\sum a_n$. Define,

$$L = \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right|$$

Then,

1. if $L < 1$ the series is absolutely convergent (and hence convergent).
2. if $L > 1$ the series is divergent.
3. if $L = 1$ the series may be divergent, conditionally convergent, or absolutely convergent.

A proof of this test is at the end of the section.

Notice that in the case of $L = 1$ the ratio test is pretty much worthless and we would need to resort to a different test to determine the convergence of the series.

Also, the absolute value bars in the definition of L are absolutely required. If they are not there it will be impossible for us to get the correct answer.

Let's take a look at some examples.

Example 1

Determine if the following series is convergent or divergent.

$$\sum_{n=1}^{\infty} \frac{(-10)^n}{4^{2n+1}(n+1)}$$

Solution

With this first example let's be a little careful and make sure that we have everything down correctly. Here are the series terms a_n.

$$a_n = \frac{(-10)^n}{4^{2n+1}(n+1)}$$

Recall that to compute a_{n+1} all that we need to do is substitute $n+1$ for all the n's in a_n.

$$a_{n+1} = \frac{(-10)^{n+1}}{4^{2(n+1)+1}\left((n+1)+1\right)} = \frac{(-10)^{n+1}}{4^{2n+3}\,(n+2)}$$

Now, to define L we will use,

$$L = \lim_{n\to\infty}\left|a_{n+1}\cdot\frac{1}{a_n}\right|$$

since this will be a little easier when dealing with fractions as we've got here. So,

$$\begin{aligned} L &= \lim_{n\to\infty}\left|\frac{(-10)^{n+1}}{4^{2n+3}\,(n+2)}\;\frac{4^{2n+1}\,(n+1)}{(-10)^{n}}\right| \\ &= \lim_{n\to\infty}\left|\frac{-10\,(n+1)}{4^2\,(n+2)}\right| \\ &= \frac{10}{16}\lim_{n\to\infty}\frac{n+1}{n+2} \\ &= \frac{10}{16} < 1 \end{aligned}$$

So, $L < 1$ and so by the Ratio Test the series converges absolutely and hence will converge.

As seen in the previous example there is usually a lot of canceling that will happen in these. Make sure that you do this canceling. If you don't do this kind of canceling it can make the limit fairly difficult.

Example 2

Determine if the following series is convergent or divergent.

$$\sum_{n=0}^{\infty}\frac{n!}{5^n}$$

Solution

Now that we've worked one in detail we won't go into quite the detail with the rest of these. Here is the limit.

$$L = \lim_{n\to\infty}\left|\frac{(n+1)!}{5^{n+1}}\;\frac{5^n}{n!}\right| = \lim_{n\to\infty}\frac{(n+1)!}{5\;n!}$$

In order to do this limit we will need to eliminate the factorials. We simply can't do the limit with the factorials in it. To eliminate the factorials we will recall from our discussion on factorials above that we can always “strip out” terms from a factorial. If we do that with the numerator (in this case because it’s the larger of the two) we get,

$$L = \lim_{n \to \infty} \frac{(n+1)\ \ n!}{5\ n!}$$

at which point we can cancel the $n!$ for the numerator an denominator to get,

$$L = \lim_{n \to \infty} \frac{(n+1)}{5} = \infty > 1$$

So, by the Ratio Test this series diverges.

Example 3

Determine if the following series is convergent or divergent.

$$\sum_{n=2}^{\infty} \frac{n^2}{(2n-1)!}$$

Solution

In this case be careful in dealing with the factorials.

$$\begin{aligned} L &= \lim_{n \to \infty} \left| \frac{(n+1)^2}{(2(n+1)-1)!} \frac{(2n-1)!}{n^2} \right| \\ &= \lim_{n \to \infty} \left| \frac{(n+1)^2}{(2n+1)!} \frac{(2n-1)!}{n^2} \right| \\ &= \lim_{n \to \infty} \frac{(n+1)^2}{(2n+1)(2n)(2n-1)!} \frac{(2n-1)!}{n^2} \\ &= \lim_{n \to \infty} \frac{(n+1)^2}{(2n+1)(2n)(n^2)} \\ &= 0 < 1 \end{aligned}$$

So, by the Ratio Test this series converges absolutely and so converges.

Example 4

Determine if the following series is convergent or divergent.

$$\sum_{n=1}^{\infty} \frac{9^n}{(-2)^{n+1} n}$$

Solution

Do not mistake this for a geometric series. The n in the denominator means that this isn't a geometric series. So, let's compute the limit.

$$\begin{aligned} L &= \lim_{n\to\infty} \left| \frac{9^{n+1}}{(-2)^{n+2}\,(n+1)} \frac{(-2)^{n+1}\,n}{9^n} \right| \\ &= \lim_{n\to\infty} \left| \frac{9\,n}{(-2)\,(n+1)} \right| \\ &= \frac{9}{2} \lim_{n\to\infty} \frac{n}{n+1} \\ &= \frac{9}{2} > 1 \end{aligned}$$

Therefore, by the Ratio Test this series is divergent.

In the previous example the absolute value bars were required to get the correct answer. If we hadn't used them we would have gotten $L = -\frac{9}{2} < 1$ which would have implied a convergent series!

Now, let's take a look at a couple of examples to see what happens when we get $L = 1$. Recall that the ratio test will not tell us anything about the convergence of these series. In both of these examples we will first verify that we get $L = 1$ and then use other tests to determine the convergence.

Example 5

Determine if the following series is convergent or divergent.

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{n^2+1}$$

Solution

Let's first get L.

$$L = \lim_{n\to\infty}\left|\frac{(-1)^{n+1}}{(n+1)^2+1}\,\frac{n^2+1}{(-1)^n}\right| = \lim_{n\to\infty}\frac{n^2+1}{(n+1)^2+1} = 1$$

So, as implied earlier we get $L = 1$ which means the ratio test is no good for determining the convergence of this series. We will need to resort to another test for this series. This series is an alternating series and so let's check the two conditions from that test.

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty}\frac{1}{n^2+1} = 0$$

$$b_n = \frac{1}{n^2+1} > \frac{1}{(n+1)^2+1} = b_{n+1}$$

The two conditions are met and so by the Alternating Series Test this series is convergent. We'll leave it to you to verify this series is also absolutely convergent.

Example 6

Determine if the following series is convergent or divergent.

$$\sum_{n=0}^{\infty}\frac{n+2}{2n+7}$$

Solution

Here's the limit.

$$L = \lim_{n\to\infty}\left|\frac{n+3}{2(n+1)+7}\,\frac{2n+7}{n+2}\right| = \lim_{n\to\infty}\frac{(n+3)(2n+7)}{(2n+9)(n+2)} = 1$$

Again, the ratio test tells us nothing here. We can however, quickly use the divergence test on this. In fact that probably should have been our first choice on this one anyway.

$$\lim_{n\to\infty}\frac{n+2}{2n+7} = \frac{1}{2} \neq 0$$

By the Divergence Test this series is divergent.

So, as we saw in the previous two examples if we get $L = 1$ from the ratio test the series can be either convergent or divergent.

There is one more thing that we should note about the ratio test before we move onto the next section. The last series was a polynomial divided by a polynomial and we saw that we got $L = 1$ from the ratio test. This will always happen with rational expression involving only polynomials or polynomials under radicals. So, in the future it isn't even worth it to try the ratio test on these kinds of problems since we now know that we will get $L = 1$.

Also, in the second to last example we saw an example of an alternating series in which the positive term was a rational expression involving polynomials and again we will always get $L = 1$ in these cases.

Let's close the section out with a proof of the Ratio Test.

Proof of Ratio Test

First note that we can assume without loss of generality that the series will start at $n = 1$ as we've done for all our series test proofs.

Let's start off the proof here by assuming that $L < 1$ and we'll need to show that $\sum a_n$ is absolutely convergent. To do this let's first note that because $L < 1$ there is some number r such that $L < r < 1$.

Now, recall that,

$$L = \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right|$$

and because we also have chosen r such that $L < r$ there is some N such that if $n \ge N$ we will have,

$$\left| \frac{a_{n+1}}{a_n} \right| < r \qquad \Rightarrow \qquad |a_{n+1}| < r\,|a_n|$$

Next, consider the following,

$$\begin{aligned}
|a_{N+1}| &< r\,|a_N| \\
|a_{N+2}| &< r\,|a_{N+1}| < r^2\,|a_N| \\
|a_{N+3}| &< r\,|a_{N+2}| < r^3\,|a_N| \\
&\vdots \\
|a_{N+k}| &< r\,|a_{N+k-1}| < r^k\,|a_N|
\end{aligned}$$

So, for $k = 1, 2, 3, \ldots$ we have $|a_{N+k}| < r^k\,|a_N|$. Just why is this important? Well we can now look at the following series.

$$\sum_{k=0}^{\infty} |a_N|\, r^k$$

This is a geometric series and because $0 < r < 1$ we in fact know that it is a convergent series. Also because $|a_{N+k}| < r^k |a_N|$ by the Comparison test the series

$$\sum_{n=N+1}^{\infty} |a_n| = \sum_{k=1}^{\infty} |a_{N+k}|$$

is convergent. However since,

$$\sum_{n=1}^{\infty} |a_n| = \sum_{n=1}^{N} |a_n| + \sum_{n=N+1}^{\infty} |a_n|$$

we know that $\sum\limits_{n=1}^{\infty} |a_n|$ is also convergent since the first term on the right is a finite sum of finite terms and hence finite. Therefore $\sum\limits_{n=1}^{\infty} a_n$ is absolutely convergent.

Next, we need to assume that $L > 1$ and we'll need to show that $\sum a_n$ is divergent. Recalling that,

$$L = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

and because $L > 1$ we know that there must be some N such that if $n \geq N$ we will have,

$$\left| \frac{a_{n+1}}{a_n} \right| > 1 \qquad \Rightarrow \qquad |a_{n+1}| > |a_n|$$

However, if $|a_{n+1}| > |a_n|$ for all $n \geq N$ then we know that,

$$\lim_{n\to\infty} |a_n| \neq 0$$

because the terms are getting larger and guaranteed to not be negative. This in turn means that,

$$\lim_{n\to\infty} a_n \neq 0$$

Therefore, by the Divergence Test $\sum a_n$ is divergent.

Finally, we need to assume that $L = 1$ and show that we could get a series that has any of the three possibilities. To do this we just need a series for each case. We'll leave the details of checking to you but all three of the following series have $L = 1$ and each one exhibits one of the possibilities.

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \qquad \text{absolutely convergent}$$

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \qquad \text{conditionally convergent}$$

$$\sum_{n=1}^{\infty} \frac{1}{n} \qquad \text{divergent}$$

10.11 Root Test

This is the last test for series convergence that we're going to be looking at. As with the Ratio Test this test will also tell whether a series is absolutely convergent or not rather than simple convergence.

Root Test

Suppose that we have the series $\sum a_n$. Define,

$$L = \lim_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} |a_n|^{\frac{1}{n}}$$

Then,

1. if $L < 1$ the series is absolutely convergent (and hence convergent).
2. if $L > 1$ the series is divergent.
3. if $L = 1$ the series may be divergent, conditionally convergent, or absolutely convergent.

A proof of this test is at the end of the section.

As with the ratio test, if we get $L = 1$ the root test will tell us nothing and we'll need to use another test to determine the convergence of the series. Also note that, generally for the series we'll be dealing with in this class, if $L = 1$ in the Ratio Test then the Root Test will also give $L = 1$.

We will also need the following fact in some of these problems.

Fact

$$\lim_{n\to\infty} n^{\frac{1}{n}} = 1$$

Let's take a look at a couple of examples.

Example 1

Determine if the following series is convergent or divergent.

$$\sum_{n=1}^{\infty} \frac{n^n}{3^{1+2n}}$$

Solution

There really isn't much to these problems other than computing the limit and then using the root test. Here is the limit for this problem.

$$L = \lim_{n\to\infty} \left|\frac{n^n}{3^{1+2n}}\right|^{\frac{1}{n}} = \lim_{n\to\infty} \frac{n}{3^{\frac{1}{n}+2}} = \frac{\infty}{3^2} = \infty > 1$$

So, by the Root Test this series is divergent.

Example 2

Determine if the following series is convergent or divergent.

$$\sum_{n=0}^{\infty} \left(\frac{5n-3n^3}{7n^3+2}\right)^n$$

Solution

Again, there isn't too much to this series.

$$L = \lim_{n\to\infty} \left|\left(\frac{5n-3n^3}{7n^3+2}\right)^n\right|^{\frac{1}{n}} = \lim_{n\to\infty} \left|\frac{5n-3n^3}{7n^3+2}\right| = \left|\frac{-3}{7}\right| = \frac{3}{7} < 1$$

Therefore, by the Root Test this series converges absolutely and hence converges.

Note that we had to keep the absolute value bars on the fraction until we'd taken the limit to get the sign correct.

Example 3

Determine if the following series is convergent or divergent.

$$\sum_{n=3}^{\infty} \frac{(-12)^n}{n}$$

Solution

Here's the limit for this series.

$$L = \lim_{n \to \infty} \left| \frac{(-12)^n}{n} \right|^{\frac{1}{n}} = \lim_{n \to \infty} \frac{12}{n^{\frac{1}{n}}} = \frac{12}{1} = 12 > 1$$

After using the fact from above we can see that the Root Test tells us that this series is divergent.

Proof of Root Test

First note that we can assume without loss of generality that the series will start at $n = 1$ as we've done for all our series test proofs. Also note that this proof is very similar to the proof of the Ratio Test.

Let's start off the proof here by assuming that $L < 1$ and we'll need to show that $\sum a_n$ is absolutely convergent. To do this let's first note that because $L < 1$ there is some number r such that $L < r < 1$.

Now, recall that,

$$L = \lim_{n \to \infty} \sqrt[n]{|a_n|} = \lim_{n \to \infty} |a_n|^{\frac{1}{n}}$$

and because we also have chosen r such that $L < r$ there is some N such that if $n \geq N$ we will have,

$$|a_n|^{\frac{1}{n}} < r \qquad \Rightarrow \qquad |a_n| < r^n$$

Now the series

$$\sum_{n=0}^{\infty} r^n$$

is a geometric series and because $0 < r < 1$ we in fact know that it is a convergent series. Also, because $|a_n| < r^n$ $n \geq N$ by the Comparison test the series

$$\sum_{n=N}^{\infty} |a_n|$$

is convergent. However since,

$$\sum_{n=1}^{\infty} |a_n| = \sum_{n=1}^{N-1} |a_n| + \sum_{n=N}^{\infty} |a_n|$$

we know that $\sum\limits_{n=1}^{\infty} |a_n|$ is also convergent since the first term on the right is a finite sum of finite terms and hence finite. Therefore $\sum\limits_{n=1}^{\infty} a_n$ is absolutely convergent.

Next, we need to assume that $L > 1$ and we'll need to show that $\sum a_n$ is divergent. Recalling that,

$$L = \lim_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} |a_n|^{\frac{1}{n}}$$

and because $L > 1$ we know that there must be some N such that if $n \geq N$ we will have,

$$|a_n|^{\frac{1}{n}} > 1 \qquad \Rightarrow \qquad |a_n| > 1^n = 1$$

However, if $|a_n| > 1$ for all $n \geq N$ then we know that,

$$\lim_{n\to\infty} |a_n| \neq 0$$

This in turn means that,

$$\lim_{n\to\infty} a_n \neq 0$$

Therefore, by the Divergence Test $\sum a_n$ is divergent.

Finally, we need to assume that $L = 1$ and show that we could get a series that has any of the three possibilities. To do this we just need a series for each case. We'll leave the details of checking to you but all three of the following series have $L = 1$ and each one exhibits one of the possibilities.

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \qquad \text{absolutely convergent}$$

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \qquad \text{conditionally convergent}$$

$$\sum_{n=1}^{\infty} \frac{1}{n} \qquad \text{divergent}$$

10.12 Strategy for Series

Now that we've got all of our tests out of the way it's time to think about organizing all of them into a general set of guidelines to help us determine the convergence of a series.

Note that these are a general set of guidelines and because some series can have more than one test applied to them we will get a different result depending on the path that we take through this set of guidelines. In fact, because more than one test may apply, you should always go completely through the guidelines and identify all possible tests that can be used on a given series. Once this has been done you can identify the test that you feel will be the easiest for you to use.

With that said here is the set of guidelines for determining the convergence of a series.

Strategy for Series

1. With a quick glance does it look like the series terms don't converge to zero in the limit, *i.e.* does $\lim\limits_{n\to\infty} a_n \neq 0$? If so, use the Divergence Test. Note that you should only do the Divergence Test if a quick glance suggests that the series terms may not converge to zero in the limit.
2. Is the series a p-series ($\sum \frac{1}{n^p}$) or a geometric series ($\sum\limits_{n=0}^{\infty} ar^n$ or $\sum\limits_{n=1}^{\infty} ar^{n-1}$)? If so use the fact that p-series will only converge if $p > 1$ and a geometric series will only converge if $|r| < 1$. Remember as well that often some algebraic manipulation is required to get a geometric series into the correct form.
3. Is the series similar to a p-series or a geometric series? If so, try the Comparison Test.
4. Is the series a rational expression involving only polynomials or polynomials under radicals (*i.e.* a fraction involving only polynomials or polynomials under radicals)? If so, try the Comparison Test and/or the Limit Comparison Test. Remember however, that in order to use the Comparison Test and the Limit Comparison Test the series terms all need to be positive.
5. Does the series contain factorials or constants raised to powers involving n? If so, then the Ratio Test may work. Note that if the series term contains a factorial then the only test that we've got that will work is the Ratio Test.
6. Can the series terms be written in the form $a_n = (-1)^n b_n$ or $a_n = (-1)^{n+1} b_n$? If so, then the Alternating Series Test may work.
7. Can the series terms be written in the form $a_n = (b_n)^n$? If so, then the Root Test may work.
8. If $a_n = f(n)$ for some positive, decreasing function and $\int_a^{\infty} f(x)\,dx$ is easy to evaluate then the Integral Test may work.

Again, remember that these are only a set of guidelines and not a set of hard and fast rules to use when trying to determine the best test to use on a series. If more than one test can be used try to use the test that will be the easiest for you to use and remember that what is easy for someone else may not be easy for you!

Also, just so we can put all the tests into one place here is a quick listing of all the tests that we've got.

Divergence Test
If $\lim\limits_{n\to\infty} a_n \neq 0$ then $\sum a_n$ will diverge

Integral Test
Suppose that $f(x)$ is a positive, decreasing function on the interval $[k,\infty)$ and that $f(n) = a_n$ then,

1. If $\displaystyle\int_k^\infty f(x)\,dx$ is convergent then so is $\displaystyle\sum_{n=k}^{\infty} a_n$.
2. If $\displaystyle\int_k^\infty f(x)\,dx$ is divergent then so is $\displaystyle\sum_{n=k}^{\infty} a_n$.

Comparison Test
Suppose that we have two series $\sum a_n$ and $\sum b_n$ with $a_n, b_n \geq 0$ for all n and $a_n \leq b_n$ for all n. Then,

1. If $\sum b_n$ is convergent then so is $\sum a_n$.
2. If $\sum a_n$ is divergent then so is $\sum b_n$.

Limit Comparison Test
Suppose that we have two series $\sum a_n$ and $\sum b_n$ with $a_n, b_n \geq 0$ for all n. Define,

$$c = \lim_{n\to\infty} \frac{a_n}{b_n}$$

If c is positive (*i.e.* $c > 0$) and is finite (*i.e.* $c < \infty$) then either both series converge or both series diverge.

Alternating Series Test
Suppose that we have a series $\sum a_n$ and either $a_n = (-1)^n b_n$ or $a_n = (-1)^{n+1} b_n$ where $b_n \geq 0$ for all n. Then if,

1. $\lim\limits_{n\to\infty} b_n = 0$ and,
2. $\{b_n\}$ is eventually a decreasing sequence

the series $\sum a_n$ is convergent

Ratio Test

Suppose we have the series $\sum a_n$. Define,

$$L = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

Then,

1. if $L < 1$ the series is absolutely convergent (and hence convergent).
2. if $L > 1$ the series is divergent.
3. if $L = 1$ the series may be divergent, conditionally convergent, or absolutely convergent.

Root Test

Suppose that we have the series $\sum a_n$. Define,

$$L = \lim_{n\to\infty} \sqrt[n]{|a_n|} = \lim_{n\to\infty} |a_n|^{\frac{1}{n}}$$

Then,

1. if $L < 1$ the series is absolutely convergent (and hence convergent).
2. if $L > 1$ the series is divergent.
3. if $L = 1$ the series may be divergent, conditionally convergent, or absolutely convergent.

10.13 Estimating the Value of a Series

We have now spent quite a few sections determining the convergence of a series, however, with the exception of geometric and telescoping series, we have not talked about finding the value of a series. This is usually a very difficult thing to do and we still aren't going to talk about how to find the value of a series. What we will do is talk about how to estimate the value of a series. Often that is all that you need to know.

Before we get into how to estimate the value of a series let's remind ourselves how series convergence works. It doesn't make any sense to talk about the value of a series that doesn't converge and so we will be assuming that the series we're working with converges. Also, as we'll see the main method of estimating the value of series will come out of this discussion.

So, let's start with the series $\sum\limits_{n=1}^{\infty} a_n$ (the starting point is not important, but we need a starting point to do the work) and let's suppose that the series converges to s. Recall that this means that if we get the partial sums,

$$s_n = \sum_{i=1}^{n} a_i$$

then they will form a convergent sequence and its limit is s. In other words,

$$\lim_{n \to \infty} s_n = s$$

Now, just what does this mean for us? Well, since this limit converges it means that we can make the partial sums, s_n, as close to s as we want simply by taking n large enough. In other words, if we take n large enough then we can say that,

$$s_n \approx s$$

This is one method of estimating the value of a series. We can just take a partial sum and use that as an estimation of the value of the series. There are now two questions that we should ask about this.

First, how good is the estimation? If we don't have an idea of how good the estimation is then it really doesn't do all that much for us as an estimation.

Secondly, is there any way to make the estimate better? Sometimes we can use this as a starting point and make the estimation better. We won't always be able to do this, but if we can that will be nice.

So, let's start with a general discussion about the determining how good the estimation is. Let's first start with the full series and strip out the first n terms.

$$\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{n} a_i + \sum_{i=n+1}^{\infty} a_i \tag{10.5}$$

Note that we converted over to an index of i in order to make the notation consistent with prior notation. Recall that we can use any letter for the index and it won't change the value.

Now, notice that the first series (the n terms that we've stripped out) is nothing more than the partial sum s_n. The second series on the right (the one starting at $i = n + 1$) is called the remainder and denoted by R_n. Finally let's acknowledge that we also know the value of the series since we are assuming it's convergent. Taking this notation into account we can rewrite Equation 10.5 as,

$$s = s_n + R_n$$

We can solve this for the remainder to get,

$$R_n = s - s_n$$

So, the remainder tells us the difference, or error, between the exact value of the series and the value of the partial sum that we are using as the estimation of the value of the series.

Of course, we can't get our hands on the actual value of the remainder because we don't have the actual value of the series. However, we can use some of the tests that we've got for convergence to get a pretty good estimate of the remainder provided we make some assumptions about the series. Once we've got an estimate on the value of the remainder we'll also have an idea on just how good a job the partial sum does of estimating the actual value of the series.

There are several tests that will allow us to get estimates of the remainder. We'll go through each one separately.

Also, when using the tests many of them had preconditions for use (*i.e.* terms had to be positive, terms had to be decreasing *etc.*) and when using the tests we noted that all we really needed was for them to eventually meet the preconditions in order for the test to work. For the following work however, we need the preconditions to always be met for all terms in the series.

If there are a few terms at the start where the preconditions aren't met we'll need to strip those terms out, do the estimate on the series that is left and then add in the terms we stripped out to get a final estimate of the series value.

Integral Test

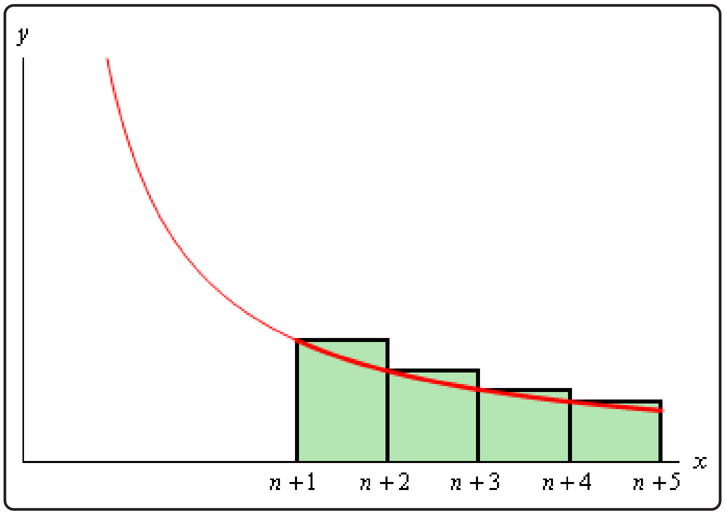
Recall that in this case we will need to assume that the series terms are all positive and be decreasing for all n. We derived the integral test by using the fact that the series could be thought of as an estimation of the area under the curve of $f(x)$ where $f(n) = a_n$. We can do something similar with the remainder.

As we'll soon see if we can get an upper and lower bound on the value of the remainder we can use these bounds to help us get upper and lower bounds on the value of the series. We can in turn use the upper and lower bounds on the series value to actually estimate the value of the series.

So, let's first recall that the remainder is,

$$R_n = \sum_{i=n+1}^{\infty} a_i = a_{n+1} + a_{n+2} + a_{n+3} + a_{n+4} + \cdots$$

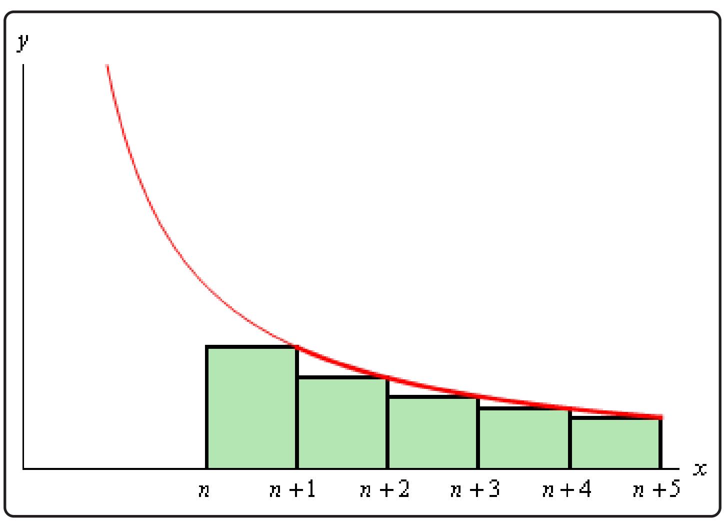
Now, if we start at $x = n + 1$, take rectangles of width 1 and use the left endpoint as the height of the rectangle we can estimate the area under $f(x)$ on the interval $[n + 1, \infty)$ as shown in the sketch below.



We can see that the remainder, R_n, is the area estimation and it will overestimate the exact area. So, we have the following inequality.

$$R_n \geq \int_{n+1}^{\infty} f(x)\, dx \tag{10.6}$$

Next, we could also estimate the area by starting at $x = n$, taking rectangles of width 1 again and then using the right endpoint as the height of the rectangle. This will give an estimation of the area under $f(x)$ on the interval $[n, \infty)$. This is shown in the following sketch.



Again, we can see that the remainder, R_n, is again this estimation and in this case it will underestimate the area. This leads to the following inequality,

$$R_n \le \int_n^\infty f(x)\,dx \tag{10.7}$$

Combining Equation 10.6 and Equation 10.7 gives,

$$\int_{n+1}^\infty f(x)\,dx \le R_n \le \int_n^\infty f(x)\,dx$$

So, provided we can do these integrals we can get both an upper and lower bound on the remainder. This will in turn give us an upper bound and a lower bound on just how good the partial sum, s_n, is as an estimation of the actual value of the series.

In this case we can also use these results to get a better estimate for the actual value of the series as well.

First, we'll start with the fact that

$$s = s_n + R_n$$

Now, if we use Equation 10.6 we get,

$$s = s_n + R_n \ge s_n + \int_{n+1}^\infty f(x)\,dx$$

Likewise if we use Equation 10.7 we get,

$$s = s_n + R_n \le s_n + \int_n^\infty f(x)\,dx$$

Putting these two together gives us,

Estimating Series with Integral Test

$$s_n + \int_{n+1}^\infty f(x)\,dx \le s \le s_n + \int_n^\infty f(x)\,dx \tag{10.8}$$

This gives an upper and a lower bound on the actual value of the series. We could then use as an estimate of the actual value of the series the average of the upper and lower bound.

Let's work an example with this.

Example 1

Using $n = 15$ to estimate the value of $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2}$.

Solution

First, for comparison purposes, we'll note that the actual value of this series is known to be,

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} = 1.644934068$$

Using $n = 15$ let's first get the partial sum.

$$s_{15} = \sum_{i=1}^{15} \frac{1}{i^2} = 1.580440283$$

Note that this is "close" to the actual value in some sense but isn't really all that close either.

Now, let's compute the integrals. These are fairly simple integrals, so we'll leave it to you to verify the values.

$$\int_{15}^{\infty} \frac{1}{x^2}\,dx = \frac{1}{15} \qquad \int_{16}^{\infty} \frac{1}{x^2}\,dx = \frac{1}{16}$$

Plugging these into Equation 10.8 gives us,

$$\begin{aligned} 1.580440283 + \frac{1}{16} &\le s \le 1.580440283 + \frac{1}{15} \\ 1.642940283 &\le s \le 1.647106950 \end{aligned}$$

Both the upper and lower bound are now very close to the actual value and if we take the average of the two we get the following estimate of the actual value.

$$s \approx 1.6450236165$$

That is pretty darn close to the actual value.

So, that is how we can use the Integral Test to estimate the value of a series. Let's move on to the next test.

Comparison Test

In this case, unlike with the integral test, we may or may not be able to get an idea of how good a particular partial sum will be as an estimate of the exact value of the series. Much of this will depend on how the comparison test is used.

First, let's remind ourselves on how the comparison test actually works. Given a series $\sum a_n$ let's assume that we've used the comparison test to show that it's convergent. Therefore, we found a second series $\sum b_n$ that converged and $a_n \le b_n$ for all n. Also recall that we need both a_n and b_n to be positive for all n.

What we want to do is determine how good of a job the partial sum,

$$s_n = \sum_{i=1}^{n} a_i$$

will do in estimating the actual value of the series $\sum a_n$. Again, we will use the remainder to do this. Let's actually write down the remainder for both series.

$$R_n = \sum_{i=n+1}^{\infty} a_i \qquad\qquad T_n = \sum_{i=n+1}^{\infty} b_i$$

Now, since $a_n \le b_n$ we also know that

$$R_n \le T_n$$

When using the comparison test it is often the case that the b_n are fairly nice terms and that we might actually be able to get an idea on the size of T_n. For instance, if our second series is a p-series we can use the results from above to get an upper bound on T_n as follows,

Estimating Series with Comparision Test

$$R_n \le T_n \le \int_n^{\infty} g\left(x\right)\,dx \qquad \text{where } g\left(n\right) = b_n$$

Also, if the second series is a geometric series then we will be able to compute T_n exactly.

If we are unable to get an idea of the size of T_n then using the comparison test to help with estimates won't do us much good.

Let's take a look at an example.

Example 2

Using $n = 15$ to estimate the value of $\displaystyle\sum_{n=0}^{\infty} \frac{2^n}{4^n + 1}$.

Solution

To do this we'll first need to go through the comparison test so we can get the second series. So,

$$\frac{2^n}{4^n + 1} \le \frac{2^n}{4^n} = \left(\frac{1}{2}\right)^n$$

and

$$\sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n$$

is a geometric series and converges because $|r| = \frac{1}{2} < 1$.

Now that we've gotten our second series let's get the estimate.

$$s_{15} = \sum_{n=0}^{15} \frac{2^n}{4^n + 1} = 1.383062486$$

So, how good is it? Well we know that,

$$R_{15} \le T_{15} = \sum_{n=16}^{\infty} \left(\frac{1}{2}\right)^n$$

will be an upper bound for the error between the actual value and the estimate. Since our second series is a geometric series we can compute this directly as follows.

$$\sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n = \sum_{n=0}^{15} \left(\frac{1}{2}\right)^n + \sum_{n=16}^{\infty} \left(\frac{1}{2}\right)^n$$

The series on the left is in the standard form and so we can compute that directly. The first series on the right has a finite number of terms and so can be computed exactly and the second series on the right is the one that we'd like to have the value for. Doing the work gives,

$$\begin{aligned} \sum_{n=16}^{\infty} \left(\frac{1}{2}\right)^n &= \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n - \sum_{n=0}^{15} \left(\frac{1}{2}\right)^n \\ &= \frac{1}{1 - \left(\frac{1}{2}\right)} - 1.999969482 \\ &= 0.000030518 \end{aligned}$$

So, according to this if we use

$$s \approx 1.383062486$$

as an estimate of the actual value we will be off from the exact value by no more than 0.000030518 and that's not too bad.

In this case it can be shown that

$$\sum_{n=0}^{\infty} \frac{2^n}{4^n + 1} = 1.383093004$$

and so we can see that the actual error in our estimation is,

$$\text{Error} = \text{Actual} - \text{Estimate} = 1.383093004 - 1.383062486 = 0.000030518$$

Note that in this case the estimate of the error is actually fairly close (and in fact exactly the same) as the actual error. This will not always happen and so we shouldn't expect that to happen in all cases. The error estimate above is simply the upper bound on the error and the actual error will often be less than this value.

Before moving on to the final part of this section let's again note that we will only be able to determine how good the estimate is using the comparison test if we can easily get our hands on the remainder of the second term. The reality is that we won't always be able to do this.

Alternating Series Test

Both of the methods that we've looked at so far have required the series to contain only positive terms. If we allow series to have negative terms in it the process is usually more difficult. However, with that said there is one case where it isn't too bad. That is the case of an alternating series.

Once again we will start off with a convergent series $\sum a_n = \sum (-1)^n b_n$ which in this case happens to be an alternating series that satisfies the conditions of the alternating series test, so we know that $b_n \geq 0$ and is decreasing for all n. Also note that we could have any power on the "-1" we just used n for the sake of convenience. We want to know how good of an estimation of the actual series value will the partial sum, s_n, be. As with the prior cases we know that the remainder, R_n, will be the error in the estimation and so if we can get a handle on that we'll know approximately how good the estimation is.

From the proof of the Alternating Series Test we can see that s will lie between s_n and s_{n+1} for any n and so,

$$|s - s_n| \leq |s_{n+1} - s_n| = b_{n+1}$$

Therefore,

Estimating Series with Alternating Series Test

$$|R_n| = |s - s_n| \leq b_{n+1}$$

We needed absolute value bars because we won't know ahead of time if the estimation is larger or smaller than the actual value and we know that the b_n's are positive.

Let's take a look at an example.

Example 3

Using $n = 15$ to estimate the value of $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2}$.

Solution

This is an alternating series and it does converge. In this case the exact value is known and so for comparison purposes,

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} = -\frac{\pi^2}{12} = -0.8224670336$$

Now, the estimation is,

$$s_{15} = \sum_{n=1}^{15} \frac{(-1)^n}{n^2} = -0.8245417574$$

From the fact above we know that

$$|R_{15}| = |s - s_{15}| \le b_{16} = \frac{1}{16^2} = 0.00390625$$

So, our estimation will have an error of no more than 0.00390625. In this case the exact value is known and so the actual error is,

$$|R_{15}| = |s - s_{15}| = 0.0020747238$$

In the previous example the estimation had only half the estimated error. It will often be the case that the actual error will be less than the estimated error. Remember that this is only an upper bound for the actual error.

Ratio Test

This will be the final case that we're going to look at for estimating series values and we are going to have to put a couple of fairly stringent restrictions on the series terms in order to do the work. One of the main restrictions we're going to make is to assume that the series terms are positive even though that is not required to actually use the test. We'll also be adding on another restriction in a bit.

In this case we've used the ratio test to show that $\sum a_n$ is convergent. To do this we computed

$$L = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

and found that $L < 1$.

As with the previous cases we are going to use the remainder, R_n, to determine how good of an estimation of the actual value the partial sum, s_n, is.

Estimating Series with Ratio Test

To get an estimate of the remainder let's first define the following sequence,

$$r_n = \frac{a_{n+1}}{a_n}$$

We now have two possible cases.

1. If $\{r_n\}$ is a decreasing sequence and $r_{n+1} < 1$ then,

$$R_n \le \frac{a_{n+1}}{1 - r_{n+1}}$$

2. If $\{r_n\}$ is an increasing sequence then,

$$R_n \le \frac{a_{n+1}}{1 - L}$$

Proof

Both parts will need the following work so we'll do it first. We'll start with the remainder.

$$\begin{aligned} R_n &= \sum_{i=n+1}^{\infty} a_i = a_{n+1} + a_{n+2} + a_{n+3} + a_{n+4} + \cdots \\ &= a_{n+1}\left(1 + \frac{a_{n+2}}{a_{n+1}} + \frac{a_{n+3}}{a_{n+1}} + \frac{a_{n+4}}{a_{n+1}} + \cdots\right) \end{aligned}$$

Next, we need to do a little work on a couple of these terms.

$$\begin{aligned} R_n &= a_{n+1}\left(1 + \frac{a_{n+2}}{a_{n+1}} + \frac{a_{n+3}}{a_{n+1}}\frac{a_{n+2}}{a_{n+2}} + \frac{a_{n+4}}{a_{n+1}}\frac{a_{n+2}}{a_{n+2}}\frac{a_{n+3}}{a_{n+3}} + \cdots\right) \\ &= a_{n+1}\left(1 + \frac{a_{n+2}}{a_{n+1}} + \frac{a_{n+2}}{a_{n+1}}\frac{a_{n+3}}{a_{n+2}} + \frac{a_{n+2}}{a_{n+1}}\frac{a_{n+3}}{a_{n+2}}\frac{a_{n+4}}{a_{n+3}} + \cdots\right) \end{aligned}$$

Now use the definition of r_n to write this as,

$$R_n = a_{n+1}\left(1 + r_{n+1} + r_{n+1}r_{n+2} + r_{n+1}r_{n+2}r_{n+3} + \cdots\right)$$

Okay now let's do the proof.

For the first part we are assuming that $\{r_n\}$ is decreasing and so we can estimate the remainder as,

$$\begin{aligned} R_n &= a_{n+1}\left(1 + r_{n+1} + r_{n+1}r_{n+2} + r_{n+1}r_{n+2}r_{n+3} + \cdots\right) \\ &\le a_{n+1}\left(1 + r_{n+1} + r_{n+1}^2 + r_{n+1}^3 + \cdots\right) \\ &= a_{n+1}\sum_{k=0}^{\infty} r_{n+1}^k \end{aligned}$$

Finally, the series here is a geometric series and because $r_{n+1} < 1$ we know that it converges and we can compute its value. So,

$$R_n \le \frac{a_{n+1}}{1 - r_{n+1}}$$

For the second part we are assuming that $\{r_n\}$ is increasing and we know that,

$$\lim_{n\to\infty} |r_n| = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = L$$

and so we know that $r_n < L$ for all n. The remainder can then be estimated as,

$$\begin{aligned} R_n &= a_{n+1}\left(1 + r_{n+1} + r_{n+1}r_{n+2} + r_{n+1}r_{n+2}r_{n+3} + \cdots\right) \\ &\le a_{n+1}\left(1 + L + L^2 + L^3 + \cdots\right) \\ &= a_{n+1}\sum_{k=0}^{\infty} L^k \end{aligned}$$

This is a geometric series and since we are assuming that our original series converges we also know that $L < 1$ and so the geometric series above converges and we can compute its value. So,

$$R_n \le \frac{a_{n+1}}{1 - L}$$

Note that there are some restrictions on the sequence $\{r_n\}$ and at least one of its terms in order to use these formulas. If the restrictions aren't met then the formulas can't be used.

Let's take a look at an example of this.

Example 4

Using $n = 15$ to estimate the value of $\displaystyle\sum_{n=0}^{\infty} \frac{n}{3^n}$.

Solution

First, let's use the ratio test to verify that this is a convergent series.

$$L = \lim_{n\to\infty} \left| \frac{n+1}{3^{n+1}} \, \frac{3^n}{n} \right| = \lim_{n\to\infty} \frac{n+1}{3n} = \frac{1}{3} < 1$$

So, it is convergent. Now let's get the estimate.

$$s_{15} = \sum_{n=0}^{15} \frac{n}{3^n} = 0.7499994250$$

To determine an estimate on the remainder, and hence the error, let's first get the sequence $\{r_n\}$.

$$r_n = \frac{n+1}{3^{n+1}} \, \frac{3^n}{n} = \frac{n+1}{3n} = \frac{1}{3}\left(1 + \frac{1}{n}\right)$$

The last rewrite was just to simplify some of the computations a little. Now, notice that,

$$f(x) = \frac{1}{3}\left(1 + \frac{1}{x}\right) \qquad\qquad f'(x) = -\frac{1}{3x^2} < 0$$

Since this function is always decreasing and $f(n) = r_n$ this sequence is decreasing. Also note that $r_{16} = \frac{1}{3}\left(1 + \frac{1}{16}\right) < 1$. Therefore, we can use the first case from the fact above to get,

$$R_{15} \le \frac{a_{16}}{1 - r_{16}} = \frac{\frac{16}{3^{16}}}{1 - \frac{1}{3}\left(1 + \frac{1}{16}\right)} = 0.0000005755187$$

So, it looks like our estimate is probably quite good. In this case the exact value is known.

$$\sum_{n=0}^{\infty} \frac{n}{3^n} = \frac{3}{4}$$

and so we can compute the actual error.

$$|R_{15}| = |s - s_{15}| = 0.000000575$$

This is less than the upper bound, but unlike in the previous example this actual error is quite close to the upper bound.

In the last two examples we've seen that the upper bound computations on the error can sometimes be quite close to the actual error and at other times they can be off by quite a bit. There is usually no way of knowing ahead of time which it will be and without the exact value in hand there will never be a way of determining which it will be.

Notice that this method did require the series terms to be positive, but that doesn't mean that we can't deal with ratio test series if they have negative terms. Often series that we used ratio test on are also alternating series and so if that is the case we can always resort to the previous material to get an upper bound on the error in the estimation, even if we didn't use the alternating series test to show convergence.

Note however that if the series does have negative terms but doesn't happen to be an alternating series then we can't use any of the methods discussed in this section to get an upper bound on the error.

10.14 Power Series

We've spent quite a bit of time talking about series now and with only a couple of exceptions we've spent most of that time talking about how to determine if a series will converge or not. It's now time to start looking at some specific kinds of series and we'll eventually reach the point where we can talk about a couple of applications of series.

In this section we are going to start talking about power series. A **power series about a**, or just **power series**, is any series that can be written in the form,

$$\sum_{n=0}^{\infty} c_n (x-a)^n$$

where a and c_n are numbers. The c_n's are often called the **coefficients** of the series. The first thing to notice about a power series is that it is a function of x. That is different from any other kind of series that we've looked at to this point. In all the prior sections we've only allowed numbers in the series and now we are allowing variables to be in the series as well. This will not change how things work however. Everything that we know about series still holds.

In the discussion of power series convergence is still a major question that we'll be dealing with. The difference is that the convergence of the series will now depend upon the values of x that we put into the series. A power series may converge for some values of x and not for other values of x.

Before we get too far into power series there is some terminology that we need to get out of the way.

First, as we will see in our examples, we will be able to show that there is a number R so that the power series will converge for, $|x-a| < R$ and will diverge for $|x-a| > R$. This number is called the **radius of convergence** for the series. Note that the series may or may not converge if $|x-a| = R$. What happens at these points will not change the radius of convergence.

Secondly, the interval of all x's, including the endpoints if need be, for which the power series converges is called the **interval of convergence** of the series.

These two concepts are fairly closely tied together. If we know that the radius of convergence of a power series is R then we have the following.

$$\begin{array}{cl} a-R < x < a+R & \text{power series converges} \\ x < a-R \text{ and } x > a+R & \text{power series diverges} \end{array}$$

The interval of convergence must then contain the interval $a-R < x < a+R$ since we know that the power series will converge for these values. We also know that the interval of convergence can't contain x's in the ranges $x < a-R$ and $x > a+R$ since we know the power series diverges for these value of x. Therefore, to completely identify the interval of convergence all that we have to do is determine if the power series will converge for $x = a-R$ or $x = a+R$. If the power series converges for one or both of these values then we'll need to include those in the interval of convergence.

Before getting into some examples let's take a quick look at the convergence of a power series for the case of $x = a$. In this case the power series becomes,

$$\sum_{n=0}^{\infty} c_n(a-a)^n = \sum_{n=0}^{\infty} c_n(0)^n = c_0(0)^0 + \sum_{n=1}^{\infty} c_n(0)^n = c_0 + \sum_{n=1}^{\infty} 0 = c_0 + 0 = c_0$$

and so the power series converges. Note that we had to strip out the first term since it was the only non-zero term in the series.

It is important to note that no matter what else is happening in the power series we are guaranteed to get convergence for $x = a$. The series may not converge for any other value of x, but it will always converge for $x = a$.

Let's work some examples. We'll put quite a bit of detail into the first example and then not put quite as much detail in the remaining examples.

Example 1

Determine the radius of convergence and interval of convergence for the following power series.

$$\sum_{n=1}^{\infty} \frac{(-1)^n n}{4^n}(x+3)^n$$

Solution

Okay, we know that this power series will converge for $x = -3$, but that's it at this point. To determine the remainder of the x's for which we'll get convergence we can use any of the tests that we've discussed to this point. After application of the test that we choose to work with we will arrive at condition(s) on x that we can use to determine the values of x for which the power series will converge and the values of x for which the power series will diverge. From this we can get the radius of convergence and most of the interval of convergence (with the possible exception of the endpoints).

With all that said, the best tests to use here are almost always the ratio or root test. Most of the power series that we'll be looking at are set up for one or the other. In this case we'll use the ratio test.

$$\begin{aligned} L &= \lim_{n\to\infty}\left|\frac{(-1)^{n+1}(n+1)(x+3)^{n+1}}{4^{n+1}}\,\frac{4^n}{(-1)^n(n)(x+3)^n}\right| \\ &= \lim_{n\to\infty}\left|\frac{-(n+1)(x+3)}{4n}\right| \end{aligned}$$

Before going any farther with the limit let's notice that since x is not dependent on the limit it can be factored out of the limit. Notice as well that in doing this we'll need to keep the absolute value bars on it since we need to make sure everything stays positive and x could well be a value that will make things negative. The limit is then,

$$\begin{aligned} L &= |x+3| \lim_{n\to\infty} \frac{n+1}{4n} \\ &= \frac{1}{4}|x+3| \end{aligned}$$

So, the ratio test tells us that if $L < 1$ the series will converge, if $L > 1$ the series will diverge, and if $L = 1$ we don't know what will happen. So, we have,

$$\begin{aligned} \frac{1}{4}|x+3| < 1 \qquad &\Rightarrow \qquad |x+3| < 4 \qquad \text{series converges} \\ \frac{1}{4}|x+3| > 1 \qquad &\Rightarrow \qquad |x+3| > 4 \qquad \text{series diverges} \end{aligned}$$

We'll deal with the $L = 1$ case in a bit. Notice that we now have the radius of convergence for this power series. These are exactly the conditions required for the radius of convergence. The radius of convergence for this power series is $R = 4$.

Now, let's get the interval of convergence. We'll get most (if not all) of the interval by solving the first inequality from above.

$$\begin{aligned} -4 < x+3 < 4 \\ -7 < x < 1 \end{aligned}$$

So, most of the interval of validity is given by $-7 < x < 1$. All we need to do is determine if the power series will converge or diverge at the endpoints of this interval. Note that these values of x will correspond to the value of x that will give $L = 1$.

The way to determine convergence at these points is to simply plug them into the original power series and see if the series converges or diverges using any test necessary.

$x = -7$:
In this case the series is,

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{(-1)^n n}{4^n}(-4)^n &= \sum_{n=1}^{\infty} \frac{(-1)^n n}{4^n}(-1)^n 4^n \\ &= \sum_{n=1}^{\infty} (-1)^n(-1)^n n \qquad\qquad (-1)^n(-1)^n = (-1)^{2n} = 1 \\ &= \sum_{n=1}^{\infty} n \end{aligned}$$

This series is divergent by the Divergence Test since $\lim\limits_{n\to\infty} n = \infty \neq 0$.

$x = 1$:
In this case the series is,

$$\sum_{n=1}^{\infty} \frac{(-1)^n n}{4^n}(4)^n = \sum_{n=1}^{\infty} (-1)^n n$$

This series is also divergent by the Divergence Test since $\lim\limits_{n\to\infty} (-1)^n n$ doesn't exist.

So, in this case the power series will not converge for either endpoint. The interval of convergence is then,

$$-7 < x < 1$$

In the previous example the power series didn't converge for either endpoint of the interval. Sometimes that will happen, but don't always expect that to happen. The power series could converge at either both of the endpoints or only one of the endpoints.

Example 2

Determine the radius of convergence and interval of convergence for the following power series.

$$\sum_{n=1}^{\infty} \frac{2^n}{n} (4x-8)^n$$

Solution

Let's jump right into the ratio test.

$$\begin{aligned} L &= \lim_{n\to\infty} \left| \frac{2^{n+1}(4x-8)^{n+1}}{n+1} \, \frac{n}{2^n (4x-8)^n} \right| \\ &= \lim_{n\to\infty} \left| \frac{2n\,(4x-8)}{n+1} \right| \\ &= |4x-8| \lim_{n\to\infty} \frac{2n}{n+1} \\ &= 2\,|4x-8| \end{aligned}$$

So we will get the following convergence/divergence information from this.

$$\begin{aligned} 2\,|4x-8| &< 1 && \text{series converges} \\ 2\,|4x-8| &> 1 && \text{series diverges} \end{aligned}$$

We need to be careful here in determining the interval of convergence. The interval of convergence requires $|x-a| < R$ and $|x-a| > R$. In other words, we need to factor a 4 out of the absolute value bars in order to get the correct radius of convergence. Doing this gives,

$$\begin{aligned} 8\,|x-2| &< 1 &\Rightarrow&& |x-2| &< \frac{1}{8} && \text{series converges} \\ 8\,|x-2| &> 1 &\Rightarrow&& |x-2| &> \frac{1}{8} && \text{series diverges} \end{aligned}$$

So, the radius of convergence for this power series is $R = \frac{1}{8}$.

Now, let's find the interval of convergence. Again, we'll first solve the inequality that gives convergence above.

$$\begin{aligned} -\frac{1}{8} &< x - 2 < \frac{1}{8} \\ \frac{15}{8} &< x < \frac{17}{8} \end{aligned}$$

Now check the endpoints.

$x = \dfrac{15}{8}$:

The series here is,

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{2^n}{n}\left(\frac{15}{2} - 8\right)^n &= \sum_{n=1}^{\infty} \frac{2^n}{n}\left(-\frac{1}{2}\right)^n \\ &= \sum_{n=1}^{\infty} \frac{2^n}{n}\frac{(-1)^n}{2^n} \\ &= \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \end{aligned}$$

This is the alternating harmonic series and we know that it converges.

$x = \dfrac{17}{8}$:

The series here is,

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{2^n}{n}\left(\frac{17}{2} - 8\right)^n &= \sum_{n=1}^{\infty} \frac{2^n}{n}\left(\frac{1}{2}\right)^n \\ &= \sum_{n=1}^{\infty} \frac{2^n}{n}\frac{1}{2^n} \\ &= \sum_{n=1}^{\infty} \frac{1}{n} \end{aligned}$$

This is the harmonic series and we know that it diverges.

So, the power series converges for one of the endpoints, but not the other. This will often happen so don't get excited about it when it does. The interval of convergence for this power series is then,

$$\frac{15}{8} \le x < \frac{17}{8}$$

We now need to take a look at a couple of special cases with radius and intervals of convergence.

Example 3

Determine the radius of convergence and interval of convergence for the following power series.

$$\sum_{n=0}^{\infty} n!(2x+1)^n$$

Solution

We'll start this example with the ratio test as we have for the previous ones.

$$\begin{aligned} L &= \lim_{n\to\infty} \left| \frac{(n+1)!(2x+1)^{n+1}}{n!(2x+1)^n} \right| \\ &= \lim_{n\to\infty} \left| \frac{(n+1)\, n!\;\, (2x+1)}{n!} \right| \\ &= |2x+1| \lim_{n\to\infty} (n+1) \end{aligned}$$

At this point we need to be careful. The limit is infinite, but there is that term with the x's in front of the limit. We'll have $L = \infty > 1$ provided $x \neq -\frac{1}{2}$.

So, this power series will only converge if $x = -\frac{1}{2}$. If you think about it we actually already knew that however. From our initial discussion we know that every power series will converge for $x = a$ and in this case $a = -\frac{1}{2}$. Remember that we get a from $(x-a)^n$, and notice the coefficient of the x must be a one!

In this case we say the radius of convergence is $R = 0$ and the interval of convergence is $x = -\frac{1}{2}$, and yes we really did mean interval of convergence even though it's only a point.

Example 4

Determine the radius of convergence and interval of convergence for the following power series.

$$\sum_{n=1}^{\infty} \frac{(x-6)^n}{n^n}$$

Solution

In this example the root test seems more appropriate. So,

$$\begin{aligned} L &= \lim_{n\to\infty} \left| \frac{(x-6)^n}{n^n} \right|^{\frac{1}{n}} \\ &= \lim_{n\to\infty} \left| \frac{x-6}{n} \right| \\ &= |x-6| \lim_{n\to\infty} \frac{1}{n} \\ &= 0 \end{aligned}$$

So, since $L = 0 < 1$ regardless of the value of x this power series will converge for every x.

In these cases, we say that the radius of convergence is $R = \infty$ and interval of convergence is $-\infty < x < \infty$.

So, let's summarize the last two examples. If the power series only converges for $x = a$ then the radius of convergence is $R = 0$ and the interval of convergence is $x = a$. Likewise, if the power series converges for every x the radius of convergence is $R = \infty$ and interval of convergence is $-\infty < x < \infty$.

Let's work one more example.

Example 5

Determine the radius of convergence and interval of convergence for the following power series.

$$\sum_{n=1}^{\infty} \frac{x^{2n}}{(-3)^n}$$

Solution

First notice that $a = 0$ in this problem. That's not really important to the problem, but it's worth pointing out so people don't get excited about it.

The important difference in this problem is the exponent on the x. In this case it is $2n$ rather than the standard n. As we will see some power series will have exponents other than an n and so we still need to be able to deal with these kinds of problems.

This one seems set up for the root test again so let's use that.

$$\begin{aligned} L &= \lim_{n\to\infty} \left| \frac{x^{2n}}{(-3)^n} \right|^{\frac{1}{n}} \\ &= \lim_{n\to\infty} \left| \frac{x^2}{-3} \right| \\ &= \frac{x^2}{3} \end{aligned}$$

So, we will get convergence if

$$\frac{x^2}{3} < 1 \qquad \Rightarrow \qquad x^2 < 3$$

The radius of convergence is NOT 3 however. The radius of convergence requires an exponent of 1 on the x. Therefore,

$$\begin{aligned} \sqrt{x^2} &< \sqrt{3} \\ |x| &< \sqrt{3} \end{aligned}$$

Be careful with the absolute value bars! In this case it looks like the radius of convergence is $R = \sqrt{3}$. Notice that we didn't bother to put down the inequality for divergence this time. The inequality for divergence is just the interval for convergence that the test gives with the inequality switched and generally isn't needed. We will usually skip that part.

Now let's get the interval of convergence. First from the inequality we get,

$$-\sqrt{3} < x < \sqrt{3}$$

Now check the endpoints.

$x = -\sqrt{3}$:

Here the power series is,

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{\left(-\sqrt{3}\right)^{2n}}{(-3)^n} &= \sum_{n=1}^{\infty} \frac{\left(\left(-\sqrt{3}\right)^2\right)^n}{(-3)^n} \\ &= \sum_{n=1}^{\infty} \frac{(3)^n}{(-1)^n(3)^n} \\ &= \sum_{n=1}^{\infty} (-1)^n \end{aligned}$$

This series is divergent by the Divergence Test since $\lim\limits_{n\to\infty} (-1)^n$ doesn't exist.

$x = \sqrt{3}$:

Because we're squaring the x this series will be the same as the previous step.

$$\sum_{n=1}^{\infty} \frac{\left(\sqrt{3}\right)^{2n}}{(-3)^n} = \sum_{n=1}^{\infty} (-1)^n$$

which is divergent.

The interval of convergence is then,

$$-\sqrt{3} < x < \sqrt{3}$$

10.15 Power Series and Functions

We opened the last section by saying that we were going to start thinking about applications of series and then promptly spent the section talking about convergence again. It's now time to actually start with the applications of series.

With this section we will start talking about how to represent functions with power series. The natural question of why we might want to do this will be answered in a couple of sections once we actually learn how to do this.

Let's start off with one that we already know how to do, although when we first ran across this series we didn't think of it as a power series nor did we acknowledge that it represented a function.

Recall that the geometric series is

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r} \qquad \text{provided } |r| < 1$$

Don't forget as well that if $|r| \geq 1$ the series diverges.

Now, if we take $a = 1$ and $r = x$ this becomes,

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \qquad \text{provided } |x| < 1 \tag{10.9}$$

Turning this around we can see that we can represent the function

$$f(x) = \frac{1}{1-x} \tag{10.10}$$

with the power series

$$\sum_{n=0}^{\infty} x^n \qquad \text{provided } |x| < 1 \tag{10.11}$$

This provision is important. We can clearly plug any number other than $x = 1$ into the function, however, we will only get a convergent power series if $|x| < 1$. This means the equality in Equation 10.9 will only hold if $|x| < 1$. For any other value of x the equality won't hold. Note as well that we can also use this to acknowledge that the radius of convergence of this power series is $R = 1$ and the interval of convergence is $|x| < 1$.

This idea of convergence is important here. We will be representing many functions as power series and it will be important to recognize that the representations will often only be valid for a range of x's and that there may be values of x that we can plug into the function that we can't plug into the power series representation.

In this section we are going to concentrate on representing functions with power series where the functions can be related back to Equation 10.10.

In this way we will hopefully become familiar with some of the kinds of manipulations that we will sometimes need to do when working with power series.

So, let's jump into a couple of examples.

Example 1

Find a power series representation for the following function and determine its interval of convergence.

$$g(x) = \frac{1}{1 + x^3}$$

Solution

What we need to do here is to relate this function back to Equation 10.10. This is actually easier than it might look. Recall that the x in Equation 10.10 is simply a variable and can represent anything. So, a quick rewrite of $g(x)$ gives,

$$g(x) = \frac{1}{1 - (-x^3)}$$

and so the $-x^3$ in $g(x)$ holds the same place as the x in Equation 10.10. Therefore, all we need to do is replace the x in Equation 10.11 and we've got a power series representation for $g(x)$.

$$g(x) = \sum_{n=0}^{\infty} \left(-x^3\right)^n \qquad \text{provided } \left|-x^3\right| < 1$$

Notice that we replaced both the x in the power series and in the interval of convergence.

All we need to do now is a little simplification.

$$g(x) = \sum_{n=0}^{\infty} (-1)^n x^{3n} \qquad \text{provided } |x|^3 < 1 \qquad \Rightarrow \qquad |x| < 1$$

So, in this case the interval of convergence is the same as the original power series. This usually won't happen. More often than not the new interval of convergence will be different from the original interval of convergence.

Example 2

Find a power series representation for the following function and determine its interval of convergence.

$$h(x) = \frac{2x^2}{1+x^3}$$

Solution

This function is similar to the previous function. The difference is the numerator and at first glance that looks to be an important difference. Since Equation 10.10 doesn't have an x in the numerator it appears that we can't relate this function back to that.

However, now that we've worked the first example this one is actually very simple since we can use the result of the answer from that example. To see how to do this let's first rewrite the function a little.

$$h(x) = 2x^2 \frac{1}{1+x^3}$$

Now, from the first example we've already got a power series for the second term so let's use that to write the function as,

$$h(x) = 2x^2 \sum_{n=0}^{\infty} (-1)^n x^{3n} \qquad \text{provided } |x| < 1$$

Notice that the presence of x's outside of the series will NOT affect its convergence and so the interval of convergence remains the same.

The last step is to bring the coefficient into the series and we'll be done. When we do this make sure and combine the x's as well. We typically only want a single x in a power series.

$$h(x) = \sum_{n=0}^{\infty} 2(-1)^n x^{3n+2} \qquad \text{provided } |x| < 1$$

As we saw in the previous example we can often use previous results to help us out. This is an important idea to remember as it can often greatly simplify our work.

Example 3

Find a power series representation for the following function and determine its interval of convergence.

$$f(x) = \frac{x}{5-x}$$

Solution

So, again, we've got an x in the numerator. So, as with the last example let's factor that out and see what we've got left.

$$f(x) = x\,\frac{1}{5-x}$$

If we had a power series representation for

$$g(x) = \frac{1}{5-x}$$

we could get a power series representation for $f(x)$.

So, let's find one. We'll first notice that in order to use Equation 10.10 we'll need the number in the denominator to be a one. That's easy enough to get.

$$g(x) = \frac{1}{5}\,\frac{1}{1-\frac{x}{5}}$$

Now all we need to do to get a power series representation is to replace the x in Equation 10.11 with $\frac{x}{5}$. Doing this gives,

$$g(x) = \frac{1}{5}\sum_{n=0}^{\infty}\left(\frac{x}{5}\right)^n \qquad \text{provided } \left|\frac{x}{5}\right| < 1$$

Now let's do a little simplification on the series.

$$\begin{aligned} g(x) &= \frac{1}{5}\sum_{n=0}^{\infty}\frac{x^n}{5^n} \\ &= \sum_{n=0}^{\infty}\frac{x^n}{5^{n+1}} \end{aligned}$$

The interval of convergence for this series is,

$$\left|\frac{x}{5}\right| < 1 \qquad \Rightarrow \qquad \frac{1}{5}|x| < 1 \qquad \Rightarrow \qquad |x| < 5$$

Okay, this was the work for the power series representation for $g(x)$ let's now find a power series representation for the original function. All we need to do for this is to multiply the power series representation for $g(x)$ by x and we'll have it.

$$\begin{aligned} f(x) &= x \frac{1}{5-x} \\ &= x \sum_{n=0}^{\infty} \frac{x^n}{5^{n+1}} \\ &= \sum_{n=0}^{\infty} \frac{x^{n+1}}{5^{n+1}} \end{aligned}$$

The interval of convergence doesn't change and so it will be $|x| < 5$.

So, hopefully we now have an idea on how to find the power series representation for some functions. Admittedly all of the functions could be related back to Equation 10.10 but it's a start.

We now need to look at some further manipulation of power series that we will need to do on occasion. We need to discuss differentiation and integration of power series.

Let's start with differentiation of the power series,

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots$$

Now, we know that if we differentiate a finite sum of terms all we need to do is differentiate each of the terms and then add them back up. With infinite sums there are some subtleties involved that we need to be careful with but are somewhat beyond the scope of this course.

Nicely enough for us however, it is known that if the power series representation of $f(x)$ has a radius of convergence of $R > 0$ then the term by term differentiation of the power series will also have a radius of convergence of R and (more importantly) will in fact be the power series representation of $f'(x)$ provided we stay within the radius of convergence.

Again, we should make the point that if we aren't dealing with a power series then we may or may not be able to differentiate each term of the series to get the derivative of the series.

So, what all this means for us is that,

$$f'(x) = \frac{d}{dx} \sum_{n=0}^{\infty} c_n (x-a)^n = c_1 + 2c_2(x-a) + 3c_3(x-a)^2 + \cdots = \sum_{n=1}^{\infty} n c_n (x-a)^{n-1}$$

Note the initial value of this series. It has been changed from $n = 0$ to $n = 1$. This is an acknowledgement of the fact that the derivative of the first term is zero and hence isn't in the derivative. Notice however, that since the $n = 0$ term of the above series is also zero, we could start the series at $n = 0$ if it was required for a particular problem. In general, however, this won't be done in this class.

We can now find formulas for higher order derivatives as well now.

$$\begin{aligned} f''(x) &= \sum_{n=2}^{\infty} n(n-1)c_n(x-a)^{n-2} \\ f'''(x) &= \sum_{n=3}^{\infty} n(n-1)(n-2)c_n(x-a)^{n-3} \\ &\quad etc. \end{aligned}$$

Once again, notice that the initial value of n changes with each differentiation in order to acknowledge that a term from the original series differentiated to zero.

Let's now briefly talk about integration. Just as with the differentiation, when we've got an infinite series we need to be careful about just integration term by term. Much like with derivatives it turns out that as long as we're working with power series we can just integrate the terms of the series to get the integral of the series itself. In other words,

$$\begin{aligned} \int f(x)\,dx &= \int \sum_{n=0}^{\infty} c_n(x-a)^n\,dx \\ &= \sum_{n=0}^{\infty} \int c_n(x-a)^n\,dx \\ &= C + \sum_{n=0}^{\infty} c_n \frac{(x-a)^{n+1}}{n+1} \end{aligned}$$

Notice that we pick up a constant of integration, C, that is outside the series here.

Let's summarize the differentiation and integration ideas before moving on to an example or two.

Fact

If $f(x) = \sum\limits_{n=0}^{\infty} c_n(x-a)^n$ has a radius of convergence of $R > 0$ then,

$$f'(x) = \sum_{n=1}^{\infty} nc_n(x-a)^{n-1} \qquad \int f(x)\,dx = C + \sum_{n=0}^{\infty} c_n \frac{(x-a)^{n+1}}{n+1}$$

and both of these also have a radius of convergence of R.

Now, let's see how we can use these facts to generate some more power series representations of functions.

Example 4

Find a power series representation for the following function and determine its radius of convergence.

$$g(x) = \frac{1}{(1-x)^2}$$

Solution

To do this problem let's notice that

$$\frac{1}{(1-x)^2} = \frac{d}{dx}\left(\frac{1}{1-x}\right)$$

Then since we've got a power series representation for

$$\frac{1}{1-x}$$

all that we'll need to do is differentiate that power series to get a power series representation for $g(x)$.

$$\begin{aligned} g(x) &= \frac{1}{(1-x)^2} \\ &= \frac{d}{dx}\left(\frac{1}{1-x}\right) \\ &= \frac{d}{dx}\left(\sum_{n=0}^{\infty} x^n\right) \\ &= \sum_{n=1}^{\infty} n x^{n-1} \end{aligned}$$

Then since the original power series had a radius of convergence of $R = 1$ the derivative, and hence $g(x)$, will also have a radius of convergence of $R = 1$.

Example 5

Find a power series representation for the following function and determine its radius of convergence.

$$h(x) = \ln(5 - x)$$

Solution

In this case we need to notice that

$$\int \frac{1}{5-x}\,dx = -\ln(5 - x)$$

and then recall that we have a power series representation for

$$\frac{1}{5-x}$$

Remember we found a representation for this in Example 3. So,

$$\begin{aligned}\ln(5-x) &= -\int \frac{1}{5-x}\,dx \\ &= -\int \sum_{n=0}^{\infty} \frac{x^n}{5^{n+1}}\,dx \\ &= C - \sum_{n=0}^{\infty} \frac{x^{n+1}}{(n+1)\,5^{n+1}}\end{aligned}$$

We can find the constant of integration, C, by plugging in a value of x. A good choice is $x = 0$ since that will make the series easy to evaluate.

$$\begin{aligned}\ln(5-0) &= C - \sum_{n=0}^{\infty} \frac{0^{n+1}}{(n+1)\,5^{n+1}} \\ \ln(5) &= C\end{aligned}$$

So, the final answer is,

$$\ln(5-x) = \ln(5) - \sum_{n=0}^{\infty} \frac{x^{n+1}}{(n+1)\,5^{n+1}}$$

Note that it is okay to have the constant sitting outside of the series like this. In fact, there is no way to bring it into the series so don't get excited about it.

Finally, because the power series representation from Example 3 had a radius of convergence of $R = 5$ this series will also have a radius of convergence of $R = 5$.

10.16 Taylor Series

In the previous section we started looking at writing down a power series representation of a function. The problem with the approach in that section is that everything came down to needing to be able to relate the function in some way to

$$\frac{1}{1-x}$$

and while there are many functions out there that can be related to this function there are many more that simply can't be related to this.

So, without taking anything away from the process we looked at in the previous section, what we need to do is come up with a more general method for writing a power series representation for a function.

So, for the time being, let's make two assumptions. First, let's assume that the function $f(x)$ does in fact have a power series representation about $x = a$,

$$f(x) = \sum_{n=0}^{\infty} c_n(x-a)^n = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + c_4(x-a)^4 + \cdots$$

Next, we will need to assume that the function, $f(x)$, has derivatives of every order and that we can in fact find them all.

Now that we've assumed that a power series representation exists we need to determine what the coefficients, c_n, are. This is easier than it might at first appear to be. Let's first just evaluate everything at $x = a$. This gives,

$$f(a) = c_0$$

So, all the terms except the first are zero and we now know what c_0 is. Unfortunately, there isn't any other value of x that we can plug into the function that will allow us to quickly find any of the other coefficients. However, if we take the derivative of the function (and its power series) then plug in $x = a$ we get,

$$\begin{aligned} f'(x) &= c_1 + 2c_2(x-a) + 3c_3(x-a)^2 + 4c_4(x-a)^3 + \cdots \\ f'(a) &= c_1 \end{aligned}$$

and we now know c_1.

Let's continue with this idea and find the second derivative.

$$\begin{aligned} f''(x) &= 2c_2 + 3(2)c_3(x-a) + 4(3)c_4(x-a)^2 + \cdots \\ f''(a) &= 2c_2 \end{aligned}$$

So, it looks like,

$$c_2 = \frac{f''(a)}{2}$$

Using the third derivative gives,

$$f'''(x) = 3(2)c_3 + 4(3)(2)c_4(x-a) + \cdots$$
$$f'''(a) = 3(2)c_3 \qquad \Rightarrow \qquad c_3 = \frac{f'''(a)}{3(2)}$$

Using the fourth derivative gives,

$$f^{(4)}(x) = 4(3)(2)c_4 + 5(4)(3)(2)c_5(x-a)\cdots$$
$$f^{(4)}(a) = 4(3)(2)c_4 \qquad \Rightarrow \qquad c_4 = \frac{f^{(4)}(a)}{4(3)(2)}$$

Hopefully by this time you've seen the pattern here. It looks like, in general, we've got the following formula for the coefficients.

$$c_n = \frac{f^{(n)}(a)}{n!}$$

This even works for $n = 0$ if you recall that $0! = 1$ and define $f^{(0)}(x) = f(x)$.

So, provided a power series representation for the function $f(x)$ about $x = a$ exists the **Taylor Series for** $f(x)$ **about** $x = a$ is,

Taylor Series

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$
$$= f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots$$

If we use $a = 0$, so we are talking about the Taylor Series about $x = 0$, we call the series a **Maclaurin Series** for $f(x)$ or,

Maclaurin Series

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n$$
$$= f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots$$

Before working any examples of Taylor Series we first need to address the assumption that a Taylor Series will in fact exist for a given function. Let's start out with some notation and definitions that we'll need.

To determine a condition that must be true in order for a Taylor series to exist for a function let's first define the n^{th} **degree Taylor polynomial of** $f(x)$ as,

$$T_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i$$

Note that this really is a polynomial of degree at most n. If we were to write out the sum without the summation notation this would clearly be an n^{th} degree polynomial. We'll see a nice application of Taylor polynomials in the next section.

Notice as well that for the full Taylor Series,

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

the n^{th} degree Taylor polynomial is just the partial sum for the series.

Next, the **remainder** is defined to be,

$$R_n(x) = f(x) - T_n(x)$$

So, the remainder is really just the *error* between the function $f(x)$ and the n^{th} degree Taylor polynomial for a given n.

With this definition note that we can then write the function as,

$$f(x) = T_n(x) + R_n(x)$$

We now have the following Theorem.

Theorem

Suppose that $f(x) = T_n(x) + R_n(x)$. Then if,

$$\lim_{n\to\infty} R_n(x) = 0$$

for $|x-a| < R$ then,

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

on $|x-a| < R$.

In general, showing that

$$\lim_{n\to\infty} R_n(x) = 0$$

is a somewhat difficult process and so we will be assuming that this can be done for some R in all of the examples that we'll be looking at.

Now let's look at some examples.

Example 1

Find the Taylor Series for $f(x) = \mathbf{e}^x$ about $x = 0$.

Solution

This is actually one of the easier Taylor Series that we'll be asked to compute. To find the Taylor Series for a function we will need to determine a general formula for $f^{(n)}(a)$. This is one of the few functions where this is easy to do right from the start.

To get a formula for $f^{(n)}(0)$ all we need to do is recognize that,

$$f^{(n)}(x) = \mathbf{e}^x \qquad n = 0, 1, 2, 3, \ldots$$

and so,

$$f^{(n)}(0) = \mathbf{e}^0 = 1 \qquad n = 0, 1, 2, 3, \ldots$$

Therefore, the Taylor series for $f(x) = \mathbf{e}^x$ about $x = 0$ is,

$$\mathbf{e}^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Example 2

Find the Taylor Series for $f(x) = \mathbf{e}^{-x}$ about $x = 0$.

Solution

There are two ways to do this problem. Both are fairly simple, however one of them requires significantly less work. We'll work both solutions since the longer one has some nice ideas that we'll see in other examples.

Solution 1

As with the first example we'll need to get a formula for $f^{(n)}(0)$. However, unlike the first one we've got a little more work to do. Let's first take some derivatives and evaluate them at $x = 0$.

$$\begin{aligned}
f^{(0)}(x) &= \mathbf{e}^{-x} & f^{(0)}(0) &= 1 \\
f^{(1)}(x) &= -\mathbf{e}^{-x} & f^{(1)}(0) &= -1 \\
f^{(2)}(x) &= \mathbf{e}^{-x} & f^{(2)}(0) &= 1 \\
f^{(3)}(x) &= -\mathbf{e}^{-x} & f^{(3)}(0) &= -1 \\
&\vdots & &\vdots \\
f^{(n)}(x) &= (-1)^n \mathbf{e}^{-x} & f^{(n)}(0) &= (-1)^n \qquad n = 0, 1, 2, 3
\end{aligned}$$

After a couple of computations we were able to get general formulas for both $f^{(n)}(x)$ and $f^{(n)}(0)$. We often won't be able to get a general formula for $f^{(n)}(x)$ so don't get too excited about getting that formula. Also, as we will see it won't always be easy to get a general formula for $f^{(n)}(a)$.

So, in this case we've got general formulas so all we need to do is plug these into the Taylor Series formula and be done with the problem.

$$\mathbf{e}^{-x} = \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{n!}$$

Solution 2

The previous solution wasn't too bad and we often have to do things in that manner. However, in this case there is a much shorter solution method. In the previous section we used series that we've already found to help us find a new series. Let's do the same thing with this one. We already know a Taylor Series for $\mathbf{e}^x$ about $x = 0$ and in this case the only difference is we've got a "$-x$" in the exponent instead of just an x.

So, all we need to do is replace the x in the Taylor Series that we found in the first example with "$-x$".

$$\mathbf{e}^{-x} = \sum_{n=0}^{\infty} \frac{(-x)^n}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n x^n}{n!}$$

This is a much shorter method of arriving at the same answer so don't forget about using previously computed series where possible (and allowed of course).

Example 3

Find the Taylor Series for $f(x) = x^4 \mathbf{e}^{-3x^2}$ about $x = 0$.

Solution

For this example, we will take advantage of the fact that we already have a Taylor Series for $\mathbf{e}^x$ about $x = 0$. In this example, unlike the previous example, doing this directly would be significantly longer and more difficult.

$$\begin{aligned} x^4 \mathbf{e}^{-3x^2} &= x^4 \sum_{n=0}^{\infty} \frac{\left(-3x^2\right)^n}{n!} \\ &= x^4 \sum_{n=0}^{\infty} \frac{(-3)^n x^{2n}}{n!} \\ &= \sum_{n=0}^{\infty} \frac{(-3)^n x^{2n+4}}{n!} \end{aligned}$$

To this point we've only looked at Taylor Series about $x = 0$ (also known as Maclaurin Series) so let's take a look at a Taylor Series that isn't about $x = 0$. Also, we'll pick on the exponential function one more time since it makes some of the work easier. This will be the final Taylor Series for exponentials in this section.

Example 4

Find the Taylor Series for $f(x) = \mathbf{e}^{-x}$ about $x = -4$.

Solution

Finding a general formula for $f^{(n)}(-4)$ is fairly simple.

$$f^{(n)}(x) = (-1)^n \mathbf{e}^{-x} \qquad\qquad f^{(n)}(-4) = (-1)^n \mathbf{e}^{4}$$

The Taylor Series is then,

$$\mathbf{e}^{-x} = \sum_{n=0}^{\infty} \frac{(-1)^n \mathbf{e}^4}{n!} (x+4)^n$$

Okay, we now need to work some examples that don't involve the exponential function since these will tend to require a little more work.

Example 5

Find the Taylor Series for $f(x) = \cos(x)$ about $x = 0$.

Solution

First, we'll need to take some derivatives of the function and evaluate them at $x = 0$.

$$\begin{aligned}
f^{(0)}(x) &= \cos(x) & f^{(0)}(0) &= 1 \\
f^{(1)}(x) &= -\sin(x) & f^{(1)}(0) &= 0 \\
f^{(2)}(x) &= -\cos(x) & f^{(2)}(0) &= -1 \\
f^{(3)}(x) &= \sin(x) & f^{(3)}(0) &= 0 \\
f^{(4)}(x) &= \cos(x) & f^{(4)}(0) &= 1 \\
f^{(5)}(x) &= -\sin(x) & f^{(5)}(0) &= 0 \\
f^{(6)}(x) &= -\cos(x) & f^{(6)}(0) &= -1 \\
&\vdots & &\vdots
\end{aligned}$$

In this example, unlike the previous ones, there is not an easy formula for either the general derivative or the evaluation of the derivative. However, there is a clear pattern to the evaluations. So, let's plug what we've got into the Taylor series and see what we get,

$$\begin{aligned}
\cos(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n \\
&= f(0) + f'(0)\,x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \frac{f^{(4)}(0)}{4!}x^4 + \frac{f^{(5)}(0)}{5!}x^5 + \cdots \\
&= \underbrace{1}_{n=0} + \underbrace{0}_{n=1} - \underbrace{\frac{1}{2!}}_{n=2}x^2 + \underbrace{0}_{n=3} + \underbrace{\frac{1}{4!}}_{n=4}x^4 + \underbrace{0}_{n=5} - \underbrace{\frac{1}{6!}}_{n=6}x^6 + \cdots
\end{aligned}$$

So, we only pick up terms with even powers on the x's. This doesn't really help us to get a general formula for the Taylor Series. However, let's drop the zeroes and "renumber" the terms as follows to see what we can get.

$$\cos(x) = \underbrace{1}_{n=0} - \underbrace{\frac{1}{2!}x^2}_{n=1} + \underbrace{\frac{1}{4!}x^4}_{n=2} - \underbrace{\frac{1}{6!}x^6}_{n=3} + \cdots$$

By renumbering the terms as we did we can actually come up with a general formula for the Taylor Series and here it is,

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$$

This idea of renumbering the series terms as we did in the previous example isn't used all that often, but occasionally is very useful. There is one more series where we need to do it so let's take a look at that so we can get one more example down of renumbering series terms.

Example 6

Find the Taylor Series for $f(x) = \sin(x)$ about $x = 0$.

Solution

As with the last example we'll start off in the same manner.

$$\begin{aligned}
f^{(0)}(x) &= \sin(x) & f^{(0)}(0) &= 0 \\
f^{(1)}(x) &= \cos(x) & f^{(1)}(0) &= 1 \\
f^{(2)}(x) &= -\sin(x) & f^{(2)}(0) &= 0 \\
f^{(3)}(x) &= -\cos(x) & f^{(3)}(0) &= -1 \\
f^{(4)}(x) &= \sin(x) & f^{(4)}(0) &= 0 \\
f^{(5)}(x) &= \cos(x) & f^{(5)}(0) &= 1 \\
f^{(6)}(x) &= -\sin(x) & f^{(6)}(0) &= 0 \\
&\vdots & &\vdots
\end{aligned}$$

So, we get a similar pattern for this one. Let's plug the numbers into the Taylor Series.

$$\begin{aligned}
\sin(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n \\
&= \frac{1}{1!}x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \frac{1}{7!}x^7 + \cdots
\end{aligned}$$

In this case we only get terms that have an odd exponent on x and as with the last problem once we ignore the zero terms there is a clear pattern and formula. So renumbering the terms as we did in the previous example we get the following Taylor Series.

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}$$

We really need to work another example or two in which $f(x)$ isn't about $x = 0$.

Example 7

Find the Taylor Series for $f(x) = \ln(x)$ about $x = 2$.

Solution

Here are the first few derivatives and the evaluations.

$$
\begin{aligned}
f^{(0)}(x) &= \ln(x) & f^{(0)}(2) &= \ln 2 \\
f^{(1)}(x) &= \frac{1}{x} & f^{(1)}(2) &= \frac{1}{2} \\
f^{(2)}(x) &= -\frac{1}{x^2} & f^{(2)}(2) &= -\frac{1}{2^2} \\
f^{(3)}(x) &= \frac{2}{x^3} & f^{(3)}(2) &= \frac{2}{2^3} \\
f^{(4)}(x) &= -\frac{2(3)}{x^4} & f^{(4)}(2) &= -\frac{2(3)}{2^4} \\
f^{(5)}(x) &= \frac{2(3)(4)}{x^5} & f^{(5)}(2) &= \frac{2(3)(4)}{2^5} \\
&\vdots & &\vdots \\
f^{(n)}(x) &= \frac{(-1)^{n+1}(n-1)!}{x^n} & f^{(n)}(2) &= \frac{(-1)^{n+1}(n-1)!}{2^n} \qquad n = 1, 2, 3, \ldots
\end{aligned}
$$

Note that while we got a general formula here it doesn't work for $n = 0$. This will happen on occasion so don't worry about it when it does.

In order to plug this into the Taylor Series formula we'll need to strip out the $n = 0$ term first.

$$
\begin{aligned}
\ln(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(2)}{n!}(x-2)^n \\
&= f(2) + \sum_{n=1}^{\infty} \frac{f^{(n)}(2)}{n!}(x-2)^n \\
&= \ln(2) + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}(n-1)!}{n!\ 2^n}(x-2)^n \\
&= \ln(2) + \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n\,2^n}(x-2)^n
\end{aligned}
$$

Notice that we simplified the factorials in this case. You should always simplify them if there are more than one and it's possible to simplify them.

Also, do not get excited about the term sitting in front of the series. Sometimes we need to do that when we can't get a general formula that will hold for all values of n.

Example 8

Find the Taylor Series for $f(x) = \dfrac{1}{x^2}$ about $x = -1$.

Solution

Again, here are the derivatives and evaluations.

$$
\begin{aligned}
f^{(0)}(x) &= \frac{1}{x^2} & f^{(0)}(-1) &= \frac{1}{(-1)^2} = 1 \\
f^{(1)}(x) &= -\frac{2}{x^3} & f^{(1)}(-1) &= -\frac{2}{(-1)^3} = 2 \\
f^{(2)}(x) &= \frac{2(3)}{x^4} & f^{(2)}(-1) &= \frac{2(3)}{(-1)^4} = 2(3) \\
f^{(3)}(x) &= -\frac{2(3)(4)}{x^5} & f^{(3)}(-1) &= -\frac{2(3)(4)}{(-1)^5} = 2(3)(4) \\
&\vdots & &\vdots \\
f^{(n)}(x) &= \frac{(-1)^n (n+1)!}{x^{n+2}} & f^{(n)}(-1) &= \frac{(-1)^n (n+1)!}{(-1)^{n+2}} = (n+1)!
\end{aligned}
$$

Notice that all the negative signs will cancel out in the evaluation. Also, this formula will work for all n, unlike the previous example.

Here is the Taylor Series for this function.

$$
\begin{aligned}
\frac{1}{x^2} &= \sum_{n=0}^{\infty} \frac{f^{(n)}(-1)}{n!}(x+1)^n \\
&= \sum_{n=0}^{\infty} \frac{(n+1)!}{n!}(x+1)^n \\
&= \sum_{n=0}^{\infty} (n+1)(x+1)^n
\end{aligned}
$$

Now, let's work one of the easier examples in this section. The problem for most students is that it may not appear to be that easy (or maybe it will appear to be too easy) at first glance.

Example 9

Find the Taylor Series for $f(x) = x^3 - 10x^2 + 6$ about $x = 3$.

Solution

Here are the derivatives for this problem.

$$\begin{aligned}
f^{(0)}(x) &= x^3 - 10x^2 + 6 & f^{(0)}(3) &= -57 \\
f^{(1)}(x) &= 3x^2 - 20x & f^{(1)}(3) &= -33 \\
f^{(2)}(x) &= 6x - 20 & f^{(2)}(3) &= -2 \\
f^{(3)}(x) &= 6 & f^{(3)}(3) &= 6 \\
&\vdots & &\vdots \\
f^{(n)}(x) &= 0 & f^{(4)}(3) &= 0 \qquad n \geq 4
\end{aligned}$$

This Taylor series will terminate after $n = 3$. This will always happen when we are finding the Taylor Series of a polynomial. Here is the Taylor Series for this one.

$$\begin{aligned}
x^3 - 10x^2 + 6 &= \sum_{n=0}^{\infty} \frac{f^{(n)}(3)}{n!}(x-3)^n \\
&= f(3) + f'(3)(x-3) + \frac{f''(3)}{2!}(x-3)^2 + \frac{f'''(3)}{3!}(x-3)^3 + 0 \\
&= -57 - 33(x-3) - (x-3)^2 + (x-3)^3
\end{aligned}$$

When finding the Taylor Series of a polynomial we don't do any simplification of the right-hand side. We leave it like it is. In fact, if we were to multiply everything out we just get back to the original polynomial!

While it's not apparent that writing the Taylor Series for a polynomial is useful there are times where this needs to be done. The problem is that they are beyond the scope of this course and so aren't covered here. For example, there is one application to series in the field of Differential Equations where this needs to be done on occasion.

So, we've seen quite a few examples of Taylor Series to this point and in all of them we were able to find general formulas for the series. This won't always be the case. To see an example of one that doesn't have a general formula check out the last example in the next section.

Before leaving this section there are three important Taylor Series that we've derived in this section that we should summarize up in one place. In my class I will assume that you know these formulas from this point on.

Fact

$$\mathbf{e}^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$
$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}$$
$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!}$$

10.17 Applications of Series

Now, that we know how to represent function as power series we can now talk about at least a couple of applications of series.

There are in fact many applications of series, unfortunately most of them are beyond the scope of this course. One application of power series (with the occasional use of Taylor Series) is in the field of Ordinary Differential Equations when finding Series Solutions to Differential Equations. If you are interested in seeing how that works you can check out that chapter of my Differential Equations notes.

Another application of series arises in the study of Partial Differential Equations. One of the more commonly used methods in that subject makes use of Fourier Series.

Many of the applications of series, especially those in the differential equations fields, rely on the fact that functions can be represented as a series. In these applications it is very difficult, if not impossible, to find the function itself. However, there are methods of determining the series representation for the unknown function.

While the differential equations applications are beyond the scope of this course there are some applications from a Calculus setting that we can look at.

Example 1

Determine a Taylor Series about $x = 0$ for the following integral.

$$\int \frac{\sin(x)}{x}\, dx$$

Solution

To do this we will first need to find a Taylor Series about $x = 0$ for the integrand. This however isn't terribly difficult. We already have a Taylor Series for sine about $x = 0$ so we'll just use that as follows,

$$\frac{\sin(x)}{x} = \frac{1}{x}\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n+1)!}$$

We can now do the problem.

$$\begin{aligned}\int \frac{\sin(x)}{x}\, dx &= \int \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n+1)!}\, dx \\ &= C + \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)\ (2n+1)!}\end{aligned}$$

So, while we can't integrate this function in terms of known functions we can come up with a series representation for the integral.

This idea of deriving a series representation for a function instead of trying to find the function itself is used quite often in several fields. In fact, there are some fields where this is one of the main ideas used and without this idea it would be very difficult to accomplish anything in those fields.

Another application of series isn't really an application of infinite series. It's more an application of partial sums. In fact, we've already seen this application in use once in this chapter. In the Estimating the Value of a Series we used a partial sum to estimate the value of a series. We can do the same thing with power series and series representations of functions. The main difference is that we will now be using the partial sum to approximate a function instead of a single value.

We will look at Taylor series for our examples, but we could just as easily use any series representation here. Recall that the **nth degree Taylor Polynomial of** $f(x)$ is given by,

$$T_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i$$

Let's take a look at example of this.

Example 2

For the function $f(x) = \cos(x)$ plot the function as well as $T_2(x)$, $T_4(x)$, and $T_8(x)$ on the same graph for the interval $[-4, 4]$.

Solution

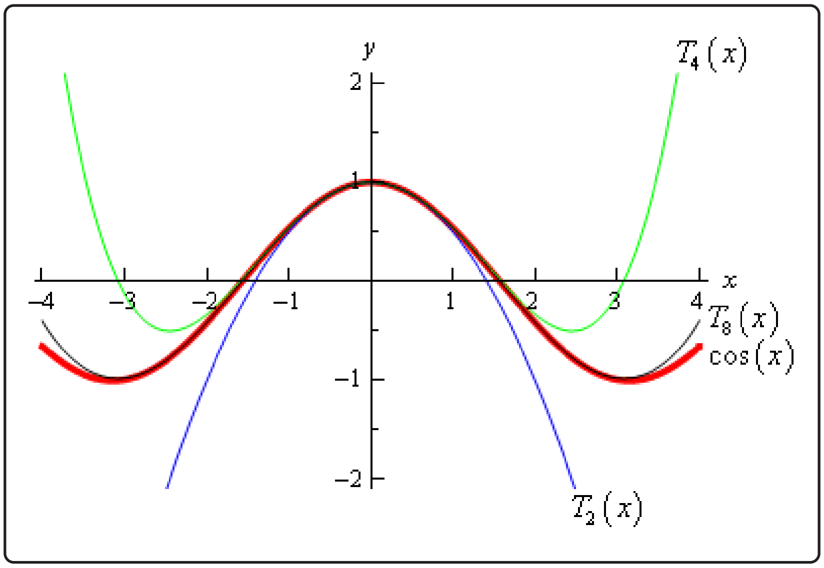
Here is the general formula for the Taylor polynomials for cosine.

$$T_{2n}(x) = \sum_{i=0}^{n} \frac{(-1)^i x^{2i}}{(2i)!}$$

The three Taylor polynomials that we've got are then,

$$\begin{aligned} T_2(x) &= 1 - \frac{x^2}{2} \\ T_4(x) &= 1 - \frac{x^2}{2} + \frac{x^4}{24} \\ T_8(x) &= 1 - \frac{x^2}{2} + \frac{x^4}{24} - \frac{x^6}{720} + \frac{x^8}{40320} \end{aligned}$$

Here is the graph of these three Taylor polynomials as well as the graph of cosine.



As we can see from this graph as we increase the degree of the Taylor polynomial it starts to look more and more like the function itself. In fact, by the time we get to $T_8(x)$ the only difference is right at the ends. The higher the degree of the Taylor polynomial the better it approximates the function.

Also, the larger the interval the higher degree Taylor polynomial we need to get a good approximation for the whole interval.

Before moving on let's write down a couple more Taylor polynomials from the previous example. Notice that because the Taylor series for cosine doesn't contain any terms with odd powers on x we get the following Taylor polynomials.

$$\begin{aligned} T_3(x) &= 1 - \frac{x^2}{2} \\ T_5(x) &= 1 - \frac{x^2}{2} + \frac{x^4}{24} \\ T_9(x) &= 1 - \frac{x^2}{2} + \frac{x^4}{24} - \frac{x^6}{720} + \frac{x^8}{40320} \end{aligned}$$

These are identical to those used in the example. Sometimes this will happen although that was not really the point of this. The point is to notice that the nth degree Taylor polynomial may actually have a degree that is less than n. It will never be more than n, but it can be less than n.

The final example in this section really isn't an application of series and probably belonged in the previous section. However, the previous section was getting too long so the example is in this section. This is an example of how to multiply series together and while this isn't an application of series it is something that does have to be done on occasion in the applications. So, in that sense it does belong in this section.

Example 3

Find the first three non-zero terms in the Taylor Series for $f(x) = \mathbf{e}^x \cos(x)$ about $x = 0$.

Solution

Before we start let's acknowledge that the easiest way to do this problem is to simply compute the first 3-4 derivatives, evaluate them at $x = 0$, plug into the formula and we'd be done. However, as we noted prior to this example we want to use this example to illustrate how we multiply series.

We will make use of the fact that we've got Taylor Series for each of these so we can use them in this problem.

$$\mathbf{e}^x \cos(x) = \left(\sum_{n=0}^{\infty} \frac{x^n}{n!}\right)\left(\sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!}\right)$$

We're not going to completely multiply out these series. We're going to do enough of the multiplication to get an answer. The problem statement says that we want the first three *non-zero* terms. That non-zero bit is important as it is possible that some of the terms will be zero. If none of the terms are zero this would mean that the first three non-zero terms would be the constant term, x term, and x^2 term. However, because some might be zero let's assume that if we get all the terms up through x^4 we'll have enough to get the answer. If we've assumed wrong it will be very easy to fix so don't worry about that.

Now, let's write down the first few terms of each series and we'll stop at the x^4 term in each.

$$\mathbf{e}^x \cos(x) = \left(1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots\right)\left(1 - \frac{x^2}{2} + \frac{x^4}{24} + \cdots\right)$$

Note that we do need to acknowledge that these series don't stop. That's the purpose of the "$+\cdots$" at the end of each. Just for a second however, let's suppose that each of these did stop and ask ourselves how we would multiply each out. If this were the case we would take every term in the second and multiply by every term in the first. In other words, we would first multiply every term in the second series by 1, then every term in the second series by x, then by x^2 *etc.*

By stopping each series at x^4 we have now guaranteed that we'll get all terms that have an exponent of 4 or less. Do you see why?

Each of the terms that we neglected to write down have an exponent of at least 5 and so multiplying by 1 or any power of x will result in a term with an exponent that is at a minimum

5. Therefore, none of the neglected terms will contribute terms with an exponent of 4 or less and so weren't needed.

So, let's start the multiplication process.

$$\begin{aligned}
\mathbf{e}^x \cos(x) &= \left(1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots\right)\left(1 - \frac{x^2}{2} + \frac{x^4}{24} + \cdots\right) \\
&= \underbrace{1 - \frac{x^2}{2} + \frac{x^4}{24} + \cdots}_{\text{Second Series} \times 1} + \underbrace{x - \frac{x^3}{2} + \frac{x^5}{24} + \cdots}_{\text{Second Series} \times x} + \underbrace{\frac{x^2}{2} - \frac{x^4}{4} + \frac{x^6}{48} + \cdots}_{\text{Second Series} \times {x^2}/{2}} \\
&\quad + \underbrace{\frac{x^3}{6} - \frac{x^5}{12} + \frac{x^7}{144} + \cdots}_{\text{Second Series} \times {x^3}/{6}} + \underbrace{\frac{x^4}{24} - \frac{x^6}{48} + \frac{x^8}{576} + \cdots}_{\text{Second Series} \times {x^4}/{24}} + \cdots
\end{aligned}$$

Now, collect like terms ignoring everything with an exponent of 5 or more since we won't have all those terms and don't want them either. Doing this gives,

$$\begin{aligned}
\mathbf{e}^x \cos(x) &= 1 + x + \left(-\frac{1}{2} + \frac{1}{2}\right)x^2 + \left(-\frac{1}{2} + \frac{1}{6}\right)x^3 + \left(\frac{1}{24} - \frac{1}{4} + \frac{1}{24}\right)x^4 + \cdots \\
&= 1 + x - \frac{x^3}{3} - \frac{x^4}{6} + \cdots
\end{aligned}$$

There we go. It looks like we over guessed and ended up with four non-zero terms, but that's okay. If we had under guessed and it turned out that we needed terms with x^5 in them all we would need to do at this point is go back and add in those terms to the original series and do a couple quick multiplications. In other words, there is no reason to completely redo all the work.

10.18 Binomial Series

In this final section of this chapter we are going to look at another series representation for a function. Before we do this let's first recall the following theorem.

Binomial Theorem

If n is any positive integer then,

$$\begin{aligned}(a+b)^n &= \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i \\ &= a^n + na^{n-1}b + \frac{n(n-1)}{2!}a^{n-2}b^2 + \cdots + nab^{n-1} + b^n\end{aligned}$$

where,

$$\binom{n}{i} = \frac{n(n-1)(n-2)\cdots(n-i+1)}{i!} \qquad i = 1, 2, 3, \ldots n$$
$$\binom{n}{0} = 1$$

This is useful for expanding $(a+b)^n$ for large n when straight forward multiplication wouldn't be easy to do. Let's take a quick look at an example.

Example 1

Use the Binomial Theorem to expand $(2x-3)^4$

Solution

There really isn't much to do other than plugging into the theorem.

$$\begin{aligned}(2x-3)^4 &= \sum_{i=0}^{4} \binom{4}{i} (2x)^{4-i} (-3)^i \\ &= \binom{4}{0}(2x)^4 + \binom{4}{1}(2x)^3(-3) + \binom{4}{2}(2x)^2(-3)^2 + \binom{4}{3}(2x)(-3)^3 + \binom{4}{4}(-3)^4 \\ &= (2x)^4 + 4(2x)^3(-3) + \frac{4(3)}{2}(2x)^2(-3)^2 + 4(2x)(-3)^3 + (-3)^4 \\ &= 16x^4 - 96x^3 + 216x^2 - 216x + 81\end{aligned}$$

Now, the Binomial Theorem required that n be a positive integer. There is an extension to this however that allows for any number at all.

Binomial Series

If k is any number and $|x| < 1$ then,

$$\begin{aligned}(1+x)^k &= \sum_{n=0}^{\infty} \binom{k}{n} x^n \\ &= 1 + kx + \frac{k(k-1)}{2!}x^2 + \frac{k(k-1)(k-2)}{3!}x^3 + \cdots\end{aligned}$$

where,

$$\begin{aligned}\binom{k}{n} &= \frac{k(k-1)(k-2)\cdots(k-n+1)}{n!} \qquad n = 1,2,3,\ldots \\ \binom{k}{0} &= 1\end{aligned}$$

So, similar to the binomial theorem except that it's an infinite series and we must have $|x| < 1$ in order to get convergence.

Let's check out an example of this.

Example 2

Write down the first four terms in the binomial series for $\sqrt{9-x}$

Solution

So, in this case $k = \frac{1}{2}$ and we'll need to rewrite the term a little to put it into the form required.

$$\sqrt{9-x} = 3\left(1 - \frac{x}{9}\right)^{\frac{1}{2}} = 3\left(1 + \left(-\frac{x}{9}\right)\right)^{\frac{1}{2}}$$

The first four terms in the binomial series is then,

$$\begin{aligned}\sqrt{9-x} &= 3\left(1 + \left(-\frac{x}{9}\right)\right)^{\frac{1}{2}} \\ &= 3\sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n} \left(-\frac{x}{9}\right)^n \\ &= 3\left[1 + \left(\frac{1}{2}\right)\left(-\frac{x}{9}\right) + \frac{\frac{1}{2}\left(-\frac{1}{2}\right)}{2}\left(-\frac{x}{9}\right)^2 + \frac{\frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)}{6}\left(-\frac{x}{9}\right)^3 + \cdots\right] \\ &= 3 - \frac{x}{6} - \frac{x^2}{216} - \frac{x^3}{3888} - \cdots\end{aligned}$$

11 Vectors

Once again we are completely changing topics from the last chapter. We are going to do a (very) brief introduction to vectors. We'll look at basic notation and concepts involving vectors as well as arithmetic involving vectors. We'll also look at the dot product and cross product of vectors as well as a couple of quick applications of the dot and cross product.

Once we get into the multi-variable Calculus (*i.e.* the topics usually taught in Calculus III) we'll run into vectors on a semi regular basis and so we'll need to be familiar with them and the common notation, concepts and arithmetic involving vectors.

11.1 Basic Concepts

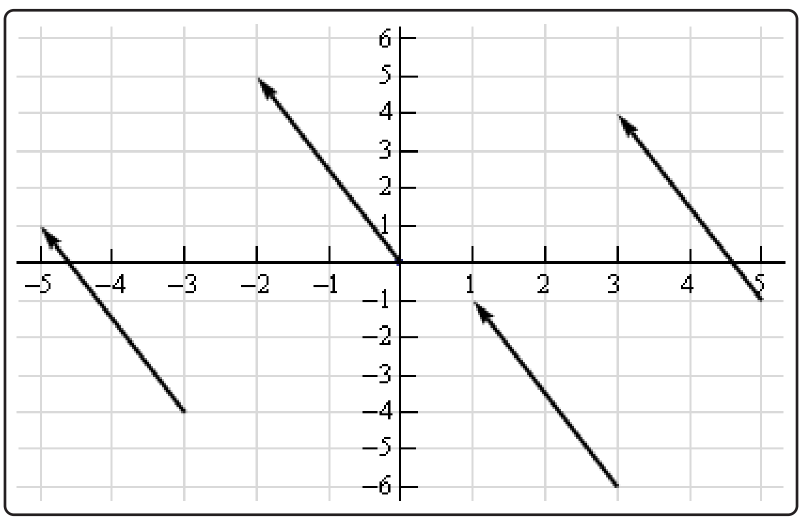
Let's start this section off with a quick discussion on what vectors are used for. Vectors are used to represent quantities that have both a magnitude and a direction. Good examples of quantities that can be represented by vectors are force and velocity. Both of these have a direction and a magnitude.

Let's consider force for a second. A force of say 5 Newtons that is applied in a particular direction can be applied at any point in space. In other words, the point where we apply the force does not change the force itself. Forces are independent of the point of application. To define a force all we need to know is the magnitude of the force and the direction that the force is applied in.

The same idea holds more generally with vectors. Vectors only impart magnitude and direction. They don't impart any information about where the quantity is applied. This is an important idea to always remember in the study of vectors.

In a graphical sense vectors are represented by directed line segments. The length of the line segment is the magnitude of the vector and the direction of the line segment is the direction of the vector. However, because vectors don't impart any information about where the quantity is applied any directed line segment with the same length and direction will represent the same vector.

Consider the sketch below.



Each of the directed line segments in the sketch represents the same vector. In each case the vector starts at a specific point then moves 2 units to the left and 5 units up. The notation that we'll use for this vector is,

$$\vec{v} = \langle -2, 5 \rangle$$

and each of the directed line segments in the sketch are called **representations** of the vector.

Be careful to distinguish vector notation, $\langle -2, 5 \rangle$, from the notation we use to represent coordinates of points, $(-2, 5)$. The vector denotes a magnitude and a direction of a quantity while the point denotes a location in space. So don't mix the notations up!

A representation of the vector $\vec{v} = \langle a_1, a_2 \rangle$ in two dimensional space is any directed line segment, $\overrightarrow{AB}$, from the point $A = (x, y)$ to the point $B = (x + a_1, y + a_2)$. Likewise a representation of the vector $\vec{v} = \langle a_1, a_2, a_3 \rangle$ in three dimensional space is any directed line segment, $\overrightarrow{AB}$, from the point $A = (x, y, z)$ to the point $B = (x + a_1, y + a_2, z + a_3)$.

Note that there is very little difference between the two dimensional and three dimensional formulas above. To get from the three dimensional formula to the two dimensional formula all we did is take out the third component/coordinate. Because of this most of the formulas here are given only in their three dimensional version. If we need them in their two dimensional form we can easily modify the three dimensional form.

There is one representation of a vector that is special in some way. The representation of the vector $\vec{v} = \langle a_1, a_2, a_3 \rangle$ that starts at the point $A = (0, 0, 0)$ and ends at the point $B = (a_1, a_2, a_3)$ is called the **position vector** of the point (a_1, a_2, a_3). So, when we talk about position vectors we are specifying the initial and final point of the vector.

Position vectors are useful if we ever need to represent a point as a vector. As we'll see there are times in which we definitely are going to want to represent points as vectors. In fact, we're going to run into topics that can only be done if we represent points as vectors.

Next, we need to discuss briefly how to generate a vector given the initial and final points of the representation. Given the two points $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$ the vector with the representation $\overrightarrow{AB}$ is,

$$\vec{v} = \langle b_1 - a_1, b_2 - a_2, b_3 - a_3 \rangle$$

Note that we have to be very careful with direction here. The vector above is the vector that starts at A and ends at B. The vector that starts at B and ends at A, *i.e.* with representation $\overrightarrow{BA}$ is,

$$\vec{w} = \langle a_1 - b_1, a_2 - b_2, a_3 - b_3 \rangle$$

These two vectors are different and so we do need to always pay attention to what point is the starting point and what point is the ending point. When determining the vector between two points we always subtract the initial point from the terminal point.

Example 1

Give the vector for each of the following.

(a) The vector from $(2, -7, 0)$ to $(1, -3, -5)$.

(b) The vector from $(1, -3, -5)$ to $(2, -7, 0)$.

(c) The position vector for $(-90, 4)$

Solution

(a) The vector from $(2, -7, 0)$ to $(1, -3, -5)$.

Remember that to construct this vector we subtract coordinates of the starting point from the ending point.

$$\langle 1 - 2, -3 - (-7), -5 - 0 \rangle = \langle -1, 4, -5 \rangle$$

(b) The vector from $(1, -3, -5)$ to $(2, -7, 0)$.

Same thing here.

$$\langle 2 - 1, -7 - (-3), 0 - (-5) \rangle = \langle 1, -4, 5 \rangle$$

Notice that the only difference between the first two is the signs are all opposite. This difference is important as it is this difference that tells us that the two vectors point in opposite directions.

(c) The position vector for $(-90, 4)$

Not much to this one other than acknowledging that the position vector of a point is nothing more than a vector with the point's coordinates as its components.

$$\langle -90, 4 \rangle$$

We now need to start discussing some of the basic concepts that we will run into on occasion.

Magnitude

The **magnitude**, or **length**, of the vector $\vec{v} = \langle a_1, a_2, a_3 \rangle$ is given by,

$$\|\vec{v}\| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

Example 2

Determine the magnitude of each of the following vectors.

(a) $\vec{a} = \langle 3, -5, 10 \rangle$

(b) $\vec{u} = \left\langle \frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}} \right\rangle$

(c) $\vec{w} = \langle 0, 0 \rangle$

(d) $\vec{i} = \langle 1, 0, 0 \rangle$

Solution

There isn't too much to these other than plug into the formula.

(a) $\|\vec{a}\| = \sqrt{9 + 25 + 100} = \sqrt{134}$

(b) $\|\vec{u}\| = \sqrt{\frac{1}{5} + \frac{4}{5}} = \sqrt{1} = 1$

(c) $\|\vec{w}\| = \sqrt{0 + 0} = 0$

(d) $\left\|\vec{i}\right\| = \sqrt{1 + 0 + 0} = 1$

We also have the following fact about the magnitude.

$$\text{If } \|\vec{a}\| = 0 \text{ then } \vec{a} = \vec{0}$$

This should make sense. Because we square all the components the only way we can get zero out of the formula was for the components to be zero in the first place.

Unit Vector

Any vector with magnitude of 1, *i.e.* $\|\vec{u}\| = 1$, is called a **unit vector**.

Example 3

Which of the vectors from Example 2 are unit vectors?

Solution

Both the second and fourth vectors had a length of 1 and so they are the only unit vectors from the first example.

Zero Vector

The vector $\vec{w} = \langle 0, 0 \rangle$ that we saw in the first example is called a zero vector since its components are all zero. Zero vectors are often denoted by $\vec{0}$. Be careful to distinguish 0 (the number) from $\vec{0}$ (the vector). The number 0 denotes the origin in space, while the vector $\vec{0}$ denotes a vector that has no magnitude or direction.

Standard Basis Vectors

The fourth vector from the second example, $\vec{i} = \langle 1, 0, 0 \rangle$, is called a **standard basis vector**. In three dimensional space there are three standard basis vectors,

$$\vec{i} = \langle 1, 0, 0 \rangle \qquad \vec{j} = \langle 0, 1, 0 \rangle \qquad \vec{k} = \langle 0, 0, 1 \rangle$$

In two dimensional space there are two standard basis vectors,

$$\vec{i} = \langle 1, 0 \rangle \qquad \vec{j} = \langle 0, 1 \rangle$$

Note that standard basis vectors are also unit vectors.

Warning

We are pretty much done with this section however, before proceeding to the next section we should point out that vectors are not restricted to two dimensional or three dimensional space. Vectors can exist in general n-dimensional space. The general notation for a n-dimensional vector is,

$$\vec{v} = \langle a_1, a_2, a_3, \ldots, a_n \rangle$$

and each of the a_i's are called **components** of the vector.

Because we will be working almost exclusively with two and three dimensional vectors in this course most of the formulas will be given for the two and/or three dimensional cases. However, most of the concepts/formulas will work with general vectors and the formulas are easily (and naturally) modified for general n-dimensional vectors. Also, because it is easier to visualize things in two dimensions most of the figures related to vectors will be two dimensional figures.

So, we need to be careful to not get too locked into the two or three dimensional cases from our discussions in this chapter. We will be working in these dimensions either because it's easier to visualize the situation or because physical restrictions of the problems will enforce a dimension upon us.

11.2 Vector Arithmetic

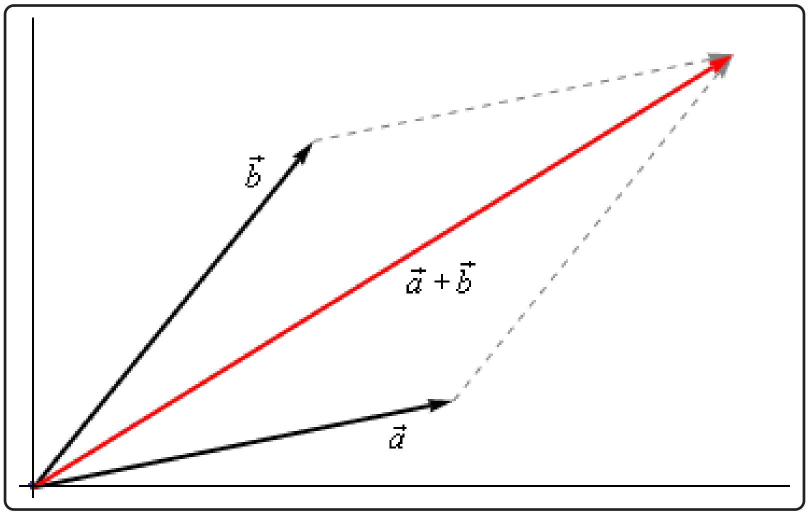
In this section we need to have a brief discussion of vector arithmetic.

We'll start with **addition** of two vectors. So, given the vectors $\vec{a} = \langle a_1, a_2, a_3 \rangle$ and $\vec{b} = \langle b_1, b_2, b_3 \rangle$ the addition of the two vectors is given by the following formula.

Vector Addition

$$\vec{a} + \vec{b} = \langle a_1 + b_1, a_2 + b_2, a_3 + b_3 \rangle$$

The following figure gives the geometric interpretation of the addition of two vectors.



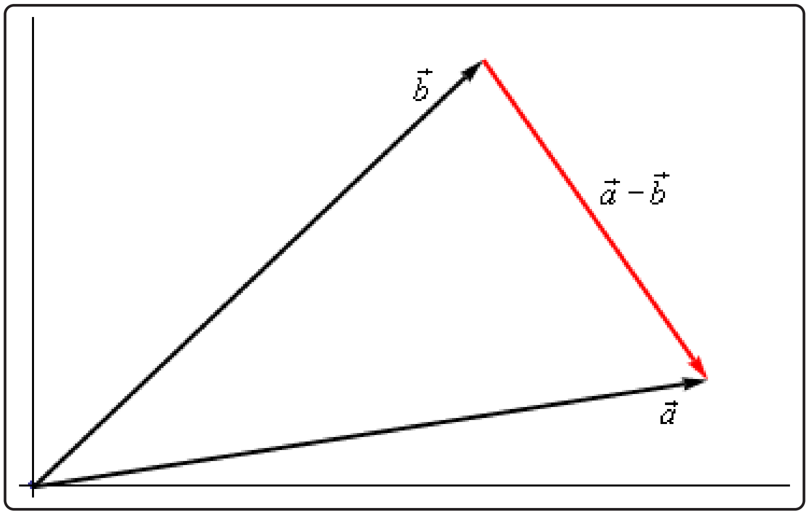
This is sometimes called the **parallelogram law** or **triangle law**.

Computationally, **subtraction** is very similar. Given the vectors $\vec{a} = \langle a_1, a_2, a_3 \rangle$ and $\vec{b} = \langle b_1, b_2, b_3 \rangle$ the difference of the two vectors is given by,

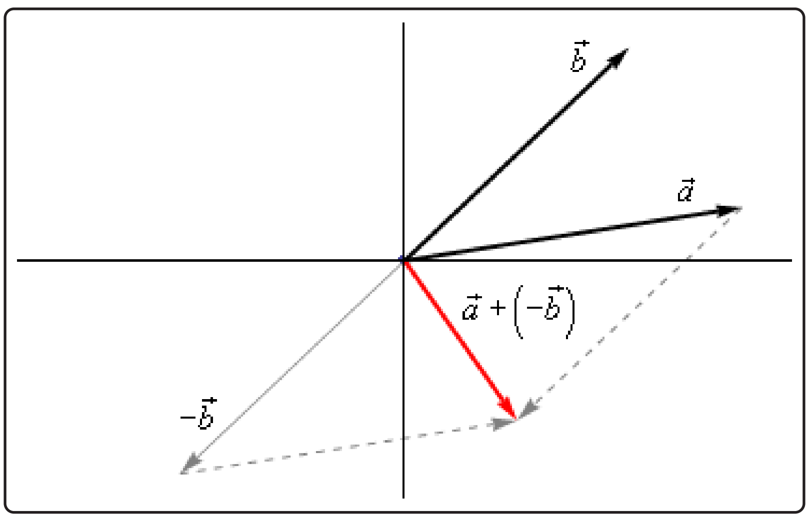
Vector Subtraction

$$\vec{a} - \vec{b} = \langle a_1 - b_1, a_2 - b_2, a_3 - b_3 \rangle$$

Here is the geometric interpretation of the difference of two vectors.



It is a little harder to see this geometric interpretation. To help see this let's instead think of subtraction as the addition of $\vec{a}$ and $-\vec{b}$. First, as we'll see in a bit $-\vec{b}$ is the same vector as $\vec{b}$ with opposite signs on all the components. In other words, $-\vec{b}$ goes in the opposite direction as $\vec{b}$. Here is the vector set up for $\vec{a} + \left(-\vec{b}\right)$.



As we can see from this figure we can move the vector representing $\vec{a} + \left(-\vec{b}\right)$ to the position we've got in the first figure showing the difference of the two vectors.

Note that we can't add or subtract two vectors unless they have the same number of components. If they don't have the same number of components then addition and subtraction can't be done.

The next arithmetic operation that we want to look at is **scalar multiplication**. Given the vector $\vec{a} = \langle a_1, a_2, a_3 \rangle$ and any number c the scalar multiplication is,

Scalar Multiplication

$$c\vec{a} = \langle ca_1, ca_2, ca_3 \rangle$$

So, we multiply all the components by the constant c. To see the geometric interpretation of scalar multiplication let's take a look at an example.

Example 1

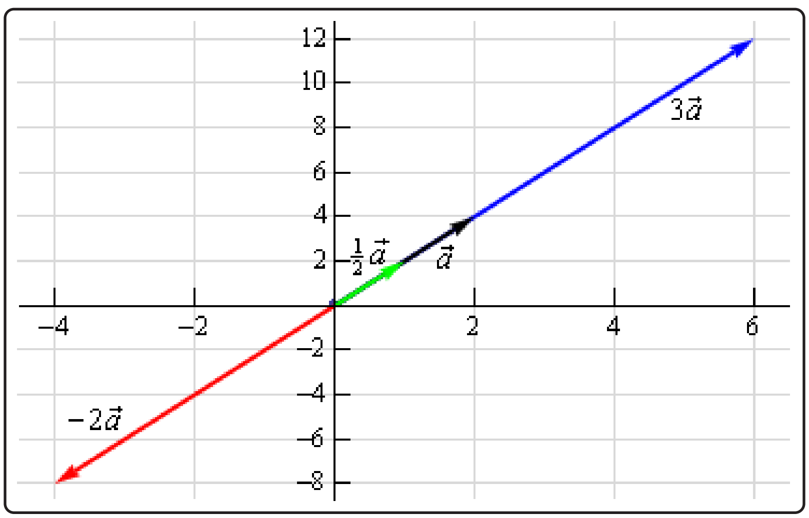
For the vector $\vec{a} = \langle 2, 4 \rangle$ compute $3\vec{a}$, $\frac{1}{2}\vec{a}$ and $-2\vec{a}$. Graph all four vectors on the same axis system.

Solution

Here are the three scalar multiplications.

$$3\vec{a} = \langle 6, 12 \rangle \qquad \frac{1}{2}\vec{a} = \langle 1, 2 \rangle \qquad -2\vec{a} = \langle -4, -8 \rangle$$

Here is the graph for each of these vectors.



In the previous example we can see that if c is positive all scalar multiplication will do is stretch (if $c > 1$) or shrink (if $c < 1$) the original vector, but it won't change the direction. Likewise, if c is negative scalar multiplication will switch the direction so that the vector will point in exactly the opposite direction and it will again stretch or shrink the magnitude of the vector depending upon the size of c.

There are several nice applications of scalar multiplication that we should now take a look at.

The first is parallel vectors. This is a concept that we will see quite a bit over the next couple of sections. Two vectors are parallel if they have the same direction or are in exactly opposite directions. Now, recall again the geometric interpretation of scalar multiplication. When we performed scalar multiplication we generated new vectors that were parallel to the original vectors (and each other for that matter).

Parallel Vectors

So, let's suppose that $\vec{a}$ and $\vec{b}$ are parallel vectors. If they are parallel then there must be a number c so that,

$$\vec{a} = c\vec{b}$$

So, two vectors are parallel if one is a scalar multiple of the other.

Example 2

Determine if the sets of vectors are parallel or not.

(a) $\vec{a} = \langle 2, -4, 1 \rangle$, $\vec{b} = \langle -6, 12, -3 \rangle$

(b) $\vec{a} = \langle 4, 10 \rangle$, $\vec{b} = \langle 2, -9 \rangle$

Solution

(a) $\vec{a} = \langle 2, -4, 1 \rangle$, $\vec{b} = \langle -6, 12, -3 \rangle$

These two vectors are parallel since $\vec{b} = -3\vec{a}$

(b) $\vec{a} = \langle 4, 10 \rangle$, $\vec{b} = \langle 2, -9 \rangle$

These two vectors aren't parallel. This can be seen by noticing that $4\left(\frac{1}{2}\right) = 2$ and yet $10\left(\frac{1}{2}\right) = 5 \neq -9$. In other words, we can't make $\vec{a}$ be a scalar multiple of $\vec{b}$.

The next application is best seen in an example.

Example 3

Find a unit vector that points in the same direction as $\vec{w} = \langle -5, 2, 1 \rangle$.

Solution

Okay, what we're asking for is a new parallel vector (points in the same direction) that happens to be a unit vector. We can do this with a scalar multiplication since all scalar multiplication does is change the length of the original vector (along with possibly flipping the direction to the opposite direction).

Here's what we'll do. First let's determine the magnitude of $\vec{w}$.

$$\|\vec{w}\| = \sqrt{25 + 4 + 1} = \sqrt{30}$$

Now, let's form the following new vector,

$$\vec{u} = \frac{1}{\|\vec{w}\|}\vec{w} = \frac{1}{\sqrt{30}} \langle -5, 2, 1 \rangle = \left\langle -\frac{5}{\sqrt{30}}, \frac{2}{\sqrt{30}}, \frac{1}{\sqrt{30}} \right\rangle$$

The claim is that this is a unit vector. That's easy enough to check

$$\|\vec{u}\| = \sqrt{\frac{25}{30} + \frac{4}{30} + \frac{1}{30}} = \sqrt{\frac{30}{30}} = 1$$

This vector also points in the same direction as $\vec{w}$ since it is only a scalar multiple of $\vec{w}$ and we used a positive multiple.

So, in general, given a vector $\vec{w}$, $\vec{u} = \frac{\vec{w}}{\|\vec{w}\|}$ will be a unit vector that points in the same direction as $\vec{w}$.

Standard Basis Vectors Revisited

In the previous section we introduced the idea of standard basis vectors without really discussing why they were important. We can now do that. Let's start with the vector

$$\vec{a} = \langle a_1, a_2, a_3 \rangle$$

We can use the addition of vectors to break this up as follows,

$$\begin{aligned} \vec{a} &= \langle a_1, a_2, a_3 \rangle \\ &= \langle a_1, 0, 0 \rangle + \langle 0, a_2, 0 \rangle + \langle 0, 0, a_3 \rangle \end{aligned}$$

Using scalar multiplication we can further rewrite the vector as,

$$\begin{aligned} \vec{a} &= \langle a_1, 0, 0 \rangle + \langle 0, a_2, 0 \rangle + \langle 0, 0, a_3 \rangle \\ &= a_1 \langle 1, 0, 0 \rangle + a_2 \langle 0, 1, 0 \rangle + a_3 \langle 0, 0, 1 \rangle \end{aligned}$$

Finally, notice that these three new vectors are simply the three standard basis vectors for three dimensional space.

$$\langle a_1, a_2, a_3 \rangle = a_1\vec{i} + a_2\vec{j} + a_3\vec{k}$$

So, we can take any vector and write it in terms of the standard basis vectors. From this point on we will use the two notations interchangeably so make sure that you can deal with both notations.

Example 4

If $\vec{a} = \langle 3, -9, 1 \rangle$ and $\vec{w} = -\vec{i} + 8\vec{k}$ compute $2\vec{a} - 3\vec{w}$.

Solution

In order to do the problem we'll convert to one notation and then perform the indicated operations.

$$\begin{aligned} 2\vec{a} - 3\vec{w} &= 2\langle 3, -9, 1 \rangle - 3\langle -1, 0, 8 \rangle \\ &= \langle 6, -18, 2 \rangle - \langle -3, 0, 24 \rangle \\ &= \langle 9, -18, -22 \rangle \end{aligned}$$

We will leave this section with some basic properties of vector arithmetic.

Properties

If $\vec{v}$, $\vec{w}$ and $\vec{u}$ are vectors (each with the same number of components) and a and b are two numbers then we have the following properties.

$$\vec{v} + \vec{w} = \vec{w} + \vec{v} \qquad \vec{u} + (\vec{v} + \vec{w}) = (\vec{u} + \vec{v}) + \vec{w}$$

$$\vec{v} + \vec{0} = \vec{v} \qquad 1\vec{v} = \vec{v}$$

$$a(\vec{v} + \vec{w}) = a\vec{v} + a\vec{w} \qquad (a + b)\vec{v} = a\vec{v} + b\vec{v}$$

The proofs of these are pretty much just "computation" proofs so we'll prove one of them and leave the others to you to prove.

Proof of $a\left(\vec{v}+\vec{w}\right)=a\vec{v}+a\vec{w}$

We'll start with the two vectors, $\vec{v}=\langle v_1,v_2,\ldots,v_n\rangle$ and $\vec{w}=\langle w_1,w_2,\ldots,w_n\rangle$ and yes we did mean for these to each have n components. The theorem works for general vectors so we may as well do the proof for general vectors.

Now, as noted above this is pretty much just a "computational" proof. What that means is that we'll compute the left side and then do some basic arithmetic on the result to show that we can make the left side look like the right side. Here is the work.

$$\begin{aligned}
a\left(\vec{v}+\vec{w}\right) &= a\left(\langle v_1,v_2,\ldots,v_n\rangle+\langle w_1,w_2,\ldots,w_n\rangle\right)\\
&= a\left\langle v_1+w_1,v_2+w_2,\ldots,v_n+w_n\right\rangle\\
&= \left\langle a\left(v_1+w_1\right),a\left(v_2+w_2\right),\ldots,a\left(v_n+w_n\right)\right\rangle\\
&= \langle av_1+aw_1,av_2+aw_2,\ldots,av_n+aw_n\rangle\\
&= \langle av_1,av_2,\ldots,av_n\rangle+\langle aw_1,aw_2,\ldots,aw_n\rangle\\
&= a\langle v_1,v_2,\ldots,v_n\rangle+a\langle w_1,w_2,\ldots,w_n\rangle=a\vec{v}+a\vec{w}
\end{aligned}$$

11.3 Dot Product

The next topic for discussion is that of the dot product. Let's jump right into the definition of the dot product. Given the two vectors $\vec{a} = \langle a_1, a_2, a_3 \rangle$ and $\vec{b} = \langle b_1, b_2, b_3 \rangle$ the dot product is,

Dot Product

$$\vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2 + a_3 b_3 \tag{11.1}$$

Sometimes the dot product is called the **scalar product**. The dot product is also an example of an inner product and so on occasion you may hear it called an inner product.

Example 1

Compute the dot product for each of the following.

(a) $\vec{v} = 5\vec{i} - 8\vec{j}, \ \ \vec{w} = \vec{i} + 2\vec{j}$

(b) $\vec{a} = \langle 0, 3, -7 \rangle, \ \vec{b} = \langle 2, 3, 1 \rangle$

Solution

Not much to do with these other than use the formula.

(a) $\vec{v} \cdot \vec{w} = 5 - 16 = -11$

(b) $\vec{a} \cdot \vec{b} = 0 + 9 - 7 = 2$

Here are some properties of the dot product.

Properties

$$\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w} \qquad (c\vec{v}) \cdot \vec{w} = \vec{v} \cdot (c\vec{w}) = c(\vec{v} \cdot \vec{w})$$

$$\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v} \qquad \vec{v} \cdot \vec{0} = 0$$

$$\vec{v} \cdot \vec{v} = \|\vec{v}\|^2 \qquad \text{If } \vec{v} \cdot \vec{v} = 0 \text{ then } \vec{v} = \vec{0}$$

The proofs of these properties are mostly "computational" proofs and so we're only going to do a couple of them and leave the rest to you to prove.

Proof of $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$

We'll start with the three vectors, $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$, $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ and $\vec{w} = \langle w_1, w_2, \ldots, w_n \rangle$ and yes we did mean for these to each have n components. The theorem works for general vectors so we may as well do the proof for general vectors.

Now, as noted above this is pretty much just a "computational" proof. What that means is that we'll compute the left side and then do some basic arithmetic on the result to show that we can make the left side look like the right side. Here is the work.

$$\begin{aligned}
\vec{u} \cdot (\vec{v} + \vec{w}) &= \langle u_1, u_2, \ldots, u_n \rangle \cdot (\langle v_1, v_2, \ldots, v_n \rangle + \langle w_1, w_2, \ldots, w_n \rangle) \\
&= \langle u_1, u_2, \ldots, u_n \rangle \cdot \langle v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n \rangle \\
&= u_1 (v_1 + w_1) + u_2 (v_2 + w_2) + \ldots + u_n (v_n + w_n) \\
&= u_1 v_1 + u_1 w_1 + u_2 v_2 + u_2 w_2 + \ldots + u_n v_n + u_n w_n \\
&= (u_1 v_1 + u_2 v_2 + \ldots + u_n v_n) + (u_1 w_1 + u_2 w_2 + \ldots + u_n w_n) \\
&= \langle u_1, u_2, \ldots, u_n \rangle \cdot \langle v_1, v_2, \ldots, v_n \rangle + \langle u_1, u_2, \ldots, u_n \rangle \cdot \langle w_1, w_2, \ldots, w_n \rangle \\
&= \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}
\end{aligned}$$

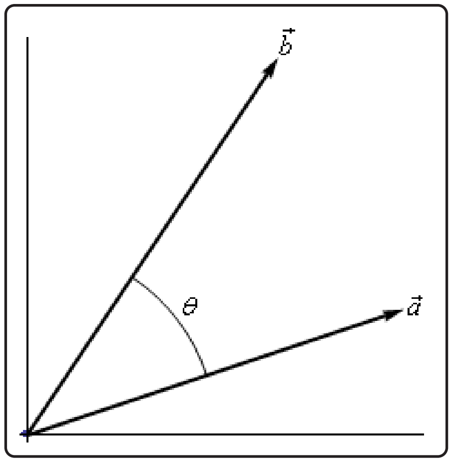
Proof of : If $\vec{v} \cdot \vec{v} = 0$ then $\vec{v} = \vec{0}$

This is a pretty simple proof. Let's start with $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ and compute the dot product.

$$\begin{aligned}
\vec{v} \cdot \vec{v} &= \langle v_1, v_2, \ldots, v_n \rangle \cdot \langle v_1, v_2, \ldots, v_n \rangle \\
&= v_1^2 + v_2^2 + \cdots + v_n^2 \\
&= 0
\end{aligned}$$

Now, since we know $v_i^2 \geq 0$ for all i then the only way for this sum to be zero is to in fact have $v_i^2 = 0$. This in turn however means that we must have $v_i = 0$ and so we must have had $\vec{v} = \vec{0}$.

There is also a nice geometric interpretation to the dot product. First suppose that θ is the angle between $\vec{a}$ and $\vec{b}$ such that $0 \leq \theta \leq \pi$ as shown in the image below.



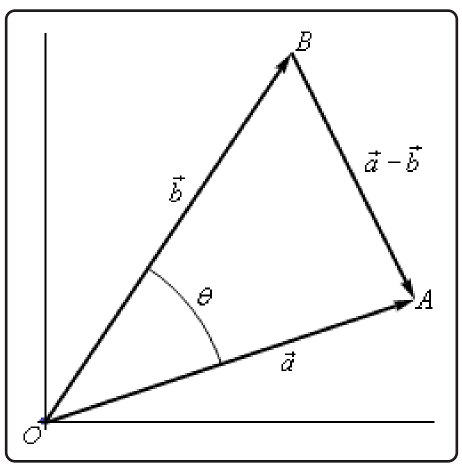
We can then have the following theorem.

Theorem

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \left\|\vec{b}\right\| \cos(\theta) \tag{11.2}$$

Proof

Let's give a modified version of the sketch above.



The three vectors above form the triangle AOB and note that the length of each side is nothing more than the magnitude of the vector forming that side.

The Law of Cosines tells us that,

$$\left\|\vec{a} - \vec{b}\right\|^2 = \|\vec{a}\|^2 + \left\|\vec{b}\right\|^2 - 2\|\vec{a}\| \left\|\vec{b}\right\| \cos(\theta)$$

Also using the properties of dot products we can write the left side as,

$$\begin{aligned}\left\|\vec{a}-\vec{b}\right\|^2 &= \left(\vec{a}-\vec{b}\right)\cdot\left(\vec{a}-\vec{b}\right)\\ &= \vec{a}\cdot\vec{a}-\vec{a}\cdot\vec{b}-\vec{b}\cdot\vec{a}+\vec{b}\cdot\vec{b}\\ &= \|\vec{a}\|^2-2\vec{a}\cdot\vec{b}+\left\|\vec{b}\right\|^2\end{aligned}$$

Our original equation is then,

$$\begin{aligned}\left\|\vec{a}-\vec{b}\right\|^2 &= \|\vec{a}\|^2+\left\|\vec{b}\right\|^2-2\,\|\vec{a}\|\,\left\|\vec{b}\right\|\cos(\theta)\\ \|\vec{a}\|^2-2\vec{a}\cdot\vec{b}+\left\|\vec{b}\right\|^2 &= \|\vec{a}\|^2+\left\|\vec{b}\right\|^2-2\,\|\vec{a}\|\,\left\|\vec{b}\right\|\cos(\theta)\\ -2\vec{a}\cdot\vec{b} &= -2\,\|\vec{a}\|\,\left\|\vec{b}\right\|\cos(\theta)\\ \vec{a}\cdot\vec{b} &= \|\vec{a}\|\,\left\|\vec{b}\right\|\cos(\theta)\end{aligned}$$

The formula from this theorem is often used not to compute a dot product but instead to find the angle between two vectors. Note as well that while the sketch of the two vectors in the proof is for two dimensional vectors the theorem is valid for vectors of any dimension (as long as they have the same dimension of course).

Let's see an example of this.

Example 2

Determine the angle between $\vec{a}=\langle 3,-4,-1\rangle$ and $\vec{b}=\langle 0,5,2\rangle$.

Solution

We will need the dot product as well as the magnitudes of each vector.

$$\vec{a}\cdot\vec{b}=-22 \qquad \|\vec{a}\|=\sqrt{26} \qquad \left\|\vec{b}\right\|=\sqrt{29}$$

The angle is then,

$$\cos(\theta)=\frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\,\left\|\vec{b}\right\|}=\frac{-22}{\sqrt{26}\sqrt{29}}=-0.8011927$$

$$\theta=\cos^{-1}\left(-0.8011927\right)=2.5\text{ radians }=143.24\text{ degrees}$$

The dot product gives us a very nice method for determining if two vectors are perpendicular and it will give another method for determining when two vectors are parallel. Note as well that often we will use the term **orthogonal** in place of perpendicular.

Orthogonal/Perpendicular Vectors

Now, if two vectors are orthogonal then we know that the angle between them is 90 degrees. From Equation 11.2 this tells us that if two vectors are orthogonal then,

$$\vec{a} \cdot \vec{b} = 0$$

Likewise, if two vectors are parallel then the angle between them is either 0 degrees (pointing in the same direction) or 180 degrees (pointing in the opposite direction). Once again using Equation 11.2 this would mean that one of the following would have to be true.

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \left\|\vec{b}\right\| \quad (\theta = 0^\circ) \qquad \text{OR} \qquad \vec{a} \cdot \vec{b} = -\|\vec{a}\| \left\|\vec{b}\right\| \quad (\theta = 180^\circ)$$

Example 3

Determine if the following vectors are parallel, orthogonal, or neither.

(a) $\vec{a} = \langle 6, -2, -1 \rangle\,,\ \vec{b} = \langle 2, 5, 2 \rangle$

(b) $\vec{u} = 2\vec{i} - \vec{j},\ \vec{v} = -\frac{1}{2}\vec{i} + \frac{1}{4}\vec{j}$

Solution

(a) $\vec{a} = \langle 6, -2, -1 \rangle\,,\ \vec{b} = \langle 2, 5, 2 \rangle$

First get the dot product to see if they are orthogonal.

$$\vec{a} \cdot \vec{b} = 12 - 10 - 2 = 0$$

The two vectors are orthogonal.

(b) $\vec{u} = 2\vec{i} - \vec{j},\ \vec{v} = -\frac{1}{2}\vec{i} + \frac{1}{4}\vec{j}$

Again, let's get the dot product first.

$$\vec{u} \cdot \vec{v} = -1 - \frac{1}{4} = -\frac{5}{4}$$

So, they aren't orthogonal. Let's get the magnitudes and see if they are parallel.

$$\|\vec{u}\| = \sqrt{5} \qquad \|\vec{v}\| = \sqrt{\frac{5}{16}} = \frac{\sqrt{5}}{4}$$

Now, notice that,

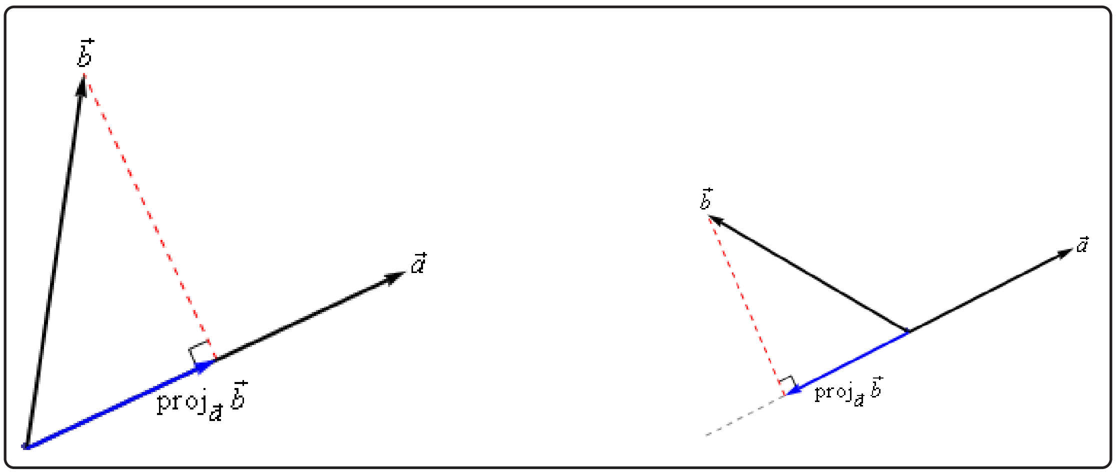
$$\vec{u}\cdot\vec{v} = -\frac{5}{4} = -\sqrt{5}\left(\frac{\sqrt{5}}{4}\right) = -\|\vec{u}\|\ \|\vec{v}\|$$

So, the two vectors are parallel.

There are several nice applications of the dot product as well that we should look at.

Projections

The best way to understand projections is to see a couple of sketches. So, given two vectors $\vec{a}$ and $\vec{b}$ we want to determine the projection of $\vec{b}$ onto $\vec{a}$. The projection is denoted by $\text{proj}_{\vec{a}}\vec{b}$. Here are a couple of sketches illustrating the projection.



So, to get the projection of $\vec{b}$ onto $\vec{a}$ we drop straight down from the end of $\vec{b}$ until we hit (and form a right angle) with the line that is parallel to $\vec{a}$. The projection is then the vector that is parallel to $\vec{a}$, starts at the same point both of the original vectors started at and ends where the dashed line hits the line parallel to $\vec{a}$.

There is a nice formula for finding the projection of $\vec{b}$ onto $\vec{a}$. Here it is,

Vector Projection of $\vec{b}$ onto $\vec{a}$

$$\text{proj}_{\vec{a}}\vec{b} = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|^2}\vec{a}$$

Note that we also need to be very careful with notation here. The projection of $\vec{a}$ onto $\vec{b}$ is given by

Vector Projection of $\vec{a}$ onto $\vec{b}$

$$\text{proj}_{\vec{b}}\vec{a} = \frac{\vec{a}\cdot\vec{b}}{\left\|\vec{b}\right\|^2}\vec{b}$$

We can see that this will be a totally different vector. This vector is parallel to $\vec{b}$, while $\text{proj}_{\vec{a}}\vec{b}$ is parallel to $\vec{a}$. So, be careful with notation and make sure you are finding the correct projection.

Here's an example.

Example 4

Determine the projection of $\vec{b} = \langle 2, 1, -1\rangle$ onto $\vec{a} = \langle 1, 0, -2\rangle$.

Solution

We need the dot product and the magnitude of $\vec{a}$.

$$\vec{a}\cdot\vec{b} = 4 \qquad \|\vec{a}\|^2 = 5$$

The projection is then,

$$\text{proj}_{\vec{a}}\vec{b} = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|^2}\vec{a} = \frac{4}{5}\langle 1, 0, -2\rangle = \left\langle \frac{4}{5}, 0, -\frac{8}{5}\right\rangle$$

For comparison purposes let's do it the other way around as well.

Example 5

Determine the projection of $\vec{a} = \langle 1, 0, -2\rangle$ onto $\vec{b} = \langle 2, 1, -1\rangle$.

Solution

We need the dot product and the magnitude of $\vec{b}$.

$$\vec{a}\cdot\vec{b} = 4 \qquad \left\|\vec{b}\right\|^2 = 6$$

The projection is then,

$$\text{proj}_{\vec{b}}\vec{a} = \frac{\vec{a}\cdot\vec{b}}{\left\|\vec{b}\right\|^2}\vec{b} = \frac{4}{6}\langle 2,1,-1\rangle = \left\langle \frac{4}{3}, \frac{2}{3}, -\frac{2}{3}\right\rangle$$

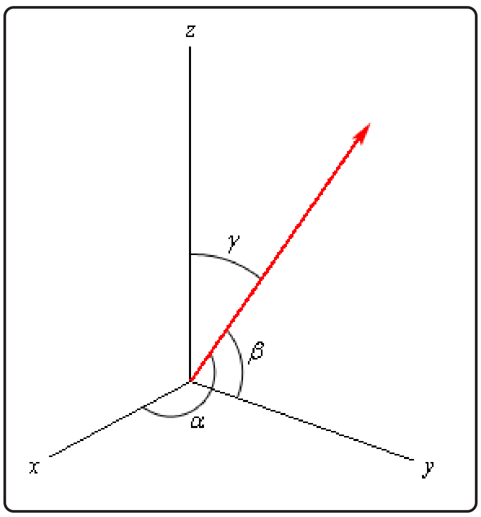
As we can see from the previous two examples the two projections are different so be careful.

Direction Cosines

This application of the dot product requires that we be in three dimensional space unlike all the other applications we've looked at to this point.

Let's start with a vector, $\vec{a}$, in three dimensional space. This vector will form angles with the x-axis (α), the y-axis (β), and the z-axis (γ). These angles are called **direction angles** and the cosines of these angles are called **direction cosines**.

Here is a sketch of a vector and the direction angles.



The formulas for the direction cosines are,

Direction Cosines

$$\cos(\alpha) = \frac{\vec{a}\cdot\vec{i}}{\|\vec{a}\|} = \frac{a_1}{\|\vec{a}\|} \qquad \cos(\beta) = \frac{\vec{a}\cdot\vec{j}}{\|\vec{a}\|} = \frac{a_2}{\|\vec{a}\|} \qquad \cos(\gamma) = \frac{\vec{a}\cdot\vec{k}}{\|\vec{a}\|} = \frac{a_3}{\|\vec{a}\|}$$

where $\vec{i}$, $\vec{j}$ and $\vec{k}$ are the standard basis vectors.

Let's verify the first dot product above. We'll leave the rest to you to verify.

$$\vec{a}\cdot\vec{i} = \langle a_1, a_2, a_3\rangle \cdot \langle 1,0,0\rangle = a_1$$

Here are a couple of nice facts about the direction cosines.

Facts

1. The vector $\vec{u} = \left\langle \cos(\alpha), \cos(\beta), \cos(\gamma) \right\rangle$ is a unit vector.
2. $\cos^2(\alpha) + \cos^2(\beta) + \cos^2(\gamma) = 1$
3. $\vec{a} = \|\vec{a}\| \left\langle \cos(\alpha), \cos(\beta), \cos(\gamma) \right\rangle$

Let's do a quick example involving direction cosines.

Example 6

Determine the direction cosines and direction angles for $\vec{a} = \langle 2, 1, -4 \rangle$.

Solution

We will need the magnitude of the vector.

$$\|\vec{a}\| = \sqrt{4 + 1 + 16} = \sqrt{21}$$

The direction cosines and angles are then,

$$\begin{aligned} \cos(\alpha) &= \frac{2}{\sqrt{21}} & \alpha &= 1.119 \text{ radians} = 64.123 \text{ degrees} \\ \cos(\beta) &= \frac{1}{\sqrt{21}} & \beta &= 1.351 \text{ radians} = 77.396 \text{ degrees} \\ \cos(\gamma) &= \frac{-4}{\sqrt{21}} & \gamma &= 2.632 \text{ radians} = 150.794 \text{ degrees} \end{aligned}$$

11.4 Cross Product

In this final section of this chapter we will look at the cross product of two vectors. We should note that the cross product requires both of the vectors to be three dimensional vectors.

Also, before getting into how to compute these we should point out a major difference between dot products and cross products. The result of a dot product is a number and the result of a cross product is a vector! Be careful not to confuse the two.

So, let's start with the two vectors $\vec{a} = \langle a_1, a_2, a_3 \rangle$ and $\vec{b} = \langle b_1, b_2, b_3 \rangle$ then the cross product is given by the formula,

Cross Product - Formula

$$\vec{a} \times \vec{b} = \langle a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1 \rangle$$

This is not an easy formula to remember. There are two ways to derive this formula. Both of them use the fact that the cross product is really the determinant of a 3x3 matrix. If you don't know what that is don't worry about it. You don't need to know anything about matrices or determinants to use either of the methods. The notation for the determinant is as follows,

Cross Product - Determinant

$$\vec{a} \times \vec{b} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

The first row is the standard basis vectors and must appear in the order given here. The second row is the components of $\vec{a}$ and the third row is the components of $\vec{b}$. Now, let's take a look at the different methods for getting the formula.

The first method uses the Method of Cofactors. If you don't know the method of cofactors that is fine, the result is all that we need. Here is the formula.

Method of Cofactors

$$\vec{a} \times \vec{b} = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \vec{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \vec{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \vec{k}$$

where,

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

This formula is not as difficult to remember as it might at first appear to be. First, the terms alternate in sign and notice that the 2x2 is missing the column below the standard basis vector that multiplies it as well as the row of standard basis vectors.

The second method is slightly easier; however, many textbooks don't cover this method as it will only work on 3x3 determinants. This method says to take the determinant as listed above and then copy the first two columns onto the end as shown below.

$$\vec{a} \times \vec{b} = \left| \begin{array}{ccc} \vec{i} & \vec{j} & \vec{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{array} \right| \begin{array}{cc} \vec{i} & \vec{j} \\ a_1 & a_2 \\ b_1 & b_2 \end{array}$$

We now have three diagonals that move from left to right and three diagonals that move from right to left. We multiply along each diagonal and add those that move from left to right and subtract those that move from right to left.

This is best seen in an example. We'll also use this example to illustrate a fact about cross products.

Example 1

If $\vec{a} = \langle 2, 1, -1 \rangle$ and $\vec{b} = \langle -3, 4, 1 \rangle$ compute each of the following.

(a) $\vec{a} \times \vec{b}$

(b) $\vec{b} \times \vec{a}$

Solution

(a) $\vec{a} \times \vec{b}$

Here is the computation for this one.

$$\begin{aligned} \vec{a} \times \vec{b} &= \left| \begin{array}{ccc} \vec{i} & \vec{j} & \vec{k} \\ 2 & 1 & -1 \\ -3 & 4 & 1 \end{array} \right| \begin{array}{cc} \vec{i} & \vec{j} \\ 2 & 1 \\ -3 & 4 \end{array} \\ &= \vec{i}\,(1)\,(1) + \vec{j}\,(-1)\,(-3) + \vec{k}\,(2)\,(4) - \vec{j}\,(2)\,(1) - \vec{i}\,(-1)\,(4) - \vec{k}\,(1)\,(-3) \\ &= 5\vec{i} + \vec{j} + 11\vec{k} \end{aligned}$$

(b) $\vec{b} \times \vec{a}$

And here is the computation for this one.

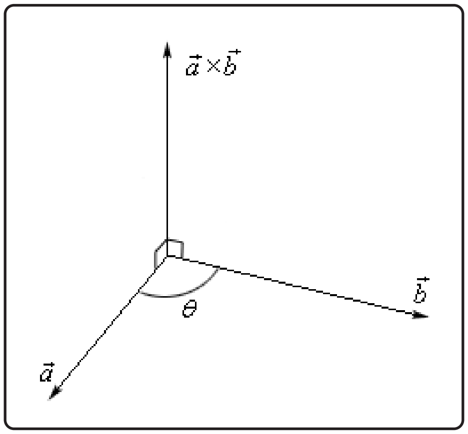
$$\begin{aligned} \vec{b} \times \vec{a} &= \left| \begin{array}{ccc} \vec{i} & \vec{j} & \vec{k} \\ -3 & 4 & 1 \\ 2 & 1 & -1 \end{array} \right| \begin{array}{cc} \vec{i} & \vec{j} \\ -3 & 4 \\ 2 & 1 \end{array} \\ &= \vec{i}\,(4)\,(-1) + \vec{j}\,(1)\,(2) + \vec{k}\,(-3)\,(1) - \vec{j}\,(-3)\,(-1) - \vec{i}\,(1)\,(1) - \vec{k}\,(4)\,(2) \\ &= -5\vec{i} - \vec{j} - 11\vec{k} \end{aligned}$$

Notice that switching the order of the vectors in the cross product simply changed all the signs in the result. Note as well that this means that the two cross products will point in exactly opposite directions since they only differ by a sign. We'll formalize up this fact shortly when we list several facts.

There is also a geometric interpretation of the cross product. First we will let θ be the angle between the two vectors $\vec{a}$ and $\vec{b}$ and assume that $0 \le \theta \le \pi$, then we have the following fact,

$$\left\|\vec{a} \times \vec{b}\right\| = \|\vec{a}\| \; \left\|\vec{b}\right\| \sin(\theta) \tag{11.3}$$

and the following figure.



There should be a natural question at this point. How did we know that the cross product pointed in the direction that we've given it here?

First, as this figure implies, the cross product is orthogonal to both of the original vectors. This will always be the case with one exception that we'll get to in a second.

Second, we knew that it pointed in the upward direction (in this case) by the "right hand rule". This says that if we take our right hand, start at $\vec{a}$ and rotate our fingers towards $\vec{b}$ our thumb will point in the direction of the cross product. Therefore, if we'd sketched in $\vec{b} \times \vec{a}$ above we would have gotten a vector in the downward direction.

Example 2

A plane is defined by any three points that are in the plane. If a plane contains the points $P = (1, 0, 0)$, $Q = (1, 1, 1)$ and $R = (2, -1, 3)$ find a vector that is orthogonal to the plane.

Solution

The one way that we know to get an orthogonal vector is to take a cross product. So, if we could find two vectors that we knew were in the plane and took the cross product of these two vectors we know that the cross product would be orthogonal to both the vectors. However, since both the vectors are in the plane the cross product would then also be orthogonal to the plane.

So, we need two vectors that are in the plane. This is where the points come into the problem. Since all three points lie in the plane any vector between them must also be in the plane. There are many ways to get two vectors between these points. We will use the following two,

$$\begin{aligned}\overrightarrow{PQ} &= \langle 1-1, 1-0, 1-0\rangle = \langle 0,1,1\rangle \\ \overrightarrow{PR} &= \langle 2-1, -1-0, 3-0\rangle = \langle 1,-1,3\rangle\end{aligned}$$

The cross product of these two vectors will be orthogonal to the plane. So, let's find the cross product.

$$\begin{aligned}\overrightarrow{PQ}\times\overrightarrow{PR} &= \left|\begin{array}{ccc|cc} \vec{i} & \vec{j} & \vec{k} & \vec{i} & \vec{j} \\ 0 & 1 & 1 & 0 & 1 \\ 1 & -1 & 3 & 1 & -1\end{array}\right| \\ &= 4\vec{i}+\vec{j}-\vec{k}\end{aligned}$$

So, the vector $4\vec{i}+\vec{j}-\vec{k}$ will be orthogonal to the plane containing the three points.

Now, let's address the one time where the cross product will not be orthogonal to the original vectors. If the two vectors, $\vec{a}$ and $\vec{b}$, are parallel then the angle between them is either 0 or 180 degrees. From Equation 11.3 this implies that,

$$\left\|\vec{a}\times\vec{b}\right\| = 0$$

From a fact about the magnitude we saw in the first section we know that this implies

$$\vec{a}\times\vec{b} = \vec{0}$$

In other words, it won't be orthogonal to the original vectors since we have the zero vector. This does give us another test for parallel vectors however.

Fact

If $\vec{a}\times\vec{b} = \vec{0}$ then $\vec{a}$ and $\vec{b}$ will be parallel vectors.

Let's also formalize up the fact about the cross product being orthogonal to the original vectors.

Fact

Provided $\vec{a} \times \vec{b} \neq \vec{0}$ then $\vec{a} \times \vec{b}$ is orthogonal to both $\vec{a}$ and $\vec{b}$.

Here are some nice properties about the cross product.

Properties

If $\vec{u}$, $\vec{v}$ and $\vec{w}$ are vectors and c is a number then,

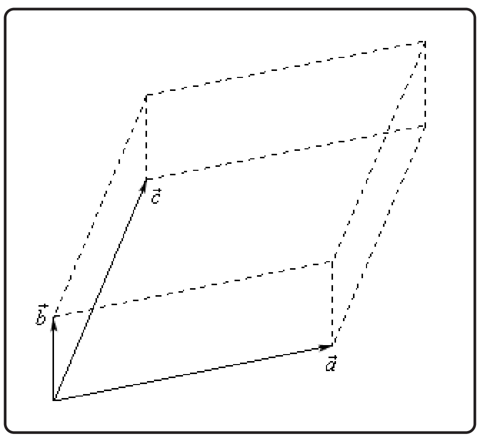
$$\vec{u} \times \vec{v} = -\vec{v} \times \vec{u} \qquad (c\vec{u}) \times \vec{v} = \vec{u} \times (c\vec{v}) = c(\vec{u} \times \vec{v})$$

$$\vec{u} \times (\vec{v} + \vec{w}) = \vec{u} \times \vec{v} + \vec{u} \times \vec{w} \qquad \vec{u} \cdot (\vec{v} \times \vec{w}) = (\vec{u} \times \vec{v}) \cdot \vec{w}$$

$$\vec{u} \cdot (\vec{v} \times \vec{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

The determinant in the last fact is computed in the same way that the cross product is computed. We will see an example of this computation shortly.

There are a couple of geometric applications to the cross product as well. Suppose we have three vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$ and we form the three dimensional figure shown below.



The area of the parallelogram (two dimensional front of this object) is given by,

$$\text{Area} = \left\| \vec{a} \times \vec{b} \right\|$$

and the volume of the parallelepiped (the whole three dimensional object) is given by,

$$\text{Volume} = \left| \vec{a} \cdot \left(\vec{b} \times \vec{c} \right) \right|$$

Note that the absolute value bars are required since the quantity could be negative and volume isn't negative.

We can use this volume fact to determine if three vectors lie in the same plane or not. If three vectors lie in the same plane then the volume of the parallelepiped will be zero.

Example 3

Determine if the three vectors $\vec{a} = \langle 1, 4, -7 \rangle$, $\vec{b} = \langle 2, -1, 4 \rangle$ and $\vec{c} = \langle 0, -9, 18 \rangle$ lie in the same plane or not.

Solution

So, as we noted prior to this example all we need to do is compute the volume of the parallelepiped formed by these three vectors. If the volume is zero they lie in the same plane and if the volume isn't zero they don't lie in the same plane.

$$\begin{aligned} \vec{a} \cdot \left(\vec{b} \times \vec{c} \right) &= \left| \begin{array}{ccc|cc} 1 & 4 & -7 & 1 & 4 \\ 2 & -1 & 4 & 2 & -1 \\ 0 & -9 & 18 & 0 & -9 \end{array} \right. \\ &= (1)(-1)(18) + (4)(4)(0) + (-7)(2)(-9) - \\ &\qquad\qquad (4)(2)(18) - (1)(4)(-9) - (-7)(-1)(0) \\ &= -18 + 126 - 144 + 36 \\ &= 0 \end{aligned}$$

So, the volume is zero and so they lie in the same plane.

12 Three Dimensional Space

In this chapter we will start looking at three dimensional space (3-D space or $\mathbb{R}^3$).As with the last chapter this is preparation for multi-variable Calculus (which we'll be starting in the next chapter) as the vast majority of the multi-variable Calculus material assumes we are in three dimensional (or higher dimensional) space.

In this chapter we will discuss the equations of lines and planes in three dimensional space as well as the equations of many of the standard quadric surfaces (*i.e* equations with at least one quadratic term in it).

We will define a vector function and discuss how to perform basic Calculus operations on vector functions. We will also discuss how to get tangent vectors (a vector tangent to a curve), normal vectors (a vector orthogonal/perpendicular) and the curvature of a curve from the vector function that defines the curve. We'll also have a quick discussion of how to get the velocity and acceleration of an object as it travels along a curve defined by a vector function.

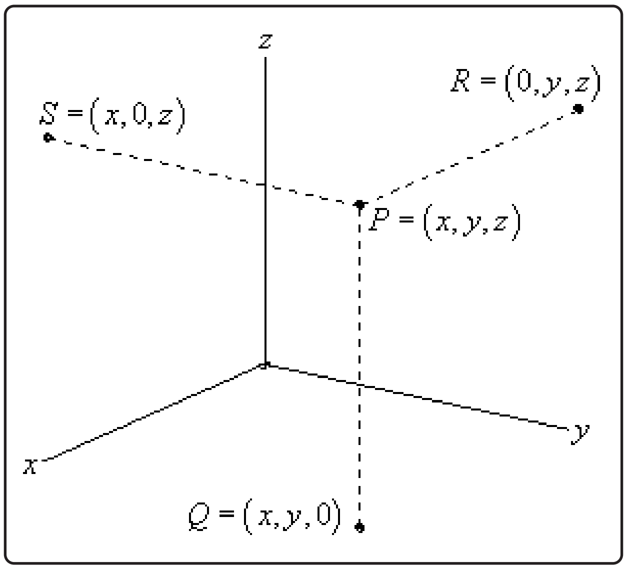
We will close out the chapter with a discussion a couple of alternative coordinates systems for three dimensional space, namely, cylindrical coordinates (a 3D extension of polar coordinates) and spherical coordinates.

12.1 The 3-D Coordinate System

We'll start the chapter off with a fairly short discussion introducing the 3-D coordinate system and the conventions that we'll be using. We will also take a brief look at how the different coordinate systems can change the graph of an equation.

Let's first get some basic notation out of the way. The 3-D coordinate system is often denoted by $\mathbb{R}^3$. Likewise, the 2-D coordinate system is often denoted by $\mathbb{R}^2$ and the 1-D coordinate system is denoted by $\mathbb{R}$. Also, as you might have guessed then a general n dimensional coordinate system is often denoted by $\mathbb{R}^n$.

Next, let's take a quick look at the basic coordinate system.



This is the standard placement of the axes in this class. It is assumed that only the positive directions are shown by the axes. If we need the negative axes for any reason we will put them in as needed.

Also note the various points on this sketch. The point P is the general point sitting out in 3-D space. If we start at P and drop straight down until we reach a z-coordinate of zero we arrive at the point Q. We say that Q sits in the xy-plane. The xy-plane corresponds to all the points which have a zero z-coordinate. We can also start at P and move in the other two directions as shown to get points in the xz-plane (this is S with a y-coordinate of zero) and the yz-plane (this is R with an x-coordinate of zero).

Collectively, the xy, xz, and yz-planes are sometimes called the coordinate planes. In the remainder of this class you will need to be able to deal with the various coordinate planes so make sure that you can.

Also, the point Q is often referred to as the projection of P in the xy-plane. Likewise, R is the projection of P in the yz-plane and S is the projection of P in the xz-plane.

Many of the formulas that you are used to working with in $\mathbb{R}^2$ have natural extensions in $\mathbb{R}^3$. For instance, the distance between two points in $\mathbb{R}^2$ is given by,

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

While the distance between any two points in $\mathbb{R}^3$ is given by,

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Likewise, the general equation for a circle with center (h, k) and radius r is given by,

$$(x - h)^2 + (y - k)^2 = r^2$$

and the general equation for a sphere with center (h, k, l) and radius r is given by,

$$(x - h)^2 + (y - k)^2 + (z - l)^2 = r^2$$

With that said we do need to be careful about just translating everything we know about $\mathbb{R}^2$ into $\mathbb{R}^3$ and assuming that it will work the same way. A good example of this is in graphing to some extent. Consider the following example.

Example 1

Graph $x = 3$ in $\mathbb{R}$, $\mathbb{R}^2$ and $\mathbb{R}^3$.

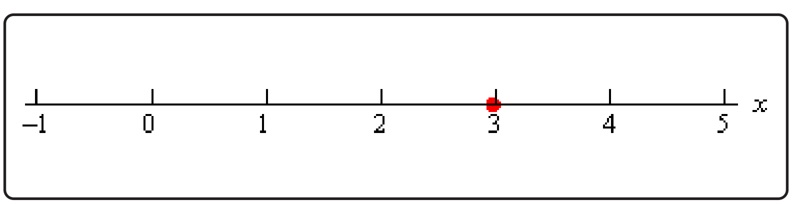
Solution

In $\mathbb{R}$ we have a single coordinate system and so $x = 3$ is a point in a 1-D coordinate system.

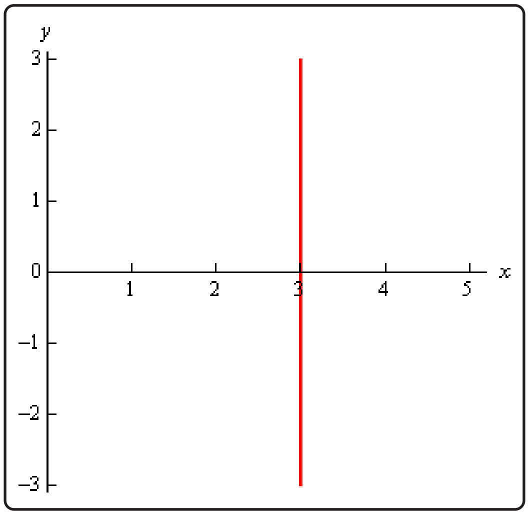
In $\mathbb{R}^2$ the equation $x = 3$ tells us to graph all the points that are in the form $(3, y)$. This is a vertical line in a 2-D coordinate system.

In $\mathbb{R}^3$ the equation $x = 3$ tells us to graph all the points that are in the form $(3, y, z)$. If you go back and look at the coordinate plane points this is very similar to the coordinates for the yz-plane except this time we have $x = 3$ instead of $x = 0$. So, in a 3-D coordinate system this is a plane that will be parallel to the yz-plane and pass through the x-axis at $x = 3$.

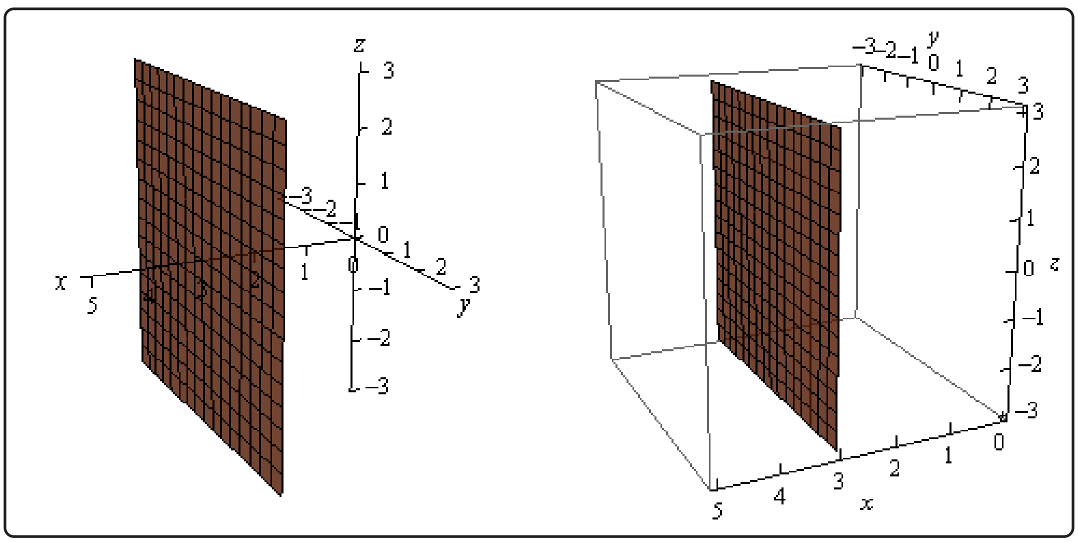
Here is the graph of $x = 3$ in $\mathbb{R}$.



Here is the graph of $x = 3$ in $\mathbb{R}^2$.



Finally, here is the graph of $x = 3$ in $\mathbb{R}^3$. Note that we've presented this graph in two different styles. On the left we've got the traditional axis system that we're used to seeing and on the right we've put the graph in a box. Both views can be convenient on occasion to help with perspective and so we'll often do this with 3D graphs and sketches.



Note that at this point we can now write down the equations for each of the coordinate planes as well using this idea.

$$z = 0 \qquad xy - \text{plane}$$
$$y = 0 \qquad xz - \text{plane}$$
$$x = 0 \qquad yz - \text{plane}$$

Let's take a look at a slightly more general example.

Example 2

Graph $y = 2x - 3$ in $\mathbb{R}^2$ and $\mathbb{R}^3$.

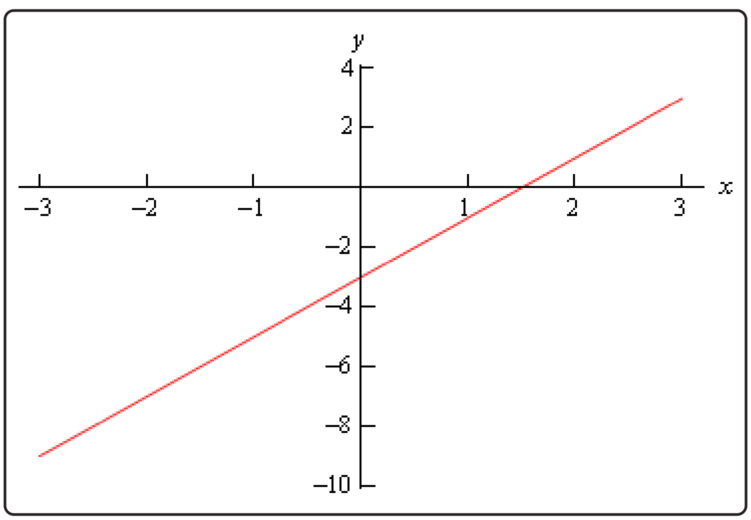
Solution

Note we had to throw out $\mathbb{R}$ for this example since there are two variables which means that we can't be in a 1-D space (1-D space has only one variable!).

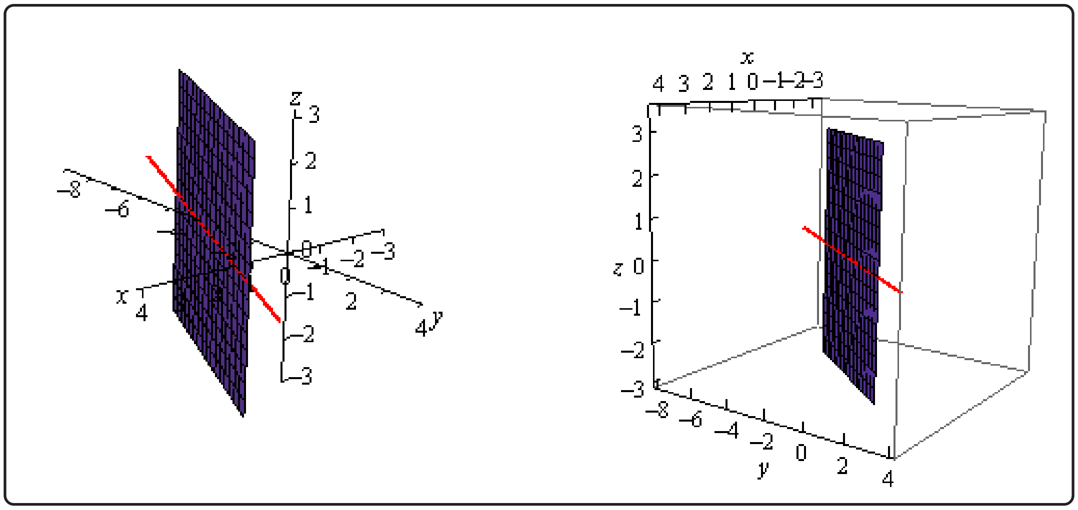
In $\mathbb{R}^2$ this is a line with slope 2 and a y intercept of -3.

However, in $\mathbb{R}^3$ this is not necessarily a line. Because we have not specified a value of z we are forced to let z take any value. This means that at any particular value of z we will get a copy of this line. So, the graph is then a vertical plane that lies over the line given by $y = 2x - 3$ in the xy-plane.

Here is the graph in $\mathbb{R}^2$.



here is the graph in $\mathbb{R}^3$.



Notice that if we look to where the plane intersects the xy-plane we will get the graph of the line in $\mathbb{R}^2$ as noted in the above graph by the red line through the plane.

Let's take a look at one more example of the difference between graphs in the different coordinate systems.

Example 3

Graph $x^2 + y^2 = 4$ in $\mathbb{R}^2$ and $\mathbb{R}^3$.

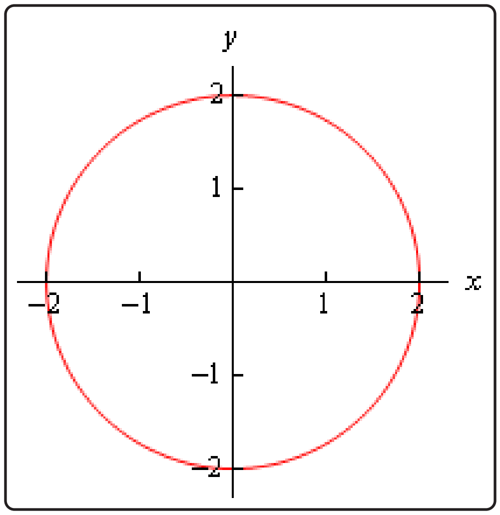
Solution

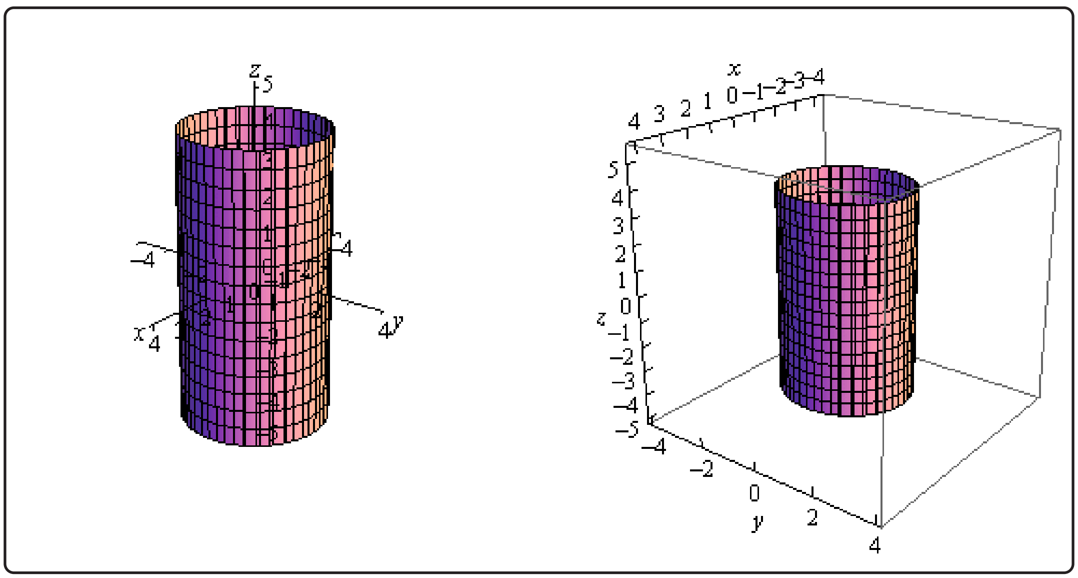
As with the previous example this won't have a 1-D graph since there are two variables.

In $\mathbb{R}^2$ this is a circle centered at the origin with radius 2.

In $\mathbb{R}^3$ however, as with the previous example, this may or may not be a circle. Since we have not specified z in any way we must assume that z can take on any value. In other words, at any value of z this equation must be satisfied and so at any value z we have a circle of radius 2 centered on the z-axis. This means that we have a cylinder of radius 2 centered on the z-axis.

Here are the graphs for this example.





Notice that again, if we look to where the cylinder intersects the xy-plane we will again get the circle from $\mathbb{R}^2$.

We need to be careful with the last two examples. It would be tempting to take the results of these and say that we can't graph lines or circles in $\mathbb{R}^3$ and yet that doesn't really make sense. There is no reason for there to not be graphs of lines or circles in $\mathbb{R}^3$. Let's think about the example of the circle. To graph a circle in $\mathbb{R}^3$ we would need to do something like $x^2 + y^2 = 4$ at $z = 5$. This would be a circle of radius 2 centered on the z-axis at the level of $z = 5$. So, as long as we specify a z we will get a circle and not a cylinder. We will see an easier way to specify circles in a later section.

We could do the same thing with the line from the second example. However, we will be looking at lines in more generality in the next section and so we'll see a better way to deal with lines in $\mathbb{R}^3$ there.

The point of the examples in this section is to make sure that we are being careful with graphing equations and making sure that we always remember which coordinate system that we are in.

Another quick point to make here is that, as we've seen in the above examples, many graphs of equations in $\mathbb{R}^3$ are surfaces. That doesn't mean that we can't graph curves in $\mathbb{R}^3$. We can and will graph curves in $\mathbb{R}^3$ as well as we'll see later in this chapter.

12.2 Equations of Lines

In this section we need to take a look at the equation of a line in $\mathbb{R}^3$. As we saw in the previous section the equation $y = mx + b$ does not describe a line in $\mathbb{R}^3$, instead it describes a plane. This doesn't mean however that we can't write down an equation for a line in 3-D space. We're just going to need a new way of writing down the equation of a curve.

So, before we get into the equations of lines we first need to briefly look at vector functions. We're going to take a more in depth look at vector functions later. At this point all that we need to worry about is notational issues and how they can be used to give the equation of a curve.

The best way to get an idea of what a vector function is and what its graph looks like is to look at an example. So, consider the following vector function.

$$\vec{r}(t) = \langle t, 1 \rangle$$

A vector function is a function that takes one or more variables, one in this case, and returns a vector. Note as well that a vector function can be a function of two or more variables. However, in those cases the graph may no longer be a curve in space.

The vector that the function gives can be a vector in whatever dimension we need it to be. In the example above it returns a vector in $\mathbb{R}^2$. When we get to the real subject of this section, equations of lines, we'll be using a vector function that returns a vector in $\mathbb{R}^3$

Now, we want to determine the graph of the vector function above. In order to find the graph of our function we'll think of the vector that the vector function returns as a position vector for points on the graph. Recall that a position vector, say $\vec{v} = \langle a, b \rangle$, is a vector that starts at the origin and ends at the point (a, b).

So, to get the graph of a vector function all we need to do is plug in some values of the variable and then plot the point that corresponds to each position vector we get out of the function and play connect the dots. Here are some evaluations for our example.

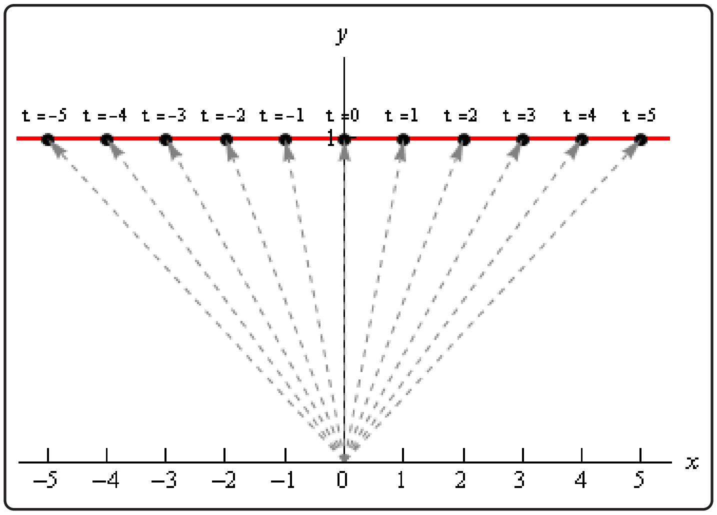
$$\vec{r}(-3) = \langle -3, 1 \rangle \qquad \vec{r}(-1) = \langle -1, 1 \rangle \qquad \vec{r}(2) = \langle 2, 1 \rangle \qquad \vec{r}(5) = \langle 5, 1 \rangle$$

So, each of these are position vectors representing points on the graph of our vector function. The points,

$$(-3, 1) \qquad (-1, 1) \qquad (2, 1) \qquad (5, 1)$$

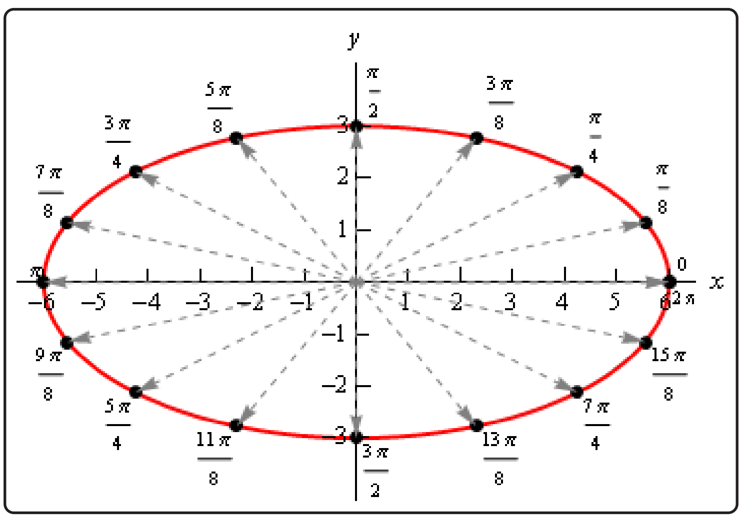
are all points that lie on the graph of our vector function.

If we do some more evaluations and plot all the points we get the following sketch.



In this sketch we've included the position vector (in gray and dashed) for several evaluations as well as the t (above each point) we used for each evaluation. It looks like, in this case the graph of the vector equation is in fact the line $y = 1$.

Here's another quick example. Here is the graph of $\vec{r}(t) = \langle 6\cos(t), 3\sin(t)\rangle$.



In this case we get an ellipse. It is important to not come away from this section with the idea that vector functions only graph out lines. We'll be looking at lines in this section, but the graphs of vector functions do not have to be lines as the example above shows.

We'll leave this brief discussion of vector functions with another way to think of the graph of a vector function. Imagine that a pencil/pen is attached to the end of the position vector and as we increase the variable the resulting position vector moves and as it moves the pencil/pen on the end sketches out the curve for the vector function.

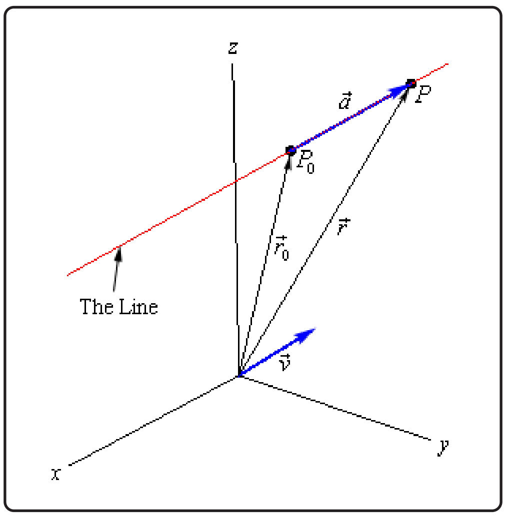
Okay, we now need to move into the actual topic of this section. We want to write down the equation of a line in $\mathbb{R}^3$ and as suggested by the work above we will need a vector function to do this. To see how we're going to do this let's think about what we need to write down the equation of a line in $\mathbb{R}^2$. In two dimensions we need the slope (m) and a point that was on the line in order to write down the equation.

In $\mathbb{R}^3$ that is still all that we need except in this case the "slope" won't be a simple number as it was in two dimensions. In this case we will need to acknowledge that a line can have a three dimensional slope. So, we need something that will allow us to describe a direction that is potentially in three dimensions. We already have a quantity that will do this for us. Vectors give directions and can be three dimensional objects.

So, let's start with the following information. Suppose that we know a point that is on the line, $P_0 = (x_0, y_0, z_0)$, and that $\vec{v} = \langle a, b, c\rangle$ is some vector that is parallel to the line. Note, in all likelihood, $\vec{v}$ will not be on the line itself. We only need $\vec{v}$ to be parallel to the line. Finally, let $P = (x, y, z)$ be any point on the line.

Now, since our "slope" is a vector let's also represent the two points on the line as vectors. We'll do this with position vectors. So, let $\vec{r_0}$ and $\vec{r}$ be the position vectors for P_0 and P respectively. Also, for no apparent reason, let's define $\vec{a}$ to be the vector with representation $\overrightarrow{P_0P}$.

We now have the following sketch with all these points and vectors on it.



Now, we've shown the parallel vector, $\vec{v}$, as a position vector but it doesn't need to be a position vector. It can be anywhere, a position vector, on the line or off the line, it just needs to be parallel to the line.

Next, notice that we can write $\vec{r}$ as follows,

$$\vec{r} = \vec{r_0} + \vec{a}$$

If you're not sure about this go back and check out the sketch for vector addition in the vector arithmetic section. Now, notice that the vectors $\vec{a}$ and $\vec{v}$ are parallel. Therefore there is a number, t, such that

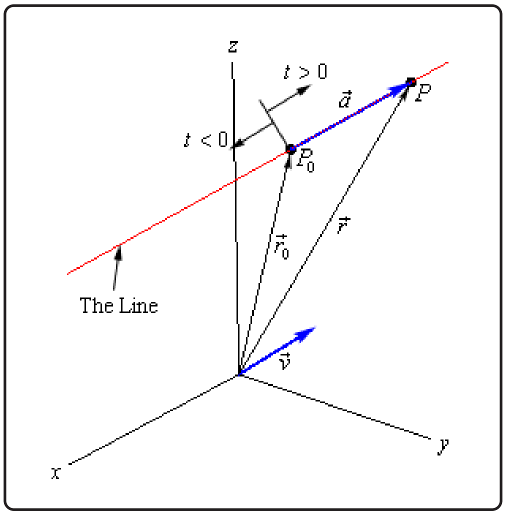
$$\vec{a} = t\,\vec{v}$$

We now have,

Vector Form of a Line

$$\vec{r} = \overrightarrow{r_0} + t\,\vec{v} = \langle x_0, y_0, z_0 \rangle + t\,\langle a, b, c \rangle$$

The only part of this equation that is not known is the t. Notice that $t\,\vec{v}$ will be a vector that lies along the line and it tells us how far from the original point that we should move. If t is positive we move away from the original point in the direction of $\vec{v}$ (right in our sketch) and if t is negative we move away from the original point in the opposite direction of $\vec{v}$ (left in our sketch). As t varies over all possible values we will completely cover the line. The following sketch shows this dependence on t of our sketch.



There are several other forms of the equation of a line. To get the first alternate form let's start with the vector form and do a slight rewrite.

$$\vec{r} = \langle x_0, y_0, z_0 \rangle + t\,\langle a, b, c \rangle$$
$$\langle x, y, z \rangle = \langle x_0 + ta, y_0 + tb, z_0 + tc \rangle$$

The only way for two vectors to be equal is for the components to be equal. In other words,

Parametric Form of a Line

$$x = x_0 + ta$$
$$y = y_0 + tb$$
$$z = z_0 + tc$$

Notice that this is really nothing more than an extension of the parametric equations we've seen previously. The only difference is that we are now working in three dimensions instead of two dimensions.

To get a point on the line all we do is pick a t and plug into either form of the line. In the vector form of the line we get a position vector for the point and in the parametric form we get the actual coordinates of the point.

There is one more form of the line that we want to look at. If we assume that a, b, and c are all non-zero numbers we can solve each of the equations in the parametric form of the line for t. We can then set all of them equal to each other since t will be the same number in each. Doing this gives the following,

Symmetric Equations of a Line

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

If one of a, b, or c does happen to be zero we can still write down the symmetric equations. To see this let's suppose that $b = 0$. In this case t will not exist in the parametric equation for y and so we will only solve the parametric equations for x and z for t. We then set those equal and acknowledge the parametric equation for y as follows,

$$\frac{x - x_0}{a} = \frac{z - z_0}{c} \qquad\qquad y = y_0$$

Let's take a look at an example.

Example 1

Write down the equation of the line that passes through the points $(2, -1, 3)$ and $(1, 4, -3)$. Write down all three forms of the equation of the line.

Solution

To do this we need the vector $\vec{v}$ that will be parallel to the line. This can be any vector as long as it's parallel to the line. In general, $\vec{v}$ won't lie on the line itself. However, in this case it will. All we need to do is let $\vec{v}$ be the vector that starts at the second point and ends at the first point. Since these two points are on the line the vector between them will also lie on the line and will hence be parallel to the line. So,

$$\vec{v} = \langle 1, -5, 6 \rangle$$

Note that the order of the points was chosen to reduce the number of minus signs in the vector. We could just have easily gone the other way.

Once we've got $\vec{v}$ there really isn't anything else to do. To use the vector form we'll need a point on the line. We've got two and so we can use either one. We'll use the first point. Here is the vector form of the line.

$$\vec{r} = \langle 2, -1, 3 \rangle + t \langle 1, -5, 6 \rangle = \langle 2 + t, -1 - 5t, 3 + 6t \rangle$$

Once we have this equation the other two forms follow. Here are the parametric equations of the line.

$$\begin{aligned} x &= 2 + t \\ y &= -1 - 5t \\ z &= 3 + 6t \end{aligned}$$

Here is the symmetric form.

$$\frac{x-2}{1} = \frac{y+1}{-5} = \frac{z-3}{6}$$

Example 2

Determine if the line that passes through the point $(0, -3, 8)$ and is parallel to the line given by $x = 10 + 3t$, $y = 12\,t$ and $z = -3 - t$ passes through the xz-plane. If it does give the coordinates of that point.

Solution

To answer this we will first need to write down the equation of the line. We know a point on the line and just need a parallel vector. We know that the new line must be parallel to the line given by the parametric equations in the problem statement. That means that any vector that is parallel to the given line must also be parallel to the new line.

Now recall that in the parametric form of the line the numbers multiplied by t are the components of the vector that is parallel to the line. Therefore, the vector,

$$\vec{v} = \langle 3, 12, -1 \rangle$$

is parallel to the given line and so must also be parallel to the new line.

The equation of new line is then,

$$\vec{r} = \langle 0, -3, 8 \rangle + t \langle 3, 12, -1 \rangle = \langle 3t, -3 + 12t, 8 - t \rangle$$

If this line passes through the xz-plane then we know that the y-coordinate of that point must be zero. So, let's set the y component of the equation equal to zero and see if we can solve for t. If we can, this will give the value of t for which the point will pass through the xz-plane.

$$-3 + 12t = 0 \qquad \Rightarrow \qquad t = \frac{1}{4}$$

So, the line does pass through the xz-plane. To get the complete coordinates of the point all we need to do is plug $t = \frac{1}{4}$ into any of the equations. We'll use the vector form.

$$\vec{r} = \left\langle 3\left(\frac{1}{4}\right), -3 + 12\left(\frac{1}{4}\right), 8 - \frac{1}{4} \right\rangle = \left\langle \frac{3}{4}, 0, \frac{31}{4} \right\rangle$$

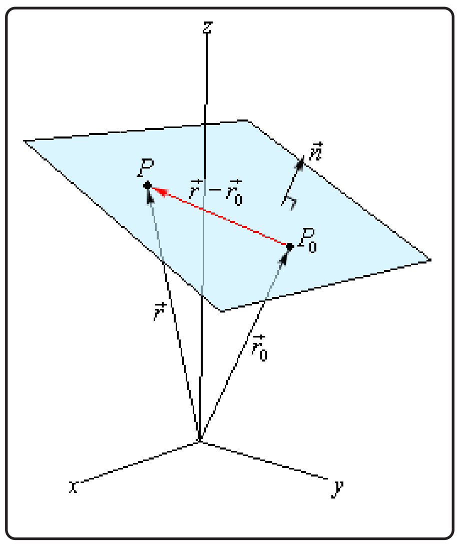
Recall that this vector is the position vector for the point on the line and so the coordinates of the point where the line will pass through the xz-plane are $\left(\frac{3}{4}, 0, \frac{31}{4}\right)$.

12.3 Equations of Planes

In the first section of this chapter we saw a couple of equations of planes. However, none of those equations had three variables in them and were really extensions of graphs that we could look at in two dimensions. We would like a more general equation for planes.

So, let's start by assuming that we know a point that is on the plane, $P_0 = (x_0, y_0, z_0)$. Let's also suppose that we have a vector that is orthogonal (perpendicular) to the plane, $\vec{n} = \langle a, b, c\rangle$. This vector is called the **normal vector**. Now, assume that $P = (x, y, z)$ is any point in the plane. Finally, since we are going to be working with vectors initially we'll let $\overrightarrow{r_0}$ and $\vec{r}$ be the position vectors for P_0 and P respectively.

Here is a sketch of all these vectors.



Notice that we added in the vector $\vec{r} - \overrightarrow{r_0}$ which will lie completely in the plane. Also notice that we put the normal vector on the plane, but there is actually no reason to expect this to be the case. We put it here to illustrate the point. It is completely possible that the normal vector does not touch the plane in any way.

Now, because $\vec{n}$ is orthogonal to the plane, it's also orthogonal to any vector that lies in the plane. In particular it's orthogonal to $\vec{r} - \overrightarrow{r_0}$. Recall from the Dot Product section that two orthogonal vectors will have a dot product of zero. In other words,

$$\vec{n} \cdot (\vec{r} - \overrightarrow{r_0}) = 0 \qquad \Rightarrow \qquad \vec{n} \cdot \vec{r} = \vec{n} \cdot \overrightarrow{r_0}$$

This is called the **vector equation of the plane**.

A slightly more useful form of the equations is as follows. Start with the first form of the vector equation and write down a vector for the difference.

$$\begin{aligned} \langle a, b, c\rangle \cdot (\langle x, y, z\rangle - \langle x_0, y_0, z_0\rangle) &= 0 \\ \langle a, b, c\rangle \cdot \langle x - x_0, y - y_0, z - z_0\rangle &= 0 \end{aligned}$$

Now, actually compute the dot product to get,

Scalar equation of the plane

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

Often this will be written as,

$$ax + by + cz = d$$

where $d = ax_0 + by_0 + cz_0$.

This second form is often how we are given equations of planes. Notice that if we are given the equation of a plane in this form we can quickly get a normal vector for the plane. A normal vector is,

$$\vec{n} = \langle a, b, c \rangle$$

Let's work a couple of examples.

Example 1

Determine the equation of the plane that contains the points $P = (1, -2, 0)$, $Q = (3, 1, 4)$ and $R = (0, -1, 2)$.

Solution

In order to write down the equation of plane we need a point (we've got three so we're cool there) and a normal vector. We need to find a normal vector. Recall however, that we saw how to do this in the Cross Product section.

We can form the following two vectors from the given points.

$$\overrightarrow{PQ} = \langle 2, 3, 4 \rangle \qquad \overrightarrow{PR} = \langle -1, 1, 2 \rangle$$

These two vectors will lie completely in the plane since we formed them from points that were in the plane. Notice as well that there are many possible vectors to use here, we just chose two of the possibilities.

Now, we know that the cross product of two vectors will be orthogonal to both of these vectors. Since both of these are in the plane any vector that is orthogonal to both of these will also be orthogonal to the plane. Therefore, we can use the cross product as the normal vector.

$$\vec{n} = \overrightarrow{PQ} \times \overrightarrow{PR} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 2 & 3 & 4 \\ -1 & 1 & 2 \end{vmatrix} \begin{matrix} \vec{i} & \vec{j} \\ 2 & 3 \\ -1 & 1 \end{matrix} = 2\vec{i} - 8\vec{j} + 5\vec{k}$$

The equation of the plane is then,

$$\begin{aligned} 2(x-1) - 8(y+2) + 5(z-0) &= 0 \\ 2x - 8y + 5z &= 18 \end{aligned}$$

We used P for the point but could have used any of the three points.

Example 2

Determine if the plane given by $-x + 2z = 10$ and the line given by $\vec{r} = \langle 5, 2-t, 10+4t \rangle$ are orthogonal, parallel or neither.

Solution

This is not as difficult a problem as it may at first appear to be. We can pick off a vector that is normal to the plane. This is $\vec{n} = \langle -1, 0, 2 \rangle$. We can also get a vector that is parallel to the line. This is $v = \langle 0, -1, 4 \rangle$.

Now, if these two vectors are parallel then the line and the plane will be orthogonal. If you think about it this makes some sense. If $\vec{n}$ and $\vec{v}$ are parallel, then $\vec{v}$ is orthogonal to the plane, but $\vec{v}$ is also parallel to the line. So, if the two vectors are parallel the line and plane will be orthogonal.

Let's check this.

$$\vec{n} \times \vec{v} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ -1 & 0 & 2 \\ 0 & -1 & 4 \end{vmatrix} \begin{matrix} \vec{i} & \vec{j} \\ -1 & 0 \\ 0 & -1 \end{matrix} = 2\vec{i} + 4\vec{j} + \vec{k} \neq \vec{0}$$

So, the vectors aren't parallel and so the plane and the line are not orthogonal.

Now, let's check to see if the plane and line are parallel. If the line is parallel to the plane then any vector parallel to the line will be orthogonal to the normal vector of the plane. In other words, if $\vec{n}$ and $\vec{v}$ are orthogonal then the line and the plane will be parallel.

Let's check this.

$$\vec{n} \cdot \vec{v} = 0 + 0 + 8 = 8 \neq 0$$

The two vectors aren't orthogonal and so the line and plane aren't parallel.

So, the line and the plane are neither orthogonal nor parallel.

12.4 Quadric Surfaces

In the previous two sections we've looked at lines and planes in three dimensions (or $\mathbb{R}^3$) and while these are used quite heavily at times in a Calculus class there are many other surfaces that are also used fairly regularly and so we need to take a look at those.

In this section we are going to be looking at quadric surfaces. Quadric surfaces are the graphs of any equation that can be put into the general form

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J = 0$$

where A, ... , J are constants.

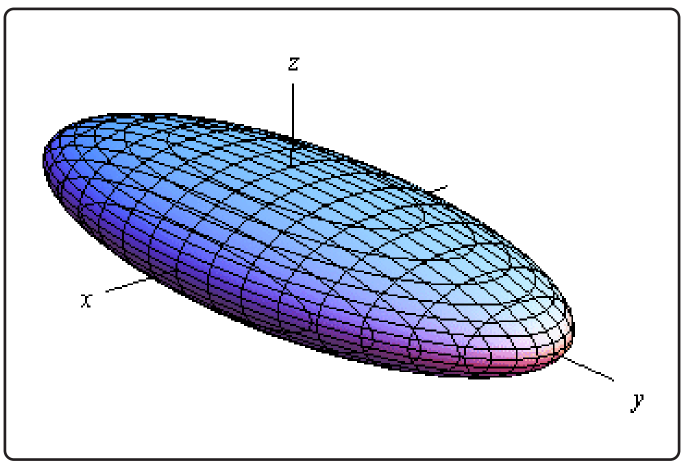
There is no way that we can possibly list all of them, but there are some standard equations so here is a list of some of the more common quadric surfaces.

Ellipsoid

Here is the general equation of an ellipsoid.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

Here is a sketch of a typical ellipsoid.



If $a = b = c$ then we will have a sphere.

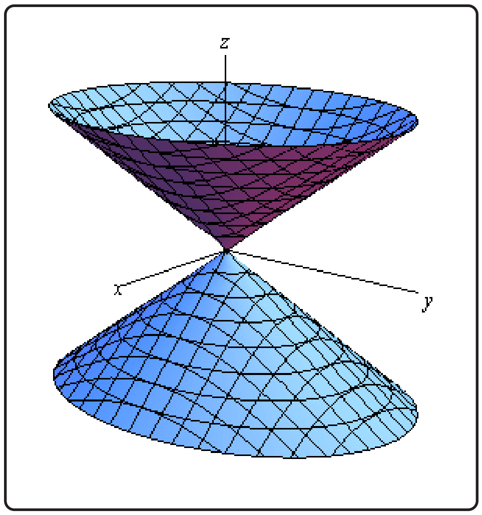
Notice that we only gave the equation for the ellipsoid that has been centered on the origin. Clearly ellipsoids don't have to be centered on the origin. However, in order to make the discussion in this section a little easier we have chosen to concentrate on surfaces that are "centered" on the origin in one way or another.

Cone

Here is the general equation of a cone.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z^2}{c^2}$$

Here is a sketch of a typical cone.



Now, note that while we called this a cone it is more of an hour glass shape rather than what most would call a cone. Of course, the upper and the lower portion of the hour glass really are cones as we would normally think of them.

That brings up the question of what if we really did just want the upper or lower portion (*i.e.* a cone in the traditional sense)? That is easy enough to answer. All we need to do is solve the given equation for z as follows,

$$z^2 = c^2\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right) = \frac{c^2}{a^2}x^2 + \frac{c^2}{b^2}y^2 = A^2x^2 + B^2y^2 \quad \rightarrow \quad z = \pm\sqrt{A^2x^2 + B^2y^2}$$

We simplified the coefficients a little to make it the equation(s) easier to deal with. Now, we know that square roots always return positive numbers and so we can then see that $z = \sqrt{A^2x^2 + B^2y^2}$ will always be positive and so be the equation for just the upper portion of the "cone" above. Likewise, $z = -\sqrt{A^2x^2 + B^2y^2}$ will always be negative and so be the equation of just the lower portion of the "cone" above.

Also, note that this is the equation of a cone that will open along the z-axis. To get the equation of a cone that opens along one of the other axes all we need to do is make a slight modification of the equation. This will be the case for the rest of the surfaces that we'll be looking at in this section as well.

In the case of a cone the variable that sits by itself on one side of the equal sign will determine the axis that the cone opens up along. For instance, a cone that opens up along the x-axis will have the equation,

$$\frac{y^2}{b^2} + \frac{z^2}{c^2} = \frac{x^2}{a^2}$$

For most of the following surfaces we will not give the other possible formulas. We will however acknowledge how each formula needs to be changed to get a change of orientation for the surface.

Cylinder

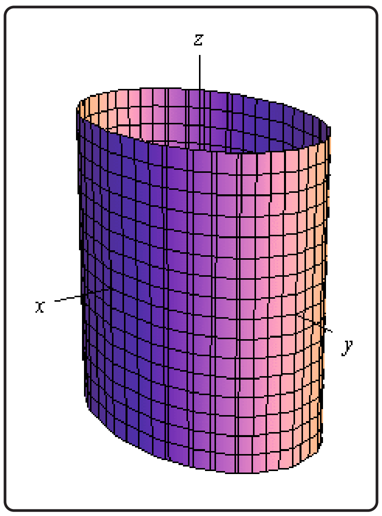
Here is the general equation of a cylinder.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

This is a cylinder whose cross section is an ellipse. If $a = b$ we have a cylinder whose cross section is a circle. We'll be dealing with those kinds of cylinders more than the general form so the equation of a cylinder with a circular cross section is,

$$x^2 + y^2 = r^2$$

Here is a sketch of typical cylinder with an ellipse cross section.



The cylinder will be centered on the axis corresponding to the variable that does not appear in the equation.

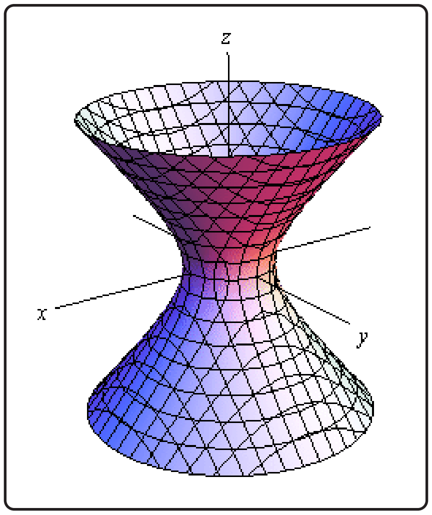
Be careful to not confuse this with a circle. In two dimensions it is a circle, but in three dimensions it is a cylinder.

Hyperboloid of One Sheet

Here is the equation of a hyperboloid of one sheet.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$$

Here is a sketch of a typical hyperboloid of one sheet.



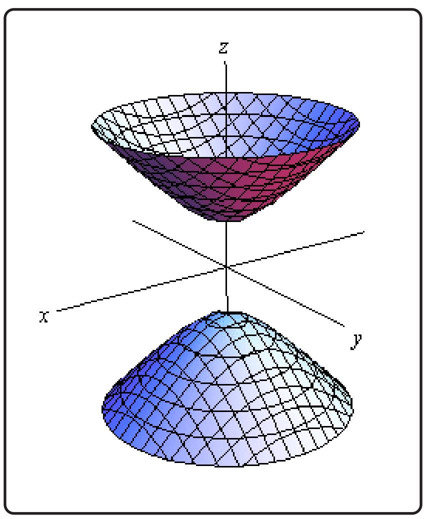
The variable with the negative in front of it will give the axis along which the graph is centered.

Hyperboloid of Two Sheets

Here is the equation of a hyperboloid of two sheets.

$$-\frac{x^2}{a^2} - \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

Here is a sketch of a typical hyperboloid of two sheets.



The variable with the positive in front of it will give the axis along which the graph is centered.

Notice that the only difference between the hyperboloid of one sheet and the hyperboloid of two sheets is the signs in front of the variables. They are exactly the opposite signs.

Also note that just as we could do with cones, if we solve the equation for z the positive portion will give the equation for the upper part of this while the negative portion will give the equation for the lower part of this.

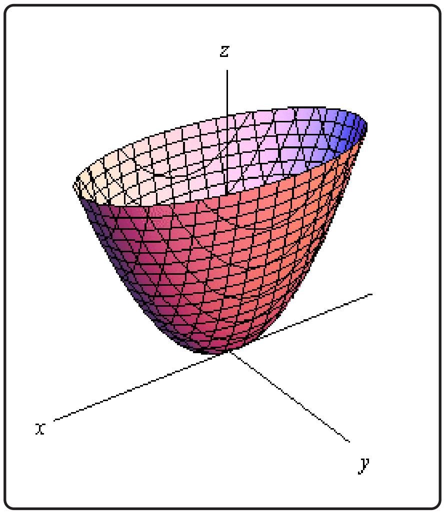
Elliptic Paraboloid

Here is the equation of an elliptic paraboloid.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z}{c}$$

As with cylinders this has a cross section of an ellipse and if $a = b$ it will have a cross section of a circle. When we deal with these we'll generally be dealing with the kind that have a circle for a cross section.

Here is a sketch of a typical elliptic paraboloid.



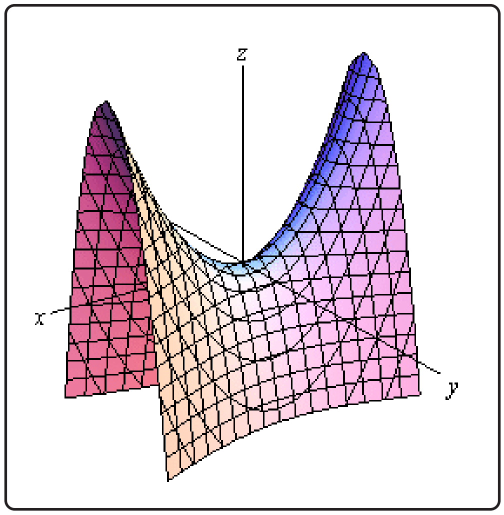
In this case the variable that isn't squared determines the axis upon which the paraboloid opens up. Also, the sign of c will determine the direction that the paraboloid opens. If c is positive then it opens up and if c is negative then it opens down.

Hyperbolic Paraboloid

Here is the equation of a hyperbolic paraboloid.

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = \frac{z}{c}$$

Here is a sketch of a typical hyperbolic paraboloid.



These graphs are vaguely saddle shaped and as with the elliptic paraboloid the sign of c will determine the direction in which the surface "opens up". The graph above is shown for c positive.

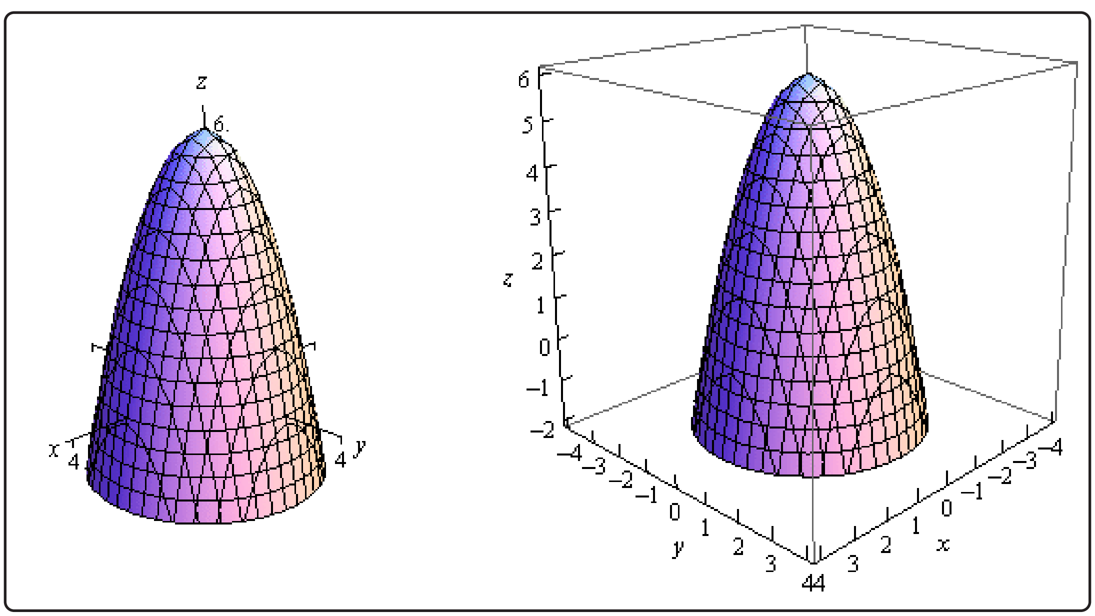
With both of the types of paraboloids discussed above note that the surface can be easily moved up or down by adding/subtracting a constant from the left side.

For instance

$$z = -x^2 - y^2 + 6$$

is an elliptic paraboloid that opens downward (be careful, the "$-$" is on the x and y instead of the z) and starts at $z = 6$ instead of $z = 0$.

Here are a couple of quick sketches of this surface.

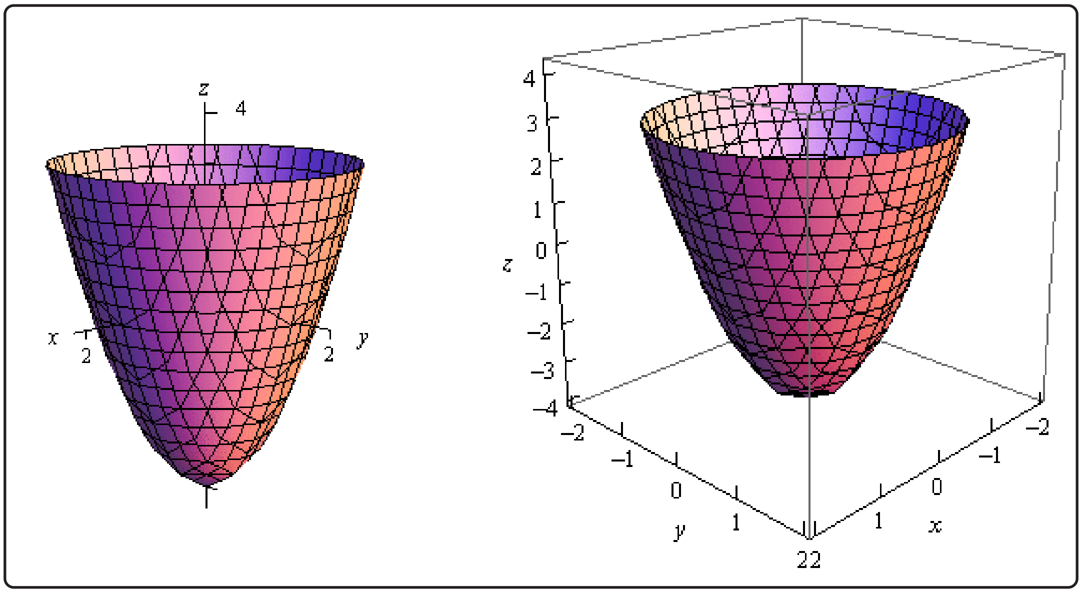


Note that we've given two forms of the sketch here. The sketch on the left has the standard set of axes but it is difficult to see the numbers on the axis. The sketch on the right has been "boxed" and this makes it easier to see the numbers to give a sense of perspective to the sketch. In most sketches that actually involve numbers on the axis system we will give both sketches to help get a feel for what the sketch looks like.

12.5 Functions of Several Variables

In this section we want to go over some of the basic ideas about functions of more than one variable.

First, remember that graphs of functions of two variables, $z = f(x, y)$ are surfaces in three dimensional space. For example, here is the graph of $z = 2x^2 + 2y^2 - 4$.



This is an elliptic paraboloid and is an example of a quadric surface. We saw several of these in the previous section. We will be seeing quadric surfaces fairly regularly later on when we start discussion Multi-variable Calculus.

Another common graph that we'll be seeing quite a bit in this course is the graph of a plane. We have a convention for graphing planes that will make them a little easier to graph and hopefully visualize.

Recall that the equation of a plane is given by

$$ax + by + cz = d$$

or if we solve this for z we can write it in terms of function notation. This gives,

$$f(x, y) = Ax + By + D$$

To graph a plane we will generally find the intersection points with the three axes and then graph the triangle that connects those three points. This triangle will be a portion of the plane and it will give us a fairly decent idea on what the plane itself should look like. For example, let's graph the plane given by,

$$f(x, y) = 12 - 3x - 4y$$

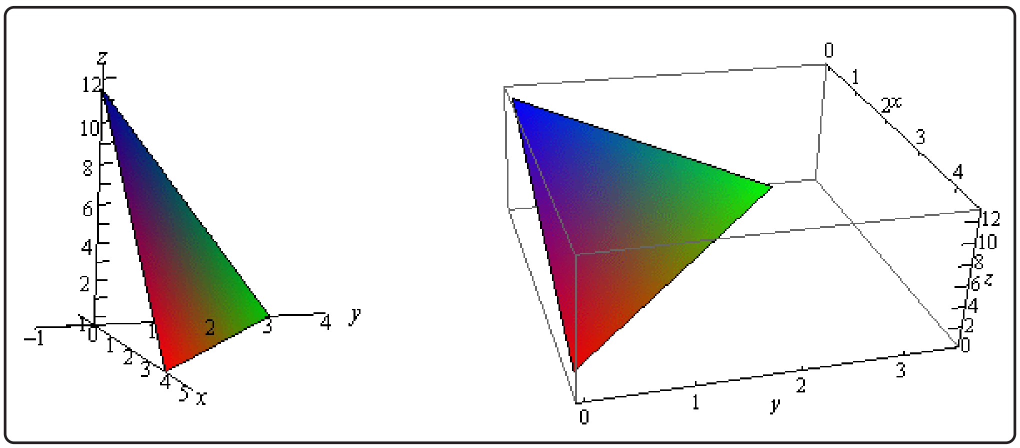
For purposes of graphing this it would probably be easier to write this as,

$$z = 12 - 3x - 4y \qquad \Rightarrow \qquad 3x + 4y + z = 12$$

Now, each of the intersection points with the three main coordinate axes is defined by the fact that two of the coordinates are zero. For instance, the intersection with the z-axis is defined by $x = y = 0$. So, the three intersection points are,

$$\begin{aligned} x-\text{axis}: &\ (4,0,0) \\ y-\text{axis}: &\ (0,3,0) \\ z-\text{axis}: &\ (0,0,12) \end{aligned}$$

Here is the graph of the plane.



Now, to extend this out, graphs of functions of the form $w = f(x, y, z)$ would be four dimensional surfaces. Of course, we can't graph them, but it doesn't hurt to point this out.

We next want to talk about the domains of functions of more than one variable. Recall that domains of functions of a single variable, $y = f(x)$, consisted of all the values of x that we could plug into the function and get back a real number. Now, if we think about it, this means that the domain of a function of a single variable is an interval (or intervals) of values from the number line, or one dimensional space.

The domain of functions of two variables, $z = f(x, y)$, are regions from two dimensional space and consist of all the coordinate pairs, (x, y), that we could plug into the function and get back a real number.

Example 1

Determine the domain of each of the following.

(a) $f(x,y) = \sqrt{x+y}$

(b) $f(x,y) = \sqrt{x} + \sqrt{y}$

(c) $f(x,y) = \ln\left(9 - x^2 - 9y^2\right)$

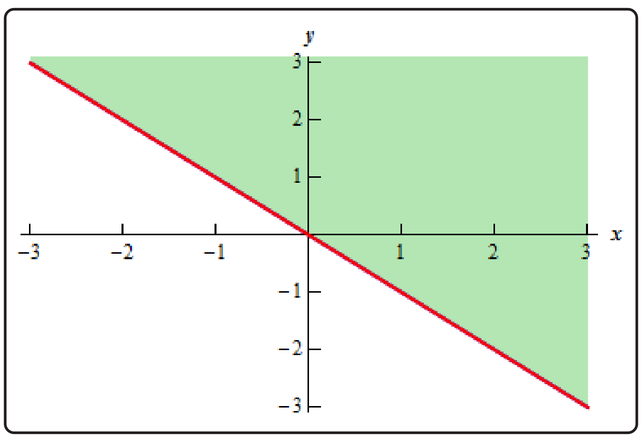
Solution

(a) $f(x,y) = \sqrt{x+y}$

In this case we know that we can't take the square root of a negative number so this means that we must require,

$$x + y \geq 0$$

Here is a sketch of the graph of this region.

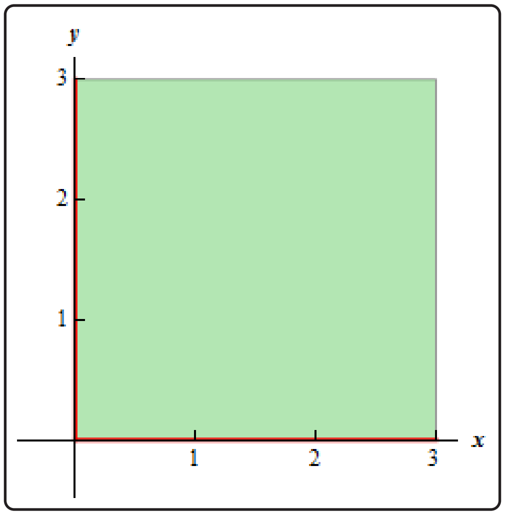


(b) $f(x,y) = \sqrt{x} + \sqrt{y}$

This function is different from the function in the previous part. Here we must require that,

$$x \geq 0 \qquad \text{and} \quad y \geq 0$$

and they really do need to be separate inequalities. There is one for each square root in the function. Here is the sketch of this region.

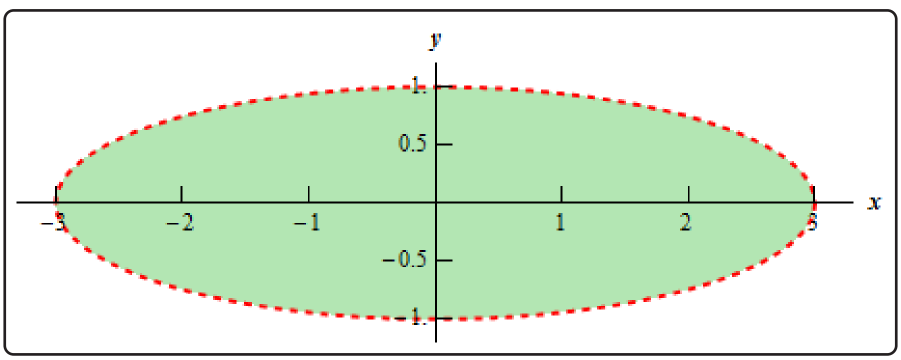


(c) $f(x,y) = \ln\left(9 - x^2 - 9y^2\right)$

In this final part we know that we can't take the logarithm of a negative number or zero. Therefore, we need to require that,

$$9 - x^2 - 9y^2 > 0 \qquad \Rightarrow \qquad \frac{x^2}{9} + y^2 < 1$$

and upon rearranging we see that we need to stay interior to an ellipse for this function. Here is a sketch of this region.



Note that domains of functions of three variables, $w = f(x,y,z)$, will be regions in three dimensional space.

Example 2

Determine the domain of the following function,

$$f(x,y,z) = \frac{1}{\sqrt{x^2+y^2+z^2-16}}$$

Solution

In this case we have to deal with the square root and division by zero issues. These will require,

$$x^2+y^2+z^2-16>0 \qquad \Rightarrow \qquad x^2+y^2+z^2>16$$

So, the domain for this function is the set of points that lies completely outside a sphere of radius 4 centered at the origin.

The next topic that we should look at is that of **level curves** or **contour curves**. The level curves of the function $z = f(x,y)$ are two dimensional curves we get by setting $z = k$, where k is any number. So the equations of the level curves are $f(x,y) = k$. Note that sometimes the equation will be in the form $f(x,y,z) = 0$ and in these cases the equations of the level curves are $f(x,y,k) = 0$.

You've probably seen level curves (or contour curves, whatever you want to call them) before. If you've ever seen the elevation map for a piece of land, this is nothing more than the contour curves for the function that gives the elevation of the land in that area. Of course, we probably don't have the function that gives the elevation, but we can at least graph the contour curves.

Let's do a quick example of this.

Example 3

Identify the level curves of $f(x,y) = \sqrt{x^2+y^2}$. Sketch a few of them.

Solution

First, for the sake of practice, let's identify what this surface given by $f(x,y)$ is. To do this let's rewrite it as,

$$z = \sqrt{x^2+y^2}$$

Recall from the Quadric Surfaces section that this the upper portion of the "cone" (or hour glass shaped surface).

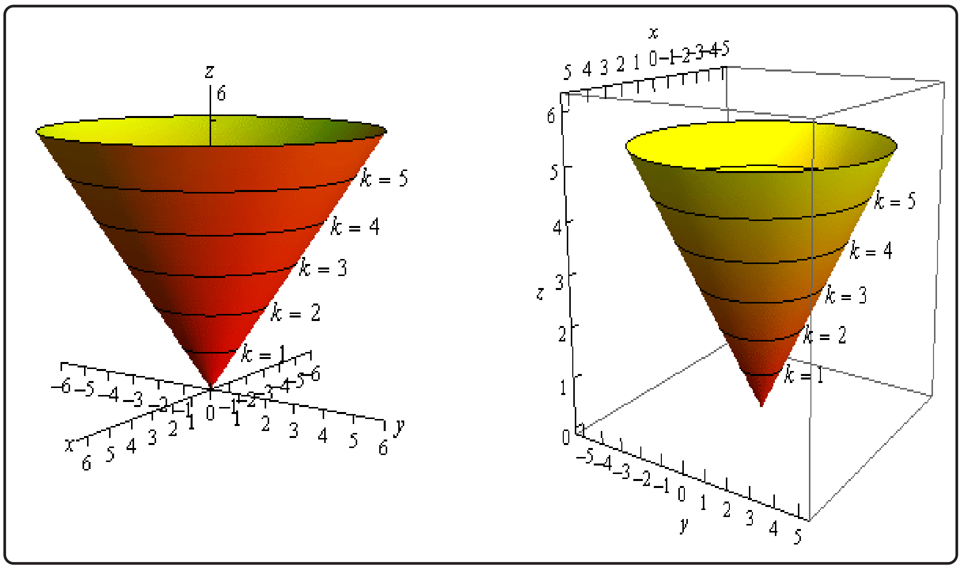
Note that this was not required for this problem. It was done for the practice of identifying the surface and this may come in handy down the road.

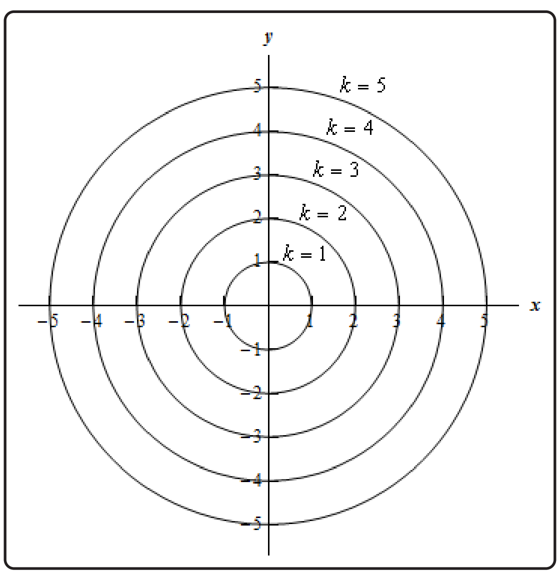
Now on to the real problem. The level curves (or contour curves) for this surface are given by the equation are found by substituting $z = k$. In the case of our example this is,

$$k = \sqrt{x^2 + y^2} \qquad \Rightarrow \qquad x^2 + y^2 = k^2$$

where k is any number. So, in this case, the level curves are circles of radius k with center at the origin.

We can graph these in one of two ways. We can either graph them on the surface itself or we can graph them in a two dimensional axis system. Here is each graph for some values of k.





Note that we can think of contours in terms of the intersection of the surface that is given by $z = f(x, y)$ and the plane $z = k$. The contour will represent the intersection of the surface and the plane.

For functions of the form $f(x, y, z)$ we will occasionally look at **level surfaces**. The equations of level surfaces are given by $f(x, y, z) = k$ where k is any number.

The final topic in this section is that of **traces**. In some ways these are similar to contours. As noted above we can think of contours as the intersection of the surface given by $z = f(x, y)$ and the plane $z = k$. Traces of surfaces are curves that represent the intersection of the surface and the plane given by $x = a$ or $y = b$.

Let's take a quick look at an example of traces.

Example 4

Sketch the traces of $f(x, y) = 10 - 4x^2 - y^2$ for the plane $x = 1$ and $y = 2$.

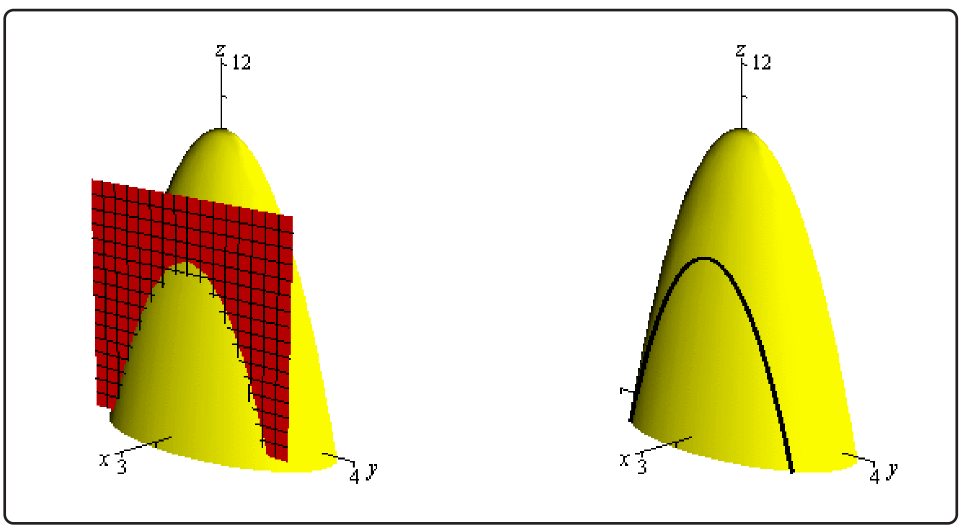
Solution

We'll start with $x = 1$. We can get an equation for the trace by plugging $x = 1$ into the equation. Doing this gives,

$$z = f(1, y) = 10 - 4(1)^2 - y^2 \qquad \Rightarrow \qquad z = 6 - y^2$$

and this will be graphed in the plane given by $x = 1$.

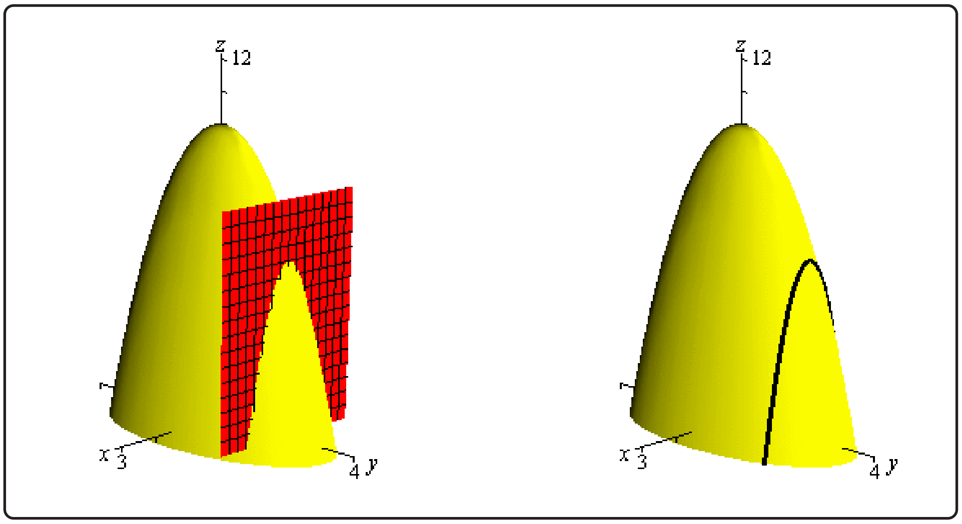
Below are two graphs. The graph on the left is a graph showing the intersection of the surface and the plane given by $x = 1$. On the right is a graph of the surface and the trace that we are after in this part.



For $y = 2$ we will do pretty much the same thing that we did with the first part. Here is the equation of the trace,

$$z = f(x, 2) = 10 - 4x^2 - (2)^2 \qquad \Rightarrow \qquad z = 6 - 4x^2$$

and here are the sketches for this case.



12.6 Vector Functions

We first saw vector functions back when we were looking at the Equation of Lines. In that section we talked about them because we wrote down the equation of a line in $\mathbb{R}^3$ in terms of a *vector function* (sometimes called a *vector-valued function*). In this section we want to look a little closer at them and we also want to look at some vector functions in $\mathbb{R}^3$ other than lines.

A vector function is a function that takes one or more variables and returns a vector. We'll spend most of this section looking at vector functions of a single variable as most of the places where vector functions show up here will be vector functions of single variables. We will however briefly look at vector functions of two variables at the end of this section.

A vector functions of a single variable in $\mathbb{R}^2$ and $\mathbb{R}^3$ have the form,

$$\vec{r}(t) = \langle f(t), g(t) \rangle \qquad\qquad \vec{r}(t) = \langle f(t), g(t), h(t) \rangle$$

respectively, where $f(t)$, $g(t)$ and $h(t)$ are called the **component functions**.

The main idea that we want to discuss in this section is that of graphing and identifying the graph given by a vector function. Before we do that however, we should talk briefly about the domain of a vector function. The **domain** of a vector function is the set of all t's for which all the component functions are defined.

Example 1

Determine the domain of the following function.

$$\vec{r}(t) = \left\langle \cos(t), \ln(4-t), \sqrt{t+1} \right\rangle$$

Solution

The first component is defined for all t's. The second component is only defined for $t < 4$. The third component is only defined for $t \geq -1$. Putting all of these together gives the following domain.

$$[-1, 4)$$

This is the largest possible interval for which all three components are defined.

Let's now move into looking at the graph of vector functions. In order to graph a vector function all we do is think of the vector returned by the vector function as a position vector for points on the graph. Recall that a position vector, say $\vec{v} = \langle a, b, c \rangle$, is a vector that starts at the origin and ends at the point (a, b, c).

So, in order to sketch the graph of a vector function all we need to do is plug in some values of t and then plot points that correspond to the resulting position vector we get out of the vector function.

Because it is a little easier to visualize things we'll start off by looking at graphs of vector functions in $\mathbb{R}^2$.

Example 2

Sketch the graph of each of the following vector functions.

(a) $\vec{r}(t) = \langle t, 1\rangle$

(b) $\vec{r}(t) = \left\langle t, t^3 - 10t + 7\right\rangle$

Solution

(a) $\vec{r}(t) = \langle t, 1\rangle$

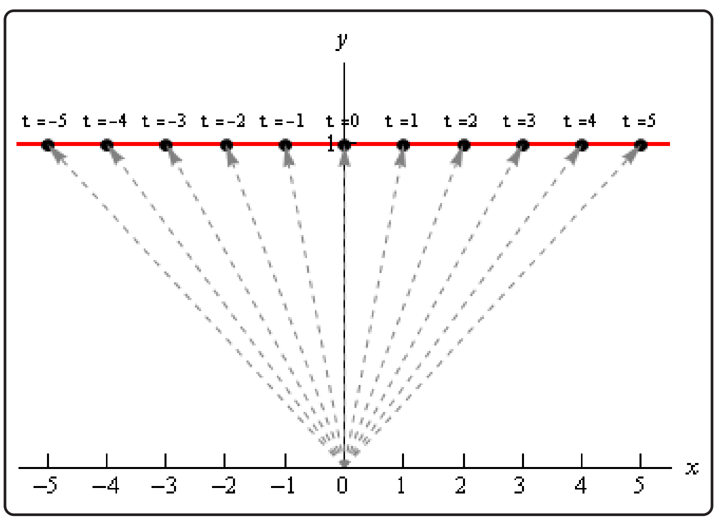
Okay, the first thing that we need to do is plug in a few values of t and get some position vectors. Here are a few,

$$\vec{r}(-3) = \langle -3, 1\rangle \qquad \vec{r}(-1) = \langle -1, 1\rangle \qquad \vec{r}(2) = \langle 2, 1\rangle \qquad \vec{r}(5) = \langle 5, 1\rangle$$

So, what this tells us is that the following points are all on the graph of this vector function.

$$(-3, 1) \qquad (-1, 1) \qquad (2, 1) \qquad (5, 1)$$

Here is a sketch of this vector function.



In this sketch we've included many more evaluations than just those above. Also note that we've put in the position vectors (in gray and dashed) so you can see how all this is working. Note however, that in practice the position vectors are generally not included in the sketch.

In this case it looks like we've got the graph of the line $y = 1$.

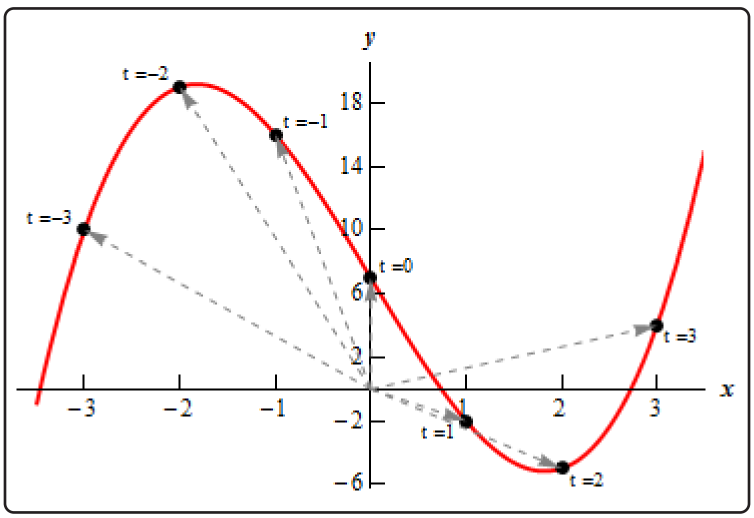
(b) $\vec{r}(t) = \langle t, t^3 - 10t + 7 \rangle$

Here are a couple of evaluations for this vector function.

$$\vec{r}(-3) = \langle -3, 10 \rangle \qquad \vec{r}(-1) = \langle -1, 16 \rangle \qquad \vec{r}(1) = \langle 1, -2 \rangle \qquad \vec{r}(3) = \langle 3, 4 \rangle$$

So, we've got a few points on the graph of this function. However, unlike the first part this isn't really going to be enough points to get a good idea of this graph. In general, it can take quite a few function evaluations to get an idea of what the graph is and it's usually easier to use a computer to do the graphing.

Here is a sketch of this graph. We've put in a few vectors/evaluations to illustrate them, but the reality is that we did have to use a computer to get a good sketch here.



Both of the vector functions in the above example were in the form,

$$\vec{r}(t) = \langle t, g(t) \rangle$$

and what we were really sketching is the graph of $y = g(x)$ as you probably caught onto. Let's graph a couple of other vector functions that do not fall into this pattern.

Example 3

Sketch the graph of each of the following vector functions.

(a) $\vec{r}(t) = \langle 6\cos(t), 3\sin(t) \rangle$

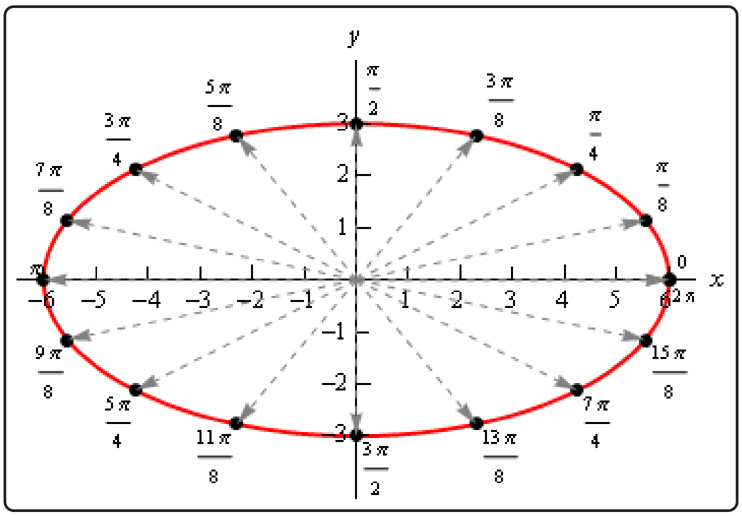
(b) $\vec{r}(t) = \left\langle t - 2\sin(t), t^2 \right\rangle$

Solution

As we saw in the last part of the previous example it can really take quite a few function evaluations to really be able to sketch the graph of a vector function. Because of that we'll be skipping all the function evaluations here and just giving the graph. The main point behind this set of examples is to not get you too locked into the form we were looking at above. The first part will also lead to an important idea that we'll discuss after this example.

So, with that said here are the sketches of each of these.

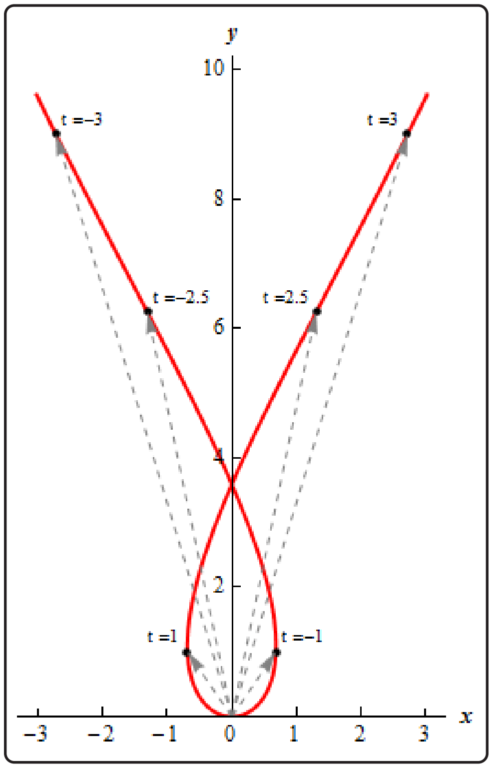
(a) $\vec{r}(t) = \langle 6\cos(t), 3\sin(t) \rangle$



So, in this case it looks like we've got an ellipse.

(b) $\vec{r}(t) = \left\langle t - 2\sin(t), t^2 \right\rangle$

Here's the sketch for this vector function.



Before we move on to vector functions in $\mathbb{R}^3$ let's go back and take a quick look at the first vector function we sketched in the previous example, $\vec{r}(t) = \langle 6\cos(t), 3\sin(t)\rangle$. The fact that we got an ellipse here should not come as a surprise to you. We know that the first component function gives the x coordinate and the second component function gives the y coordinates of the point that we graph. If we strip these out to make this clear we get,

$$x = 6\cos(t) \qquad\qquad y = 3\sin(t)$$

This should look familiar to you. Back when we were looking at Parametric Equations we saw that this was nothing more than one of the sets of parametric equations that gave an ellipse.

This is an important idea in the study of vector functions. Any vector function can be broken down into a set of parametric equations that represent the same graph. In general, the two dimensional vector function, $\vec{r}(t) = \langle f(t), g(t)\rangle$, can be broken down into the parametric equations,

$$x = f(t) \qquad\qquad y = g(t)$$

Likewise, a three dimensional vector function, $\vec{r}(t) = \langle f(t), g(t), h(t)\rangle$, can be broken down into the parametric equations,

$$x = f(t) \qquad y = g(t) \qquad z = h(t)$$

Do not get too excited about the fact that we're now looking at parametric equations in $\mathbb{R}^3$. They work in exactly the same manner as parametric equations in $\mathbb{R}^2$ which we're used to dealing with already. The only difference is that we now have a third component.

Let's take a look at a couple of graphs of vector functions.

Example 4

Sketch the graph of the following vector function.

$$\vec{r}(t) = \langle 2 - 4t, -1 + 5t, 3 + t \rangle$$

Solution

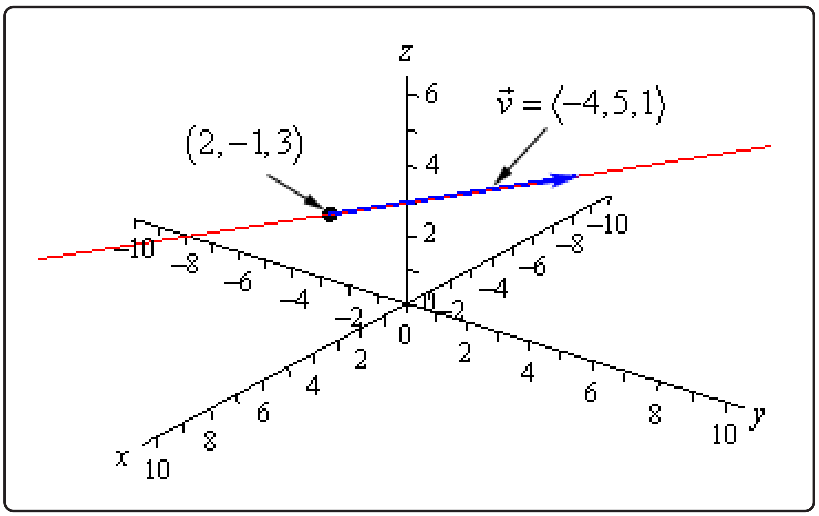
Notice that this is nothing more than a line. It might help if we rewrite it a little.

$$\vec{r}(t) = \langle 2, -1, 3 \rangle + t \langle -4, 5, 1 \rangle$$

In this form we can see that this is the equation of a line that goes through the point $(2, -1, 3)$ and is parallel to the vector $\vec{v} = \langle -4, 5, 1 \rangle$.

To graph this line all that we need to do is plot the point and then sketch in the parallel vector. In order to get the sketch will assume that the vector is on the line and will start at the point in the line. To sketch in the line all we do this is extend the parallel vector into a line.

Here is a sketch.



Example 5

Sketch the graph of the following vector function.

$$\vec{r}(t) = \langle 2\cos(t), 2\sin(t), 3\rangle$$

Solution

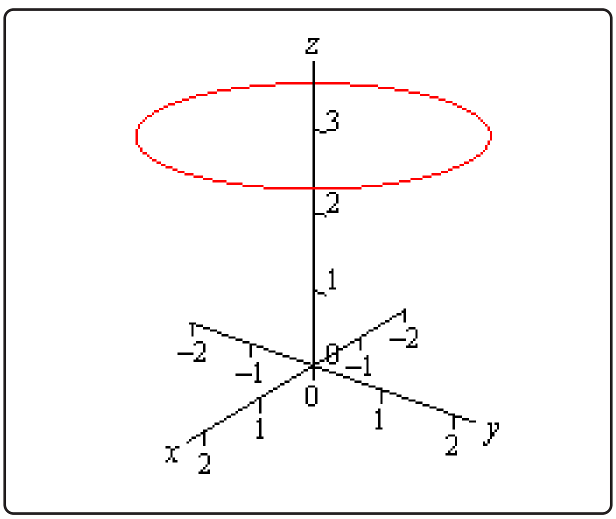
In this case to see what we've got for a graph let's get the parametric equations for the curve.

$$x = 2\cos(t) \qquad y = 2\sin(t) \qquad z = 3$$

If we ignore the z equation for a bit we'll recall (hopefully) that the parametric equations for x and y give a circle of radius 2 centered on the origin (or about the z-axis since we are in $\mathbb{R}^3$).

Now, all the parametric equations here tell us is that no matter what is going on in the graph all the z coordinates must be 3. So, we get a circle of radius 2 centered on the z-axis and at the level of $z = 3$.

Here is a sketch.



Note that it is very easy to modify the above vector function to get a circle centered on the x or y-axis as well. For instance,

$$\vec{r}(t) = \langle 10\sin(t), -3, 10\cos(t)\rangle$$

will be a circle of radius 10 centered on the y-axis and at $y = -3$. In other words, as long as two of the terms are a sine and a cosine (with the same coefficient) and the other is a fixed number then we will have a circle that is centered on the axis that is given by the fixed number.

Let's take a look at a modification of this.

Example 6

Sketch the graph of the following vector function.

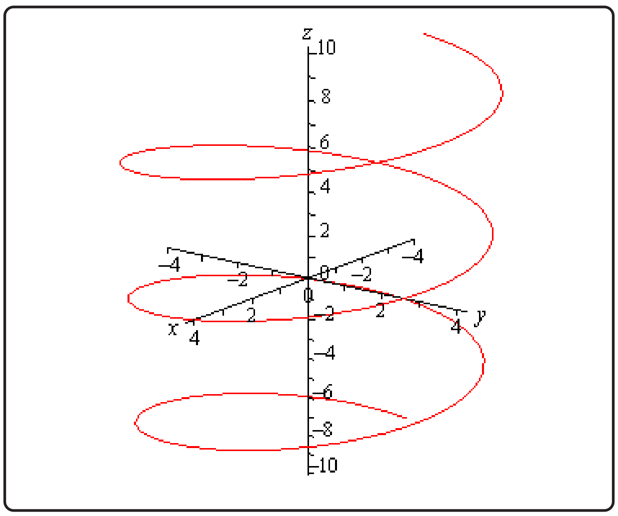
$$\vec{r}(t) = \langle 4\cos(t), 4\sin(t), t\rangle$$

Solution

If this one had a constant in the z component we would have another circle. However, in this case we don't have a constant. Instead we've got a t and that will change the curve. However, because the x and y component functions are still a circle in parametric equations our curve should have a circular nature to it in some way.

In fact, the only change is in the z component and as t increases the z coordinate will increase. Also, as t increases the x and y coordinates will continue to form a circle centered on the z-axis. Putting these two ideas together tells us that as we increase t the circle that is being traced out in the x and y directions should also be rising.

Here is a sketch of this curve.



So, we've got a helix (or spiral, depending on what you want to call it) here.

As with circles the component that has the t will determine the axis that the helix rotates about. For instance,

$$\vec{r}(t) = \langle t, 6\cos(t), 6\sin(t)\rangle$$

is a helix that rotates around the x-axis.

Also note that if we allow the coefficients on the sine and cosine for both the circle and helix to be different we will get ellipses.

For example,

$$\vec{r}(t) = \langle 9\cos(t), t, 2\sin(t) \rangle$$

will be a helix that rotates about the y-axis and is in the shape of an ellipse.

There is a nice formula that we should derive before moving onto vector functions of two variables.

Example 7

Determine the vector equation for the line segment starting at the point $P = (x_1, y_1, z_1)$ and ending at the point $Q = (x_2, y_2, z_2)$.

Solution

It is important to note here that we only want the equation of the line segment that starts at P and ends at Q. We don't want any other portion of the line and we do want the direction of the line segment preserved as we increase t. With all that said, let's not worry about that and just find the vector equation of the line that passes through the two points. Once we have this we will be able to get what we're after.

So, we need a point on the line. We've got two and we will use P. We need a vector that is parallel to the line and since we've got two points we can find the vector between them. This vector will lie on the line and hence be parallel to the line. Also, let's remember that we want to preserve the starting and ending point of the line segment so let's construct the vector using the same "orientation".

$$\vec{v} = \langle x_2 - x_1, y_2 - y_1, z_2 - z_1 \rangle$$

Using this vector and the point P we get the following vector equation of the line.

$$\vec{r}(t) = \langle x_1, y_1, z_1 \rangle + t \langle x_2 - x_1, y_2 - y_1, z_2 - z_1 \rangle$$

While this is the vector equation of the line, let's rewrite the equation slightly.

$$\begin{aligned} \vec{r}(t) &= \langle x_1, y_1, z_1 \rangle + t \langle x_2, y_2, z_2 \rangle - t \langle x_1, y_1, z_1 \rangle \\ &= (1 - t) \langle x_1, y_1, z_1 \rangle + t \langle x_2, y_2, z_2 \rangle \end{aligned}$$

This is the equation of the line that contains the points P and Q. We of course just want the line segment that starts at P and ends at Q. We can get this by simply restricting the values of t.

Notice that

$$\vec{r}(0) = \langle x_1, y_1, z_1 \rangle \qquad \vec{r}(1) = \langle x_2, y_2, z_2 \rangle$$

So, if we restrict t to be between zero and one we will cover the line segment and we will start and end at the correct point.

So, the vector equation of the line segment that starts at $P = (x_1, y_1, z_1)$ and ends at $Q = (x_2, y_2, z_2)$ is,

$$\vec{r}(t) = (1-t)\langle x_1, y_1, z_1\rangle + t\langle x_2, y_2, z_2\rangle \qquad 0 \le t \le 1$$

As noted briefly at the beginning of this section we can also have vector functions of two variables. In these cases the graphs of vector function of two variables are surfaces. So, to make sure that we don't forget that let's work an example with that as well.

Example 8

Identify the surface that is described by $\vec{r}(x,y) = x\,\vec{i} + y\,\vec{j} + \left(x^2+y^2\right)\vec{k}$.

Solution

First, notice that in this case the vector function will in fact be a function of two variables. This will always be the case when we are using vector functions to represent surfaces.

To identify the surface let's go back to parametric equations.

$$x = x \qquad y = y \qquad z = x^2 + y^2$$

The first two are really only acknowledging that we are picking x and y for free and then determining z from our choices of these two. The last equation is the one that we want. We should recognize that function from the section on quadric surfaces. The third equation is the equation of an elliptic paraboloid and so the vector function represents an elliptic paraboloid.

As a final topic for this section let's generalize the idea from the previous example and note that given any function of one variable ($y = f(x)$ or $x = h(y)$) or any function of two variables ($z = g(x,y)$, $x = g(y,z)$, or $y = g(x,z)$) we can always write down a vector form of the equation.

For a function of one variable this will be,

$$\vec{r}(x) = x\,\vec{i} + f(x)\,\vec{j} \qquad \vec{r}(y) = h(y)\,\vec{i} + y\,\vec{j}$$

and for a function of two variables the vector form will be,

$$\vec{r}(x,y) = x\,\vec{i} + y\,\vec{j} + g(x,y)\,\vec{k} \qquad \vec{r}(y,z) = g(y,z)\,\vec{i} + y\,\vec{j} + z\,\vec{k}$$

$$\vec{r}(x,z) = x\,\vec{i} + g(x,z)\,\vec{j} + z\,\vec{k}$$

depending upon the original form of the function.

For example, the hyperbolic paraboloid $y = 2x^2 - 5z^2$ can be written as the following vector function.

$$\vec{r}(x,z) = x\,\vec{i} + \left(2x^2 - 5z^2\right)\,\vec{j} + z\,\vec{k}$$

This is a fairly important idea and we will be doing quite a bit of this kind of thing in multi-variable Calculus.

12.7 Calculus with Vector Functions

In this section we need to talk briefly about limits, derivatives and integrals of vector functions. As you will see, these behave in a fairly predictable manner. We will be doing all of the work in $\mathbb{R}^3$ but we can naturally extend the formulas/work in this section to $\mathbb{R}^n$ (*i.e.* n-dimensional space).

Let's start with limits. Here is the limit of a vector function.

Vector Function Limits

$$\begin{aligned}\lim_{t\to a}\vec{r}(t) &= \lim_{t\to a}\langle f(t), g(t), h(t)\rangle \\ &= \left\langle \lim_{t\to a} f(t), \lim_{t\to a} g(t), \lim_{t\to a} h(t)\right\rangle \\ &= \lim_{t\to a} f(t)\,\vec{i} + \lim_{t\to a} g(t)\,\vec{j} + \lim_{t\to a} h(t)\,\vec{k}\end{aligned}$$

So, all that we do is take the limit of each of the component's functions and leave it as a vector.

Example 1

Compute $\lim\limits_{t\to 1}\vec{r}(t)$ where $\vec{r}(t) = \left\langle t^3, \dfrac{\sin(3t-3)}{t-1}, \mathbf{e}^{2t}\right\rangle$.

Solution

There really isn't all that much to do here.

$$\begin{aligned}\lim_{t\to 1}\vec{r}(t) &= \left\langle \lim_{t\to 1} t^3, \lim_{t\to 1}\frac{\sin(3t-3)}{t-1}, \lim_{t\to 1}\mathbf{e}^{2t}\right\rangle \\ &= \left\langle \lim_{t\to 1} t^3, \lim_{t\to 1}\frac{3\cos(3t-3)}{1}, \lim_{t\to 1}\mathbf{e}^{2t}\right\rangle \\ &= \langle 1, 3, \mathbf{e}^2\rangle\end{aligned}$$

Notice that we had to use L'Hospital's Rule on the y component.

Now let's take care of derivatives and after seeing how limits work it shouldn't be too surprising that we have the following for derivatives.

Vector Function Derivative

$$\vec{r}'(t) = \langle f'(t), g'(t), h'(t)\rangle = f'(t)\,\vec{i} + g'(t)\,\vec{j} + h'(t)\,\vec{k}$$

Example 2

Compute $\vec{r}'(t)$ for $\vec{r}(t) = t^6\,\vec{i} + \sin(2t)\,\vec{j} - \ln(t+1)\,\vec{k}$.

Solution

There really isn't too much to this problem other than taking the derivatives.

$$\vec{r}'(t) = 6t^5\,\vec{i} + 2\cos(2t)\,\vec{j} - \frac{1}{t+1}\vec{k}$$

Most of the basic facts that we know about derivatives still hold however, just to make it clear here are some facts about derivatives of vector functions.

Facts

$$\frac{d}{dt}\left(\vec{u} + \vec{v}\right) = \vec{u}' + \vec{v}'$$

$$\frac{d}{dt}\left(c\vec{u}\right) = c\,\vec{u}'$$

$$\frac{d}{dt}\left(f(t)\,\vec{u}(t)\right) = f'(t)\,\vec{u}(t) + f(t)\,\vec{u}'(t)$$

$$\frac{d}{dt}\left(\vec{u}\cdot\vec{v}\right) = \vec{u}'\cdot\vec{v} + \vec{u}\cdot\vec{v}'$$

$$\frac{d}{dt}\left(\vec{u}\times\vec{v}\right) = \vec{u}'\times\vec{v} + \vec{u}\times\vec{v}'$$

$$\frac{d}{dt}\left(\vec{u}(f(t))\right) = f'(t)\,\vec{u}'\left(f(t)\right)$$

There is also one quick definition that we should get out of the way so that we can use it when we need to.

A **smooth curve** is any curve for which $\vec{r}'(t)$ is continuous and $\vec{r}'(t) \neq 0$ for any t except possibly at the endpoints. A helix is a smooth curve, for example.

Finally, we need to discuss integrals of vector functions. Using both limits and derivatives as a guide it shouldn't be too surprising that we also have the following for integration for indefinite integrals

Vector Function Indefinite Integral

$$\int \vec{r}(t)\, dt = \left\langle \int f(t)\, dt, \int g(t)\, dt, \int h(t)\, dt \right\rangle + \vec{c}$$
$$\int \vec{r}(t)\, dt = \int f(t)\, dt\, \vec{i} + \int g(t)\, dt\, \vec{j} + \int h(t)\, dt\, \vec{k} + \vec{c}$$

and the following for definite integrals.

Vector Function Definite Integral

$$\int_a^b \vec{r}(t)\, dt = \left\langle \int_a^b f(t)\, dt, \int_a^b g(t)\, dt, \int_a^b h(t)\, dt \right\rangle$$
$$\int_a^b \vec{r}(t)\, dt = \int_a^b f(t)\, dt\, \vec{i} + \int_a^b g(t)\, dt\, \vec{j} + \int_a^b h(t)\, dt\, \vec{k}$$

With the indefinite integrals we put in a constant of integration to make sure that it was clear that the constant in this case needs to be a vector instead of a regular constant.

Also, for the definite integrals we will sometimes write it as follows,

$$\int_a^b \vec{r}(t)\, dt = \left(\left\langle \int f(t)\, dt, \int g(t)\, dt, \int h(t)\, dt \right\rangle\right)\Bigg|_a^b$$
$$\int_a^b \vec{r}(t)\, dt = \left(\int f(t)\, dt\, \vec{i} + \int g(t)\, dt\, \vec{j} + \int h(t)\, dt\, \vec{k}\right)\Bigg|_a^b$$

In other words, we will do the indefinite integral and then do the evaluation of the vector as a whole instead of on a component by component basis.

Example 3

Compute $\int \vec{r}(t)\, dt$ for $\vec{r}(t) = \langle \sin(t), 6, 4t \rangle$.

Solution

All we need to do is integrate each of the components and be done with it.

$$\int \vec{r}(t)\, dt = \left\langle -\cos(t), 6t, 2t^2 \right\rangle + \vec{c}$$

Example 4

Compute $\displaystyle\int_0^1 \vec{r}(t)\,dt$ for $\vec{r}(t) = \langle \sin(t), 6, 4t \rangle$.

Solution

In this case all that we need to do is reuse the result from the previous example and then do the evaluation.

$$\begin{aligned}\int_0^1 \vec{r}(t)\,dt &= \left(\left\langle -\cos(t), 6t, 2t^2 \right\rangle\right)\Big|_0^1 \\ &= \langle -\cos(1), 6, 2 \rangle - \langle -1, 0, 0 \rangle \\ &= \langle 1 - \cos(1), 6, 2 \rangle\end{aligned}$$

12.8 Tangent, Normal and Binormal Vectors

In this section we want to look at an application of derivatives for vector functions. Actually, there are a couple of applications, but they all come back to needing the first one.

In the past we've used the fact that the derivative of a function was the slope of the tangent line. With vector functions we get exactly the same result, with one exception.

Given the vector function, $\vec{r}(t)$, we call $\vec{r}'(t)$ the **tangent vector** provided it exists and provided $\vec{r}'(t) \neq \vec{0}$. The tangent line to $\vec{r}(t)$ at P is then the line that passes through the point P and is parallel to the tangent vector, $\vec{r}'(t)$. Note that we really do need to require $\vec{r}'(t) \neq \vec{0}$ in order to have a tangent vector. If we had

$$\vec{r}'(t) = \vec{0}$$

we would have a vector that had no magnitude and so couldn't give us the direction of the tangent.

Also, provided $\vec{r}'(t) \neq \vec{0}$, the **unit tangent vector** to the curve is given by,

Unit Tangent Vector

$$\vec{T}(t) = \frac{\vec{r}'(t)}{\|\vec{r}'(t)\|}$$

While, the components of the unit tangent vector can be somewhat messy on occasion there are times when we will need to use the unit tangent vector instead of the tangent vector.

Example 1

Find the general formula for the tangent vector and unit tangent vector to the curve given by $\vec{r}(t) = t^2\,\vec{i} + 2\sin(t)\,\vec{j} + 2\cos(t)\,\vec{k}$.

Solution

First, by general formula we mean that we won't be plugging in a specific t and so we will be finding a formula that we can use at a later date if we'd like to find the tangent at any point on the curve. With that said there really isn't all that much to do at this point other than to do the work.

Here is the tangent vector to the curve.

$$\vec{r}'(t) = 2t\,\vec{i} + 2\cos(t)\,\vec{j} - 2\sin(t)\,\vec{k}$$

To get the unit tangent vector we need the length of the tangent vector.

$$\begin{aligned}\left\|\vec{r}\,'(t)\right\| &= \sqrt{4t^2 + 4\cos^2(t) + 4\sin^2(t)} \\ &= \sqrt{4t^2+4}\end{aligned}$$

The unit tangent vector is then,

$$\begin{aligned}\vec{T}(t) &= \frac{1}{\sqrt{4t^2+4}}\left(2t\,\vec{i} + 2\cos\,t\,\vec{j} - 2\sin(t)\,\vec{k}\right) \\ &= \frac{2t}{\sqrt{4t^2+4}}\,\vec{i} + \frac{2\cos(t)}{\sqrt{4t^2+4}}\vec{j} - \frac{2\sin(t)}{\sqrt{4t^2+4}}\,\vec{k}\end{aligned}$$

Example 2

Find the vector equation of the tangent line to the curve given by $\vec{r}(t) = t^2\,\vec{i} + 2\sin(t)\,\vec{j} + 2\cos(t)\,\vec{k}$ at $t = \frac{\pi}{3}$.

Solution

First, we need the tangent vector and since this is the function we were working with in the previous example we can just reuse the tangent vector from that example and plug in $t = \frac{\pi}{3}$.

$$\vec{r}\,'\left(\frac{\pi}{3}\right) = \frac{2\pi}{3}\,\vec{i} + 2\cos\left(\frac{\pi}{3}\right)\,\vec{j} - 2\sin\left(\frac{\pi}{3}\right)\,\vec{k} = \frac{2\pi}{3}\,\vec{i} + \vec{j} - \sqrt{3}\,\vec{k}$$

We'll also need the point on the line at $t = \frac{\pi}{3}$ so,

$$\vec{r}\left(\frac{\pi}{3}\right) = \frac{\pi^2}{9}\,\vec{i} + \sqrt{3}\,\vec{j} + \vec{k}$$

The vector equation of the line is then,

$$\vec{r}(t) = \left\langle \frac{\pi^2}{9}, \sqrt{3}, 1\right\rangle + t\left\langle \frac{2\pi}{3}, 1, -\sqrt{3}\right\rangle$$

Before moving on let's note a couple of things about the previous example. First, we could have used the unit tangent vector had we wanted to for the parallel vector. However, that would have made for a more complicated equation for the tangent line.

Second, notice that we used $\vec{r}(t)$ to represent the tangent line despite the fact that we used that as well for the function. Do not get excited about that. The $\vec{r}(t)$ here is much like y is with normal functions. With normal functions, y is the generic letter that we used to represent functions and

$\vec{r}\,(t)$ tends to be used in the same way with vector functions.

Next, we need to talk about the **unit normal** and the **binormal** vectors.

The unit normal vector is defined to be,

Unit Normal Vector

$$\vec{N}\,(t) = \frac{\vec{T}'\,(t)}{\left\|\vec{T}'\,(t)\right\|}$$

The unit normal is orthogonal (or normal, or perpendicular) to the unit tangent vector and hence to the curve as well. We've already seen normal vectors when we were dealing with Equations of Planes. They will show up with some regularity in several Calculus III topics.

The definition of the unit normal vector always seems a little mysterious when you first see it. It follows directly from the following fact.

Fact

Suppose that $\vec{r}\,(t)$ is a vector such that $\|\vec{r}\,(t)\| = c$ for all t. Then $\vec{r}\,'\,(t)$ is orthogonal to $\vec{r}\,(t)$.

Proof

To prove this fact is pretty simple. From the fact statement and the relationship between the magnitude of a vector and the dot product we have the following.

$$\vec{r}\,(t) \cdot \vec{r}\,(t) = \|\vec{r}\,(t)\|^2 = c^2 \qquad \text{for all } t$$

Now, because this is true for all t we can see that,

$$\frac{d}{dt}\left(\vec{r}\,(t) \cdot \vec{r}\,(t)\right) = \frac{d}{dt}\left(c^2\right) = 0$$

Also, recalling the fact from the previous section about differentiating a dot product we see that,

$$\frac{d}{dt}\left(\vec{r}\,(t) \cdot \vec{r}\,(t)\right) = \vec{r}\,'\,(t) \cdot \vec{r}\,(t) + \vec{r}\,(t) \cdot \vec{r}\,'\,(t) = 2\vec{r}\,'\,(t) \cdot \vec{r}\,(t)$$

Or, upon putting all this together we get,

$$2\vec{r}\,'\,(t) \cdot \vec{r}\,(t) = 0 \qquad \Rightarrow \qquad \vec{r}\,'\,(t) \cdot \vec{r}\,(t) = 0$$

Therefore $\vec{r}\,'\,(t)$ is orthogonal to $\vec{r}\,(t)$.

The definition of the unit normal then falls directly from this. Because $\vec{T}(t)$ is a unit vector we know that $\left\|\vec{T}(t)\right\| = 1$ for all t and hence by the Fact $\vec{T}'(t)$ is orthogonal to $\vec{T}(t)$. However, because $\vec{T}(t)$ is tangent to the curve, $\vec{T}'(t)$ must be orthogonal, or normal, to the curve as well and so be a normal vector for the curve. All we need to do then is divide by $\left\|\vec{T}'(t)\right\|$ to arrive at a unit normal vector.

Next, is the binormal vector. The binormal vector is defined to be,

Binormal Vector

$$\vec{B}(t) = \vec{T}(t) \times \vec{N}(t)$$

Because the binormal vector is defined to be the cross product of the unit tangent and unit normal vector we then know that the binormal vector is orthogonal to both the tangent vector and the normal vector.

Example 3

Find the normal and binormal vectors for $\vec{r}(t) = \langle t, 3\sin(t), 3\cos(t)\rangle$.

Solution

We first need the unit tangent vector so first get the tangent vector and its magnitude.

$$\vec{r}'(t) = \langle 1, 3\cos(t), -3\sin(t)\rangle$$
$$\left\|\vec{r}'(t)\right\| = \sqrt{1 + 9\cos^2(t) + 9\sin^2(t)} = \sqrt{10}$$

The unit tangent vector is then,

$$\vec{T}(t) = \left\langle \frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\cos(t), -\frac{3}{\sqrt{10}}\sin(t)\right\rangle$$

The unit normal vector will now require the derivative of the unit tangent and its magnitude.

$$\vec{T}'(t) = \left\langle 0, -\frac{3}{\sqrt{10}}\sin(t), -\frac{3}{\sqrt{10}}\cos(t)\right\rangle$$
$$\left\|\vec{T}'(t)\right\| = \sqrt{\frac{9}{10}\sin^2(t) + \frac{9}{10}\cos^2(t)} = \sqrt{\frac{9}{10}} = \frac{3}{\sqrt{10}}$$

The unit normal vector is then,

$$\vec{N}(t) = \frac{\sqrt{10}}{3}\left\langle 0, -\frac{3}{\sqrt{10}}\sin(t), -\frac{3}{\sqrt{10}}\cos(t)\right\rangle = \langle 0, -\sin(t), -\cos(t)\rangle$$

Finally, the binormal vector is,

$$\begin{aligned}
\vec{B}(t) &= \vec{T}(t) \times \vec{N}(t) \\
&= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}}\cos(t) & -\frac{3}{\sqrt{10}}\sin(t) \\ 0 & -\sin(t) & -\cos(t) \end{vmatrix} \\
&= -\frac{3}{\sqrt{10}}\cos^2(t)\,\vec{i} - \frac{1}{\sqrt{10}}\sin(t)\,\vec{k} + \frac{1}{\sqrt{10}}\cos(t)\,\vec{j} - \frac{3}{\sqrt{10}}\sin^2(t)\,\vec{i} \\
&= -\frac{3}{\sqrt{10}}\,\vec{i} + \frac{1}{\sqrt{10}}\cos(t)\,\vec{j} - \frac{1}{\sqrt{10}}\sin(t)\,\vec{k}
\end{aligned}$$

12.9 Arc Length with Vector Functions

In this section we'll recast an old formula into terms of vector functions. We want to determine the length of a vector function,

$$\vec{r}(t) = \langle f(t), g(t), h(t) \rangle$$

on the interval $a \le t \le b$.

We actually already know how to do this. Recall that we can write the vector function into the parametric form,

$$x = f(t) \qquad y = g(t) \qquad z = h(t)$$

Also, recall that with two dimensional parametric curves the arc length is given by,

$$L = \int_a^b \sqrt{[f'(t)]^2 + [g'(t)]^2}\, dt$$

There is a natural extension of this to three dimensions. So, the length of the curve $\vec{r}(t)$ on the interval $a \le t \le b$ is,

$$L = \int_a^b \sqrt{[f'(t)]^2 + [g'(t)]^2 + [h'(t)]^2}\, dt$$

There is a nice simplification that we can make for this. Notice that the integrand (the function we're integrating) is nothing more than the magnitude of the tangent vector,

$$\|\vec{r}\,'(t)\| = \sqrt{[f'(t)]^2 + [g'(t)]^2 + [h'(t)]^2}$$

Therefore, the arc length can be written as,

Arc Length

$$L = \int_a^b \|\vec{r}\,'(t)\|\, dt$$

Let's work a quick example of this.

Example 1

Determine the length of the curve $\vec{r}(t) = \langle 2t, 3\sin(2t), 3\cos(2t) \rangle$ on the interval $0 \le t \le 2\pi$.

Solution

We will first need the tangent vector and its magnitude.

$$\vec r\,'(t) = \langle 2, 6\cos(2t), -6\sin(2t)\rangle$$
$$\|\vec r\,'(t)\| = \sqrt{4 + 36\cos^2(2t) + 36\sin^2(2t)} = \sqrt{4+36} = 2\sqrt{10}$$

The length is then,

$$\begin{aligned} L &= \int_a^b \|\vec r\,'(t)\|\,dt \\ &= \int_0^{2\pi} 2\sqrt{10}\,dt \\ &= 4\pi\sqrt{10} \end{aligned}$$

We need to take a quick look at another concept here. We define the **arc length function** as,

Arc Length Function

$$s(t) = \int_0^t \|\vec r\,'(u)\|\,du$$

Before we look at why this might be important let's work a quick example.

Example 2

Determine the arc length function for $\vec r(t) = \langle 2t, 3\sin(2t), 3\cos(2t)\rangle$.

Solution

From the previous example we know that,

$$\|\vec r\,'(t)\| = 2\sqrt{10}$$

The arc length function is then,

$$s(t) = \int_0^t 2\sqrt{10}\,du = \left(2\sqrt{10}\,u\right)\Big|_0^t = 2\sqrt{10}\,t$$

Okay, just why would we want to do this? Well let's take the result of the example above and solve it for t.

$$t = \frac{s}{2\sqrt{10}}$$

Now, taking this and plugging it into the original vector function and we can **reparametrize** the function into the form, $\vec{r}\left(t\left(s\right)\right)$. For our function this is,

$$\vec{r}\left(t\left(s\right)\right) = \left\langle \frac{s}{\sqrt{10}}, 3\sin\left(\frac{s}{\sqrt{10}}\right), 3\cos\left(\frac{s}{\sqrt{10}}\right)\right\rangle$$

So, why would we want to do this? Well with the reparameterization we can now tell where we are on the curve after we've traveled a distance of s along the curve. Note as well that we will start the measurement of distance from where we are at $t = 0$.

Example 3

Where on the curve $\vec{r}\left(t\right) = \left\langle 2t, 3\sin\left(2t\right), 3\cos\left(2t\right)\right\rangle$ are we after traveling for a distance of $\dfrac{\pi\sqrt{10}}{3}$?

Solution

To determine this we need the reparameterization, which we have from above.

$$\vec{r}\left(t\left(s\right)\right) = \left\langle \frac{s}{\sqrt{10}}, 3\sin\left(\frac{s}{\sqrt{10}}\right), 3\cos\left(\frac{s}{\sqrt{10}}\right)\right\rangle$$

Then, to determine where we are all that we need to do is plug in $s = \frac{\pi\sqrt{10}}{3}$ into this and we'll get our location.

$$\vec{r}\left(t\left(\frac{\pi\sqrt{10}}{3}\right)\right) = \left\langle \frac{\pi}{3}, 3\sin\left(\frac{\pi}{3}\right), 3\cos\left(\frac{\pi}{3}\right)\right\rangle = \left\langle \frac{\pi}{3}, \frac{3\sqrt{3}}{2}, \frac{3}{2}\right\rangle$$

So, after traveling a distance of $\frac{\pi\sqrt{10}}{3}$ along the curve we are at the point $\left(\frac{\pi}{3}, \frac{3\sqrt{3}}{2}, \frac{3}{2}\right)$.

12.10 Curvature

In this section we want to briefly discuss the **curvature** of a smooth curve (recall that for a smooth curve we require $\vec{r}'(t)$ is continuous and $\vec{r}'(t) \neq 0$). The curvature measures how fast a curve is changing direction at a given point.

There are several formulas for determining the curvature for a curve. The formal definition of curvature is,

$$\kappa = \left\| \frac{d\vec{T}}{ds} \right\|$$

where $\vec{T}$ is the unit tangent and s is the arc length. Recall that we saw in a previous section how to reparametrize a curve to get it into terms of the arc length.

In general the formal definition of the curvature is not easy to use so there are two alternate formulas that we can use. Here they are.

Curvature

$$\kappa = \frac{\left\| \vec{T}'(t) \right\|}{\left\| \vec{r}'(t) \right\|} \qquad \kappa = \frac{\left\| \vec{r}'(t) \times \vec{r}''(t) \right\|}{\left\| \vec{r}'(t) \right\|^3}$$

These may not be particularly easy to deal with either, but at least we don't need to reparametrize the unit tangent.

Example 1

Determine the curvature for $\vec{r}(t) = \langle t, 3\sin(t), 3\cos(t) \rangle$.

Solution

Back in the section when we introduced the tangent vector we computed the tangent and unit tangent vectors for this function. These were,

$$\vec{r}'(t) = \langle 1, 3\cos(t), -3\sin(t) \rangle$$
$$\vec{T}(t) = \left\langle \frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\cos(t), -\frac{3}{\sqrt{10}}\sin(t) \right\rangle$$

The derivative of the unit tangent is,

$$\vec{T}'(t) = \left\langle 0, -\frac{3}{\sqrt{10}}\sin(t), -\frac{3}{\sqrt{10}}\cos(t) \right\rangle$$

The magnitudes of the two vectors are,

$$\left\|\vec{r}'(t)\right\| = \sqrt{1 + 9\cos^2(t) + 9\sin^2(t)} = \sqrt{10}$$
$$\left\|\vec{T}'(t)\right\| = \sqrt{0 + \frac{9}{10}\sin^2(t) + \frac{9}{10}\cos^2(t)} = \sqrt{\frac{9}{10}} = \frac{3}{\sqrt{10}}$$

The curvature is then,

$$\kappa = \frac{\left\|\vec{T}'(t)\right\|}{\left\|\vec{r}'(t)\right\|} = \frac{{}^{3}/_{\sqrt{10}}}{\sqrt{10}} = \frac{3}{10}$$

In this case the curvature is constant. This means that the curve is changing direction at the same rate at every point along it. Recalling that this curve is a helix this result makes sense.

Example 2

Determine the curvature of $\vec{r}(t) = t^2\,\vec{i} + t\,\vec{k}$.

Solution

In this case the second form of the curvature would probably be easiest. Here are the first couple of derivatives.

$$\vec{r}'(t) = 2t\,\vec{i} + \vec{k} \qquad\qquad \vec{r}''(t) = 2\,\vec{i}$$

Next, we need the cross product.

$$\begin{aligned}\vec{r}'(t) \times \vec{r}''(t) &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 2t & 0 & 1 \\ 2 & 0 & 0 \end{vmatrix} \\ &= 2\vec{j}\end{aligned}$$

The magnitudes are,

$$\left\|\vec{r}'(t) \times \vec{r}''(t)\right\| = 2 \qquad\qquad \left\|\vec{r}'(t)\right\| = \sqrt{4t^2 + 1}$$

The curvature at any value of t is then,

$$\kappa = \frac{2}{\left(4t^2 + 1\right)^{\frac{3}{2}}}$$

There is a special case that we can look at here as well. Suppose that we have a curve given by

$y = f(x)$ and we want to find its curvature.

As we saw when we first looked at vector functions we can write this as follows,

$$\vec{r}(x) = x\,\vec{i} + f(x)\,\vec{j}$$

If we then use the second formula for the curvature we will arrive at the following formula for the curvature.

$$\kappa = \frac{|f''(x)|}{\left(1 + \left[f'(x)\right]^2\right)^{\frac{3}{2}}}$$

12.11 Velocity and Acceleration

In this section we need to take a look at the velocity and acceleration of a moving object.

From Calculus I we know that given the position function of an object that the velocity of the object is the first derivative of the position function and the acceleration of the object is the second derivative of the position function.

So, given this it shouldn't be too surprising that if the position function of an object is given by the vector function $\vec{r}(t)$ then the velocity and acceleration of the object is given by,

$$\vec{v}(t) = \vec{r}\,'(t) \qquad\qquad \vec{a}(t) = \vec{r}\,''(t)$$

Notice that the velocity and acceleration are also going to be vectors as well.

In the study of the motion of objects the acceleration is often broken up into a **tangential component**, a_T, and a **normal component**, a_N. The tangential component is the part of the acceleration that is tangential to the curve and the normal component is the part of the acceleration that is normal (or orthogonal) to the curve. If we do this we can write the acceleration as,

$$\vec{a} = a_T\vec{T} + a_N\vec{N}$$

where $\vec{T}$ and $\vec{N}$ are the unit tangent and unit normal for the position function.

If we define $v = \|\vec{v}(t)\|$ then the tangential and normal components of the acceleration are given by,

Tangential and Normal Acceleration

$$a_T = v' = \frac{\vec{r}\,'(t)\cdot\vec{r}\,''(t)}{\|r'(t)\|} \qquad\qquad a_N = \kappa v^2 = \frac{\|\vec{r}\,'(t)\times\vec{r}\,''(t)\|}{\|r'(t)\|}$$

where κ is the curvature for the position function.

There are two formulas to use here for each component of the acceleration and while the second formula may seem overly complicated it is often the easier of the two. In the tangential component, v, may be messy and computing the derivative may be unpleasant. In the normal component we will already be computing both of these quantities in order to get the curvature and so the second formula in this case is definitely the easier of the two.

Let's take a quick look at a couple of examples.

Example 1

If the acceleration of an object is given by $\vec{a} = \vec{i} + 2\vec{j} + 6t\vec{k}$ find the object's velocity and position functions given that the initial velocity is $\vec{v}(0) = \vec{j} - \vec{k}$ and the initial position is $\vec{r}(0) = \vec{i} - 2\vec{j} + 3\vec{k}$.

Solution

We'll first get the velocity. To do this all (well almost all) we need to do is integrate the acceleration.

$$\begin{aligned}\vec{v}(t) &= \int \vec{a}(t)\,dt\\ &= \int \vec{i} + 2\vec{j} + 6t\vec{k}\,dt\\ &= t\,\vec{i} + 2t\,\vec{j} + 3t^2\,\vec{k} + \vec{c}\end{aligned}$$

To completely get the velocity we will need to determine the "constant" of integration. We can use the initial velocity to get this.

$$\vec{j} - \vec{k} = \vec{v}(0) = \vec{c}$$

The velocity of the object is then,

$$\begin{aligned}\vec{v}(t) &= t\,\vec{i} + 2t\,\vec{j} + 3t^2\,\vec{k} + \vec{j} - \vec{k}\\ &= t\,\vec{i} + (2t+1)\,\vec{j} + \left(3t^2 - 1\right)\,\vec{k}\end{aligned}$$

We will find the position function by integrating the velocity function.

$$\begin{aligned}\vec{r}(t) &= \int \vec{v}(t)\,dt\\ &= \int t\,\vec{i} + (2t+1)\,\vec{j} + \left(3t^2 - 1\right)\,\vec{k}\,dt\\ &= \frac{1}{2}t^2\,\vec{i} + \left(t^2 + t\right)\vec{j} + \left(t^3 - t\right)\vec{k} + \vec{c}\end{aligned}$$

Using the initial position gives us,

$$\vec{i} - 2\vec{j} + 3\vec{k} = \vec{r}(0) = \vec{c}$$

So, the position function is,

$$\vec{r}(t) = \left(\frac{1}{2}t^2 + 1\right)\,\vec{i} + \left(t^2 + t - 2\right)\vec{j} + \left(t^3 - t + 3\right)\vec{k}$$

Example 2

For the object in the previous example determine the tangential and normal components of the acceleration.

Solution

There really isn't much to do here other than plug into the formulas. To do this we'll need to notice that,

$$\begin{aligned} \vec{r}'(t) &= t\,\vec{i} + (2t+1)\,\vec{j} + \left(3t^2-1\right)\,\vec{k} \\ \vec{r}''(t) &= \vec{i} + 2\vec{j} + 6t\vec{k} \end{aligned}$$

Let's first compute the dot product and cross product that we'll need for the formulas.

$$\vec{r}'(t)\cdot\vec{r}''(t) = t + 2(2t+1) + 6t\left(3t^2-1\right) = 18t^3 - t + 2$$

$$\begin{aligned} \vec{r}'(t)\times\vec{r}''(t) &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ t & 2t+1 & 3t^2-1 \\ 1 & 2 & 6t \end{vmatrix} \\ &= (6t)(2t+1)\vec{i} + \left(3t^2-1\right)\vec{j} + 2t\vec{k} - 6t^2\vec{j} - 2\left(3t^2-1\right)\vec{i} - (2t+1)\vec{k} \\ &= \left(6t^2+6t+2\right)\vec{i} - \left(3t^2+1\right)\vec{j} - \vec{k} \end{aligned}$$

Next, we also need a couple of magnitudes.

$$\begin{aligned} \left\|\vec{r}'(t)\right\| &= \sqrt{t^2 + (2t+1)^2 + \left(3t^2-1\right)^2} = \sqrt{9t^4 - t^2 + 4t + 2} \\ \left\|\vec{r}'(t)\times\vec{r}''(t)\right\| &= \sqrt{\left(6t^2+6t+2\right)^2 + \left(3t^2+1\right)^2 + 1} = \sqrt{45t^4 + 72t^3 + 66t^2 + 24t + 6} \end{aligned}$$

The tangential component of the acceleration is then,

$$a_T = \frac{18t^3 - t + 2}{\sqrt{9t^4 - t^2 + 4t + 2}}$$

The normal component of the acceleration is,

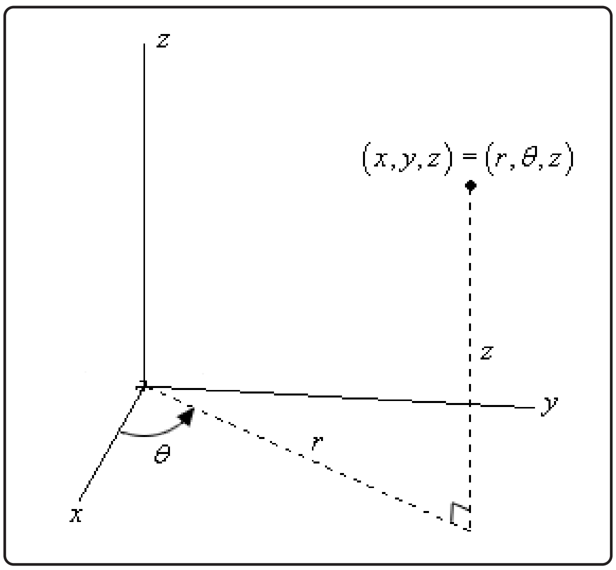
$$a_N = \frac{\sqrt{45t^4 + 72t^3 + 66t^2 + 24t + 6}}{\sqrt{9t^4 - t^2 + 4t + 2}} = \sqrt{\frac{45t^4 + 72t^3 + 66t^2 + 24t + 6}{9t^4 - t^2 + 4t + 2}}$$

12.12 Cylindrical Coordinates

As with two dimensional space the standard (x, y, z) coordinate system is called the Cartesian coordinate system. In the last two sections of this chapter we'll be looking at some alternate coordinate systems for three dimensional space.

We'll start off with the cylindrical coordinate system. This one is fairly simple as it is nothing more than an extension of polar coordinates into three dimensions. Not only is it an extension of polar coordinates, but we extend it into the third dimension just as we extend Cartesian coordinates into the third dimension. All that we do is add a z on as the third coordinate. The r and θ are the same as with polar coordinates.

Here is a sketch of a point in $\mathbb{R}^3$.



The conversions for x and y are the same conversions that we used back when we were looking at polar coordinates. So, if we have a point in cylindrical coordinates the Cartesian coordinates can be found by using the following conversions.

Cylindrical to Cartesian Converstion Formulas

$$x = r\cos(\theta)$$
$$y = r\sin(\theta)$$
$$z = z$$

The third equation is just an acknowledgement that the z-coordinate of a point in Cartesian and polar coordinates is the same.

Likewise, if we have a point in Cartesian coordinates the cylindrical coordinates can be found by using the following conversions.

Cartesian to Cylindrical Converstion Formulas

$$r = \sqrt{x^2 + y^2} \qquad \text{OR} \qquad r^2 = x^2 + y^2$$
$$\theta = \tan^{-1}\left(\frac{y}{x}\right)$$
$$z = z$$

Let's take a quick look at some surfaces in cylindrical coordinates.

Example 1

Identify the surface for each of the following equations.

(a) $r = 5$

(b) $r^2 + z^2 = 100$

(c) $z = r$

Solution

(a) $r = 5$

In two dimensions we know that this is a circle of radius 5. Since we are now in three dimensions and there is no z in equation this means it is allowed to vary freely. So, for any given z we will have a circle of radius 5 centered on the z-axis.

In other words, we will have a cylinder of radius 5 centered on the z-axis.

(b) $r^2 + z^2 = 100$

This equation will be easy to identify once we convert back to Cartesian coordinates.

$$\begin{aligned} r^2 + z^2 &= 100 \\ x^2 + y^2 + z^2 &= 100 \end{aligned}$$

So, this is a sphere centered at the origin with radius 10.

(c) $z = r$

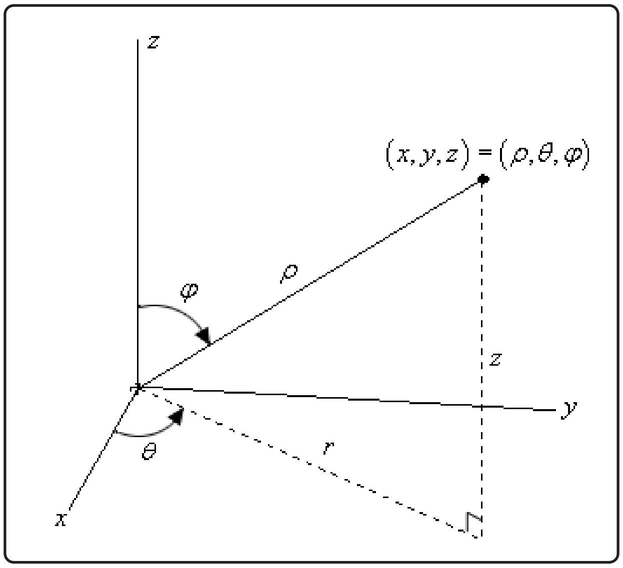
Again, this one won't be too bad if we convert back to Cartesian. For reasons that will be apparent eventually, we'll first square both sides, then convert.

$$\begin{aligned} z^2 &= r^2 \\ z^2 &= x^2 + y^2 \end{aligned}$$

From the section on quadric surfaces we know that this is the equation of a cone.

12.13 Spherical Coordinates

In this section we will introduce spherical coordinates. Spherical coordinates can take a little getting used to. It's probably easiest to start things off with a sketch.



Spherical coordinates consist of the following three quantities.

First there is ρ. This is the distance from the origin to the point and we will require $\rho \geq 0$.

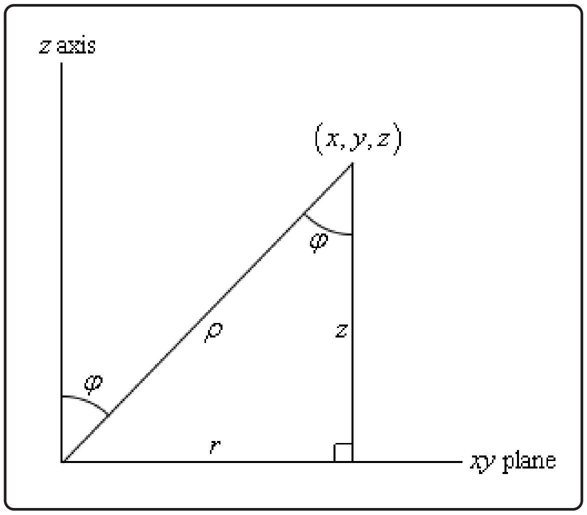
Next there is θ. This is the same angle that we saw in polar/cylindrical coordinates. It is the angle between the positive x-axis and the line above denoted by r (which is also the same r as in polar/cylindrical coordinates). There are no restrictions on θ.

Finally, there is φ. This is the angle between the positive z-axis and the line from the origin to the point. We will require $0 \leq \varphi \leq \pi$.

In summary, ρ is the distance from the origin to the point, φ is the angle that we need to rotate down from the positive *z*-axis to get to the point and θ is how much we need to rotate around the z-axis to get to the point.

We should first derive some conversion formulas. Let's first start with a point in spherical coordinates and ask what the cylindrical coordinates of the point are. So, we know (ρ, θ, φ) and want to find (r, θ, z). Of course, we really only need to find r and z since θ is the same in both coordinate systems.

If we look at the sketch above from directly in front of the triangle we get the following sketch,



We know that the angle between the z-axis and ρ is φ and with a little geometry we also know that the angle between ρ and the vertical side of the right triangle is also φ.

Then, with a little right triangle trig we get,

$$z = \rho\cos(\varphi)$$
$$r = \rho\sin(\varphi)$$

and these are exactly the formulas that we were looking for. So, given a point in spherical coordinates the cylindrical coordinates of the point will be,

Spherical to Cylindrical Conversion Formulas

$$r = \rho\sin(\varphi)$$
$$\theta = \theta$$
$$z = \rho\cos(\varphi)$$

Note as well from the Pythagorean theorem we also get,

Fact

$$\rho^2 = r^2 + z^2$$

Next, let's find the Cartesian coordinates of the same point. To do this we'll start with the cylindrical conversion formulas from the previous section.

$$\begin{aligned} x &= r\cos(\theta) \\ y &= r\sin(\theta) \\ z &= z \end{aligned}$$

Now all that we need to do is use the formulas from above for r and z to get,

Spherical to Cartesian Conversion Formulas

$$\begin{aligned} x &= \rho\sin(\varphi)\cos(\theta) \\ y &= \rho\sin(\varphi)\sin(\theta) \\ z &= \rho\cos(\varphi) \end{aligned}$$

Also note that since we know that $r^2 = x^2 + y^2$ we get,

Fact

$$\rho^2 = x^2 + y^2 + z^2$$

Converting points from Cartesian or cylindrical coordinates into spherical coordinates is usually done with the same conversion formulas. To see how this is done let's work an example of each.

Example 1

Perform each of the following conversions.

(a) Convert the point $\left(\sqrt{6}, \frac{\pi}{4}, \sqrt{2}\right)$ from cylindrical to spherical coordinates.

(b) Convert the point $\left(-1, 1, -\sqrt{2}\right)$ from Cartesian to spherical coordinates.

Solution

(a) Convert the point $\left(\sqrt{6}, \frac{\pi}{4}, \sqrt{2}\right)$ from cylindrical to spherical coordinates.

We'll start by acknowledging that θ is the same in both coordinate systems and so we don't need to do anything with that.

Next, let's find ρ.

$$\rho = \sqrt{r^2 + z^2} = \sqrt{6+2} = \sqrt{8} = 2\sqrt{2}$$

Finally, let's get φ. To do this we can use either the conversion for r or z. We'll use the conversion for z.

$$z = \rho\cos(\varphi) \qquad \Rightarrow \qquad \cos(\varphi) = \frac{z}{\rho} = \frac{\sqrt{2}}{2\sqrt{2}} \qquad \Rightarrow \qquad \varphi = \cos^{-1}\left(\frac{1}{2}\right) = \frac{\pi}{3}$$

Notice that there are many possible values of φ that will give $\cos(\varphi) = \frac{1}{2}$, however, we have restricted φ to the range $0 \le \varphi \le \pi$ and so this is the only possible value in that range.

So, the spherical coordinates of this point will are $\left(2\sqrt{2}, \frac{\pi}{4}, \frac{\pi}{3}\right)$.

(b) Convert the point $\left(-1, 1, -\sqrt{2}\right)$ from Cartesian to spherical coordinates.

The first thing that we'll do here is find ρ.

$$\rho = \sqrt{x^2 + y^2 + z^2} = \sqrt{1+1+2} = 2$$

Now we'll need to find φ. We can do this using the conversion for z.

$$z = \rho\cos(\varphi) \qquad \Rightarrow \qquad \cos(\varphi) = \frac{z}{\rho} = \frac{-\sqrt{2}}{2} \qquad \Rightarrow \qquad \varphi = \cos^{-1}\left(\frac{-\sqrt{2}}{2}\right) = \frac{3\pi}{4}$$

As with the last parts this will be the only possible φ in the range allowed.

Finally, let's find θ. To do this we can use the conversion for x or y. We will use the conversion for y in this case.

$$\sin(\theta) = \frac{y}{\rho\sin(\varphi)} = \frac{1}{2\left(\frac{\sqrt{2}}{2}\right)} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2} \qquad \Rightarrow \qquad \theta = \frac{\pi}{4} \text{ or } \theta = \frac{3\pi}{4}$$

Now, we actually have more possible choices for θ but all of them will reduce down to one of the two angles above since they will just be one of these two angles with one or more complete rotations around the unit circle added on.

We will however, need to decide which one is the correct angle since only one will be. To do this let's notice that, in two dimensions, the point with coordinates $x = -1$ and $y = 1$ lies in the second quadrant. This means that θ must be angle that will put the point into the second quadrant. Therefore, the second angle, $\theta = \frac{3\pi}{4}$, must be the correct one.

The spherical coordinates of this point are then $\left(2, \frac{3\pi}{4}, \frac{3\pi}{4}\right)$.

Now, let's take a look at some equations and identify the surfaces that they represent.

Example 2

Identify the surface for each of the following equations.

(a) $\rho = 5$

(b) $\varphi = \dfrac{\pi}{3}$

(c) $\theta = \dfrac{2\pi}{3}$

(d) $\rho \sin(\varphi) = 2$

Solution

(a) $\rho = 5$

There are a couple of ways to think about this one.

First, think about what this equation is saying. This equation says that, no matter what θ and φ are, the distance from the origin must be 5. So, we can rotate as much as we want away from the z-axis and around the z-axis, but we must always remain at a fixed distance from the origin. This is exactly what a sphere is. So, this is a sphere of radius 5 centered at the origin.

The other way to think about it is to just convert to Cartesian coordinates.

$$\begin{aligned} \rho &= 5 \\ \rho^2 &= 25 \\ x^2 + y^2 + z^2 &= 25 \end{aligned}$$

Sure enough a sphere of radius 5 centered at the origin.

(b) $\varphi = \dfrac{\pi}{3}$

In this case there isn't an easy way to convert to Cartesian coordinates so we'll just need to think about this one a little. This equation says that no matter how far away from the origin that we move and no matter how much we rotate around the z-axis the point must always be at an angle of $\frac{\pi}{3}$ from the z-axis.

This is exactly what happens in a cone. All of the points on a cone are a fixed angle from the z-axis. So, we have a cone whose points are all at an angle of $\frac{\pi}{3}$ from the z-axis.

(c) $\theta = \dfrac{2\pi}{3}$

As with the last part we won't be able to easily convert to Cartesian coordinates here. In this case no matter how far from the origin we get or how much we rotate down from the positive z-axis the points must always form an angle of $\frac{2\pi}{3}$ with the x-axis.

Points in a vertical plane will do this. So, we have a vertical plane that forms an angle of $\frac{2\pi}{3}$ with the positive x-axis.

(d) $\rho \sin(\varphi) = 2$

In this case we can convert to Cartesian coordinates so let's do that. There are actually two ways to do this conversion. We will look at both since both will be used on occasion.

Solution 1
In this solution method we will convert directly to Cartesian coordinates. To do this we will first need to square both sides of the equation.

$$\rho^2 \sin^2(\varphi) = 4$$

Now, for no apparent reason add $\rho^2 \cos^2(\varphi)$ to both sides.

$$\begin{aligned} \rho^2 \sin^2(\varphi) + \rho^2 \cos^2(\varphi) &= 4 + \rho^2 \cos^2(\varphi) \\ \rho^2 \left(\sin^2(\varphi) + \cos^2(\varphi)\right) &= 4 + \rho^2 \cos^2(\varphi) \\ \rho^2 &= 4 + (\rho \cos(\varphi))^2 \end{aligned}$$

Now we can convert to Cartesian coordinates.

$$\begin{aligned} x^2 + y^2 + z^2 &= 4 + z^2 \\ x^2 + y^2 &= 4 \end{aligned}$$

So, we have a cylinder of radius 2 centered on the z-axis.

This solution method wasn't too bad, but it did require some not so obvious steps to complete.

Solution 2
This method is much shorter, but also involves something that you may not see the first time around. In this case instead of going straight to Cartesian coordinates we'll first convert to cylindrical coordinates.

This won't always work, but in this case all we need to do is recognize that $r = \rho \sin(\varphi)$ and we will get something we can recognize. Using this we get,

$$\begin{aligned} \rho \sin(\varphi) &= 2 \\ r &= 2 \end{aligned}$$

At this point we know this is a cylinder (remember that we're in three dimensions and so this isn't a circle!). However, let's go ahead and finish the conversion process out.

$$\begin{aligned} r^2 &= 4 \\ x^2 + y^2 &= 4 \end{aligned}$$

So, as we saw in the last part of the previous example it will *sometimes* be easier to convert equations in spherical coordinates into cylindrical coordinates before converting into Cartesian coordinates. This won't always be easier, but it can make some of the conversions quicker and easier.

The last thing that we want to do in this section is generalize the first three parts of the previous example.

$\rho = a$	sphere of radius a centered at the origin
$\varphi = \alpha$	cone that makes an angle of α with the positive $z-$ axis
$\theta = \beta$	vertical plane that makes an angle of β with the positive $x-$ axis

13 Partial Derivatives

To this point, with the exception of the occasional section in the last chapter, we've been working almost exclusively with functions of a single variable. It is now time to formally start multi-variable Calculus, *i.e.* Calculus involving functions of two or more variables. We will be covering the same basic topics as we do with single variable Calculus. Namely, limits, derivatives and integrals.

In this chapter we will open up with a quick section discussing taking limits of multi-variable functions. We will only be covering limits of multi-variable functions with a single chapter because, as we'll see, many of the concepts from single variable limits still hold, with some natural extensions of course. However, as we'll also see the work will often be significantly longer/harder and so we won't be spending a lot of time discussing limits of multi-variable functions. Luckily enough for us we also won't need to worry all that much about limits of multi-variable functions so the quick discussion of limits in this chapter will suffice.

The rest of the chapter will be discussing how to take derivatives of multi-variable functions. We want to keep the "main" interpretation of derivatives, namely the derivative will still give the rate of change of the function. The issue here is that because we have multiple variables the function can have differing rates of change depending on how we allow the various variables to change.

So, to start out the derivative discussion we will start by defining the partial derivative. These will restrict just how we allow the various variables to change. We will eventually introduce the directional derivative which will allow the variables to change in any arbitrary manner. In the process of introducing the idea of a directional derivative we'll also introduce the concept of a gradient of a function. The gradient will arise in quite a few sections throughout the rest of this multi-variable Calculus material, including integrals.

Finally, as we'll see, if you can take derivatives of single variable functions then you have the majority of the knowledge that you need to take derivatives of multi-variable functions. There are, however, some subtleties that we'll need to remember to deal with. Those subtleties are, generally, the issues that most students run into when taking derivatives of multi-variable functions.

13.1 Limits

In this section we will take a look at limits involving functions of more than one variable. In fact, we will concentrate mostly on limits of functions of two variables, but the ideas can be extended out to functions with more than two variables.

Before getting into this let's briefly recall how limits of functions of one variable work. We say that,

$$\lim_{x \to a} f(x) = L$$

provided,

$$\lim_{x \to a^+} f(x) = \lim_{x \to a^-} f(x) = L$$

Also, recall that,

$$\lim_{x \to a^+} f(x)$$

is a right hand limit and requires us to only look at values of x that are greater than a. Likewise,

$$\lim_{x \to a^-} f(x)$$

is a left hand limit and requires us to only look at values of x that are less than a.

In other words, we will have $\lim\limits_{x \to a} f(x) = L$ provided $f(x)$ approaches L as we move in towards $x = a$ (without letting $x = a$) from both sides.

Now, notice that in this case there are only two paths that we can take as we move in towards $x = a$. We can either move in from the left or we can move in from the right. Then in order for the limit of a function of one variable to exist the function must be approaching the same value as we take each of these paths in towards $x = a$.

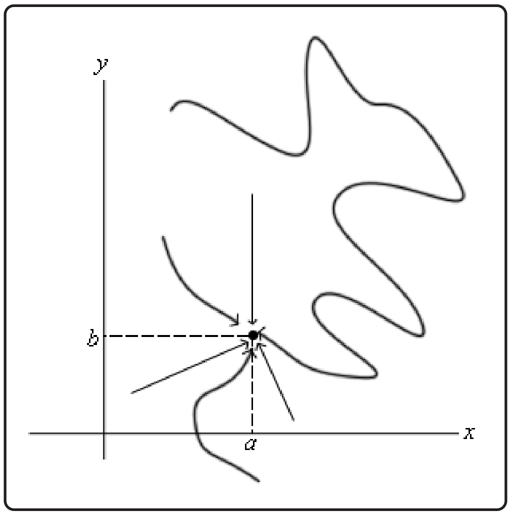
With functions of two variables we will have to do something similar, except this time there is (potentially) going to be a lot more work involved. Let's first address the notation and get a feel for just what we're going to be asking for in these kinds of limits.

We will be asking to take the limit of the function $f(x, y)$ as x approaches a and as y approaches b. This can be written in several ways. Here are a couple of the more standard notations.

$$\lim_{\substack{x \to a \\ y \to b}} f(x, y) \qquad\qquad \lim_{(x,y) \to (a,b)} f(x, y)$$

We will use the second notation more often than not in this course. The second notation is also a little more helpful in illustrating what we are really doing here when we are taking a limit. In taking a limit of a function of two variables we are really asking what the value of $f(x, y)$ is doing as we move the point (x, y) in closer and closer to the point (a, b) without actually letting it be (a, b).

Just like with limits of functions of one variable, in order for this limit to exist, the function must be approaching the same value regardless of the path that we take as we move in towards (a, b). The problem that we are immediately faced with is that there are literally an infinite number of paths that we can take as we move in towards (a, b). Here are a few examples of paths that we could take.



We put in a couple of straight line paths as well as a couple of "stranger" paths that aren't straight line paths. Also, we only included 6 paths here and as you can see simply by varying the slope of the straight line paths there are an infinite number of these and then we would need to consider paths that aren't straight line paths.

In other words, to show that a limit exists we would technically need to check an infinite number of paths and verify that the function is approaching the same value regardless of the path we are using to approach the point.

Luckily for us however we can use one of the main ideas from Calculus I limits to help us take limits here.

Definition

A function $f(x, y)$ is **continuous** at the point (a, b) if,

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b)$$

From a graphical standpoint this definition means the same thing as it did when we first saw continuity in Calculus I. A function will be continuous at a point if the graph doesn't have any holes or breaks at that point.

How can this help us take limits? Well, just as in Calculus I, if you know that a function is continuous at (a, b) then you also know that

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b)$$

must be true. So, if we know that a function is continuous at a point then all we need to do to take the limit of the function at that point is to plug the point into the function.

All the standard functions that we know to be continuous are still continuous even if we are plugging in more than one variable now. We just need to watch out for division by zero, square roots of negative numbers, logarithms of zero or negative numbers, *etc.*

Note that the idea about paths is one that we shouldn't forget since it is a nice way to determine if a limit doesn't exist. If we can find two paths upon which the function approaches different values as we get near the point then we will know that the limit doesn't exist.

Let's take a look at a couple of examples.

Example 1

Determine if the following limits exist or not. If they do exist give the value of the limit.

(a) $\displaystyle\lim_{(x,y,z)\to(2,1,-1)}\left(3x^2z + yx\cos\left(\pi x - \pi z\right)\right)$

(b) $\displaystyle\lim_{(x,y)\to(5,1)}\frac{xy}{x+y}$

Solution

(a) $\displaystyle\lim_{(x,y,z)\to(2,1,-1)}\left(3x^2z + yx\cos\left(\pi x - \pi z\right)\right)$

Okay, in this case the function is continuous at the point in question and so all we need to do is plug in the values and we're done.

$$\lim_{(x,y,z)\to(2,1,-1)}\left(3x^2z + yx\cos\left(\pi x - \pi z\right)\right) = 3(2)^2\left(-1\right) + \left(1\right)\left(2\right)\cos\left(2\pi + \pi\right) = -14$$

(b) $\displaystyle\lim_{(x,y)\to(5,1)}\frac{xy}{x+y}$

In this case the function will not be continuous along the line $y = -x$ since we will get division by zero when this is true. However, for this problem that is not something that we will need to worry about since the point that we are taking the limit at isn't on this line.

Therefore, all that we need to do is plug in the point since the function is continuous at this point.

$$\lim_{(x,y)\to(5,1)}\frac{xy}{x+y} = \frac{5}{6}$$

In the previous example there wasn't really anything to the limits. The functions were continuous at the point in question and so all we had to do was plug in the point. That, of course, will not always be the case so let's work a few examples that are more typical of those you'll see here.

Example 2

Determine if the following limit exist or not. If they do exist give the value of the limit.

$$\lim_{(x,y)\to(1,1)} \frac{2x^2 - xy - y^2}{x^2 - y^2}$$

Solution

In this case the function is not continuous at the point in question (clearly division by zero). However, that does not mean that the limit can't be done. We saw many examples of this in Calculus I where the function was not continuous at the point we were looking at and yet the limit did exist.

In the case of this limit notice that we can factor both the numerator and denominator of the function as follows,

$$\lim_{(x,y)\to(1,1)} \frac{2x^2 - xy - y^2}{x^2 - y^2} = \lim_{(x,y)\to(1,1)} \frac{(2x+y)(x-y)}{(x-y)(x+y)} = \lim_{(x,y)\to(1,1)} \frac{2x+y}{x+y}$$

So, just as we saw in many examples in Calculus I, upon factoring and canceling common factors we arrive at a function that in fact we can take the limit of. So, to finish out this example all we need to do is actually take the limit.

Taking the limit gives,

$$\lim_{(x,y)\to(1,1)} \frac{2x^2 - xy - y^2}{x^2 - y^2} = \lim_{(x,y)\to(1,1)} \frac{2x+y}{x+y} = \frac{3}{2}$$

Before we move on to the next set of examples we should note that the situation in the previous example is what generally happened in many limit examples/problems in Calculus I. In Calculus III however, this tends to be the exception in the examples/problems as the next set of examples will show. In other words, do not expect most of these types of limits to just factor and then exist as they did in Calculus I.

Example 3

Determine if the following limits exist or not. If they do exist give the value of the limit.

(a) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{x^4+3y^4}$

(b) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^3y}{x^6+y^2}$

Solution

(a) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{x^4+3y^4}$

In this case the function is not continuous at the point in question and so we can't just plug in the point. Also, note that, unlike the previous example, we can't factor this function and do some canceling so that the limit can be taken.

Therefore, since the function is not continuous at the point and because there is no factoring we can do, there is at least a chance that the limit doesn't exist. If we could find two different paths to approach the point that gave different values for the limit then we would know that the limit didn't exist. Two of the more common paths to check are the x and y-axis so let's try those.

Before actually doing this we need to address just what exactly do we mean when we say that we are going to approach a point along a path. When we approach a point along a path we will do this by either fixing x or y or by relating x and y through some function. In this way we can reduce the limit to just a limit involving a single variable which we know how to do from Calculus I.

So, let's see what happens along the x-axis. If we are going to approach $(0,0)$ along the x-axis we can take advantage of the fact that that along the x-axis we know that $y=0$. This means that, along the x-axis, we will plug in $y=0$ into the function and then take the limit as x approaches zero.

$$\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{x^4+3y^4} = \lim_{(x,0)\to(0,0)} \frac{x^2(0)^2}{x^4+3(0)^4} = \lim_{(x,0)\to(0,0)} 0 = 0$$

So, along the x-axis the function will approach zero as we move in towards the origin.

Now, let's try the y-axis. Along this axis we have $x=0$ and so the limit becomes,

$$\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{x^4+3y^4} = \lim_{(0,y)\to(0,0)} \frac{(0)^2y^2}{(0)^4+3y^4} = \lim_{(0,y)\to(0,0)} 0 = 0$$

So, the same limit along two paths. Don't misread this. This does NOT say that the limit exists and has a value of zero. This only means that the limit happens to have the same value along two paths.

Let's take a look at a third fairly common path to take a look at. In this case we'll move in towards the origin along the path $y = x$. This is what we meant previously about relating x and y through a function.

To do this we will replace all the y's with x's and then let x approach zero. Let's take a look at this limit.

$$\lim_{(x,y)\to(0,0)} \frac{x^2y^2}{x^4+3y^4} = \lim_{(x,x)\to(0,0)} \frac{x^2x^2}{x^4+3x^4} = \lim_{(x,x)\to(0,0)} \frac{x^4}{4x^4} = \lim_{(x,x)\to(0,0)} \frac{1}{4} = \frac{1}{4}$$

So, a different value from the previous two paths and this means that the limit can't possibly exist.

Note that we can use this idea of moving in towards the origin along a line with the more general path $y = mx$ if we need to.

(b) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^3y}{x^6+y^2}$

Okay, with this last one we again have continuity problems at the origin and again there is no factoring we can do that will allow the limit to be taken. So, again let's see if we can find a couple of paths that give different values of the limit.

First, we will use the path $y = x$. Along this path we have,

$$\lim_{(x,y)\to(0,0)} \frac{x^3y}{x^6+y^2} = \lim_{(x,x)\to(0,0)} \frac{x^3x}{x^6+x^2} = \lim_{(x,x)\to(0,0)} \frac{x^4}{x^6+x^2} = \lim_{(x,x)\to(0,0)} \frac{x^2}{x^4+1} = 0$$

Now, let's try the path $y = x^3$. Along this path the limit becomes,

$$\lim_{(x,y)\to(0,0)} \frac{x^3y}{x^6+y^2} = \lim_{(x,x^3)\to(0,0)} \frac{x^3x^3}{x^6+(x^3)^2} = \lim_{(x,x^3)\to(0,0)} \frac{x^6}{2x^6} = \lim_{(x,x^3)\to(0,0)} \frac{1}{2} = \frac{1}{2}$$

We now have two paths that give different values for the limit and so the limit doesn't exist.

As this limit has shown us we can, and often need, to use paths other than lines like we did in the first part of this example.

So, as we've seen in the previous example limits are a little different here from those we saw in Calculus I. Limits in multiple variables can be quite difficult to evaluate and we've shown several examples where it took a little work just to show that the limit does not exist.

13.2 Partial Derivatives

Now that we have the brief discussion on limits out of the way we can proceed into taking derivatives of functions of more than one variable. Before we actually start taking derivatives of functions of more than one variable let's recall an important interpretation of derivatives of functions of one variable.

Recall that given a function of one variable, $f(x)$, the derivative, $f'(x)$, represents the rate of change of the function as x changes. This is an important interpretation of derivatives and we are not going to want to lose it with functions of more than one variable. The problem with functions of more than one variable is that there is more than one variable. In other words, what do we do if we only want one of the variables to change, or if we want more than one of them to change? In fact, if we're going to allow more than one of the variables to change there are then going to be an infinite amount of ways for them to change. For instance, one variable could be changing faster than the other variable(s) in the function. Notice as well that it will be completely possible for the function to be changing differently depending on how we allow one or more of the variables to change.

We will need to develop ways, and notations, for dealing with all of these cases. In this section we are going to concentrate exclusively on only changing one of the variables at a time, while the remaining variable(s) are held fixed. We will deal with allowing multiple variables to change in a later section.

Because we are going to only allow one of the variables to change taking the derivative will now become a fairly simple process. Let's start off this discussion with a fairly simple function.

Let's start with the function $f(x,y) = 2x^2y^3$ and let's determine the rate at which the function is changing at a point, (a,b), if we hold y fixed and allow x to vary and if we hold x fixed and allow y to vary.

We'll start by looking at the case of holding y fixed and allowing x to vary. Since we are interested in the rate of change of the function at (a,b) and are holding y fixed this means that we are going to always have $y = b$ (if we didn't have this then eventually y would have to change in order to get to the point...). Doing this will give us a function involving only x's and we can define a new function as follows,

$$g(x) = f(x,b) = 2x^2b^3$$

Now, this is a function of a single variable and at this point all that we are asking is to determine the rate of change of $g(x)$ at $x = a$. In other words, we want to compute $g'(a)$ and since this is a function of a single variable we already know how to do that. Here is the rate of change of the function at (a,b) if we hold y fixed and allow x to vary.

$$g'(a) = 4ab^3$$

We will call $g'(a)$ the **partial derivative** of $f(x,y)$ with respect to x at (a,b) and we will denote it in the following way,

$$f_x(a,b) = 4ab^3$$

Now, let's do it the other way. We will now hold x fixed and allow y to vary. We can do this in a similar way. Since we are holding x fixed it must be fixed at $x = a$ and so we can define a new function of y and then differentiate this as we've always done with functions of one variable.

Here is the work for this,

$$h(y) = f(a, y) = 2a^2y^3 \qquad \Rightarrow \qquad h'(b) = 6a^2b^2$$

In this case we call $h'(b)$ the **partial derivative** of $f(x, y)$ with respect to y at (a, b) and we denote it as follows,

$$f_y(a, b) = 6a^2b^2$$

Note that these two partial derivatives are sometimes called the **first order partial derivatives**. Just as with functions of one variable we can have derivatives of all orders. We will be looking at higher order derivatives in a later section.

Note that the notation for partial derivatives is different than that for derivatives of functions of a single variable. With functions of a single variable we could denote the derivative with a single prime. However, with partial derivatives we will always need to remember the variable that we are differentiating with respect to and so we will subscript the variable that we differentiated with respect to. We will shortly be seeing some alternate notation for partial derivatives as well.

Note as well that we usually don't use the (a, b) notation for partial derivatives as that implies we are working with a specific point which we usually are not doing. The more standard notation is to just continue to use (x, y). So, the partial derivatives from above will more commonly be written as,

$$f_x(x, y) = 4xy^3 \qquad \text{and} \qquad f_y(x, y) = 6x^2y^2$$

Now, as this quick example has shown taking derivatives of functions of more than one variable is done in pretty much the same manner as taking derivatives of a single variable. To compute $f_x(x, y)$ all we need to do is treat all the y's as constants (or numbers) and then differentiate the x's as we've always done. Likewise, to compute $f_y(x, y)$ we will treat all the x's as constants and then differentiate the y's as we are used to doing.

Before we work any examples let's get the formal definition of the partial derivative out of the way as well as some alternate notation.

Since we can think of the two partial derivatives above as derivatives of single variable functions it shouldn't be too surprising that the definition of each is very similar to the definition of the derivative for single variable functions. Here are the formal definitions of the two partial derivatives we looked at above.

$$f_x(x, y) = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h} \qquad f_y(x, y) = \lim_{h \to 0} \frac{f(x, y+h) - f(x, y)}{h}$$

If you recall the Calculus I definition of the limit these should look familiar as they are very close to the Calculus I definition with a (possibly) obvious change.

Now let's take a quick look at some of the possible alternate notations for partial derivatives. Given the function $z = f(x, y)$ the following are all equivalent notations,

$$f_x(x,y) = f_x = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\big(f(x,y)\big) = z_x = \frac{\partial z}{\partial x} = D_x f$$
$$f_y(x,y) = f_y = \frac{\partial f}{\partial y} = \frac{\partial}{\partial y}\big(f(x,y)\big) = z_y = \frac{\partial z}{\partial y} = D_y f$$

For the fractional notation for the partial derivative notice the difference between the partial derivative and the ordinary derivative from single variable calculus.

$$f(x) \quad \Rightarrow \quad f'(x) = \frac{df}{dx}$$
$$f(x,y) \quad \Rightarrow \quad f_x(x,y) = \frac{\partial f}{\partial x} \text{ \& } f_y(x,y) = \frac{\partial f}{\partial y}$$

Okay, now let's work some examples. When working these examples always keep in mind that we need to pay very close attention to which variable we are differentiating with respect to. This is important because we are going to treat all other variables as constants and then proceed with the derivative as if it was a function of a single variable. If you can remember this you'll find that doing partial derivatives are not much more difficult that doing derivatives of functions of a single variable as we did in Calculus I.

Example 1

Find all of the first order partial derivatives for the following functions.

(a) $f(x,y) = x^4 + 6\sqrt{y} - 10$

(b) $w = x^2 y - 10y^2 z^3 + 43x - 7\tan(4y)$

(c) $h(s,t) = t^7 \ln(s^2) + \dfrac{9}{t^3} - \sqrt[7]{s^4}$

(d) $f(x,y) = \cos\left(\dfrac{4}{x}\right) \mathbf{e}^{x^2 y - 5y^3}$

Solution

(a) $f(x,y) = x^4 + 6\sqrt{y} - 10$

Let's first take the derivative with respect to x and remember that as we do so all the y's will be treated as constants. The partial derivative with respect to x is,

$$f_x(x,y) = 4x^3$$

Notice that the second and the third term differentiate to zero in this case. It should be clear why the third term differentiated to zero. It's a constant and we know that constants always differentiate to zero. This is also the reason that the second term differentiated to zero. Remember that since we are differentiating with respect to x here we are going to treat all y's as constants. That means that terms that only involve y's will be treated as constants and hence will differentiate to zero.

Now, let's take the derivative with respect to y. In this case we treat all x's as constants and so the first term involves only x's and so will differentiate to zero, just as the third term will. Here is the partial derivative with respect to y.

$$f_y(x,y) = \frac{3}{\sqrt{y}}$$

(b) $w = x^2y - 10y^2z^3 + 43x - 7\tan(4y)$

With this function we've got three first order derivatives to compute. Let's do the partial derivative with respect to x first. Since we are differentiating with respect to x we will treat all y's and all z's as constants. This means that the second and fourth terms will differentiate to zero since they only involve y's and z's.

This first term contains both x's and y's and so when we differentiate with respect to x the y will be thought of as a multiplicative constant and so the first term will be differentiated just as the third term will be differentiated.

Here is the partial derivative with respect to x.

$$\frac{\partial w}{\partial x} = 2xy + 43$$

Let's now differentiate with respect to y. In this case all x's and z's will be treated as constants. This means the third term will differentiate to zero since it contains only x's while the x's in the first term and the z's in the second term will be treated as multiplicative constants. Here is the derivative with respect to y.

$$\frac{\partial w}{\partial y} = x^2 - 20yz^3 - 28\sec^2(4y)$$

Finally, let's get the derivative with respect to z. Since only one of the terms involve z's this will be the only non-zero term in the derivative. Also, the y's in that term will be treated as multiplicative constants. Here is the derivative with respect to z.

$$\frac{\partial w}{\partial z} = -30y^2z^2$$

(c) $h(s,t) = t^7 \ln(s^2) + \dfrac{9}{t^3} - \sqrt[7]{s^4}$

With this one we'll not put in the detail of the first two. Before taking the derivative let's rewrite the function a little to help us with the differentiation process.

$$h(s,t) = t^7 \ln(s^2) + 9t^{-3} - s^{\frac{4}{7}}$$

Now, the fact that we're using s and t here instead of the "standard" x and y shouldn't be a problem. It will work the same way. Here are the two derivatives for this function.

$$h_s(s,t) = \frac{\partial h}{\partial s} = t^7\left(\frac{2s}{s^2}\right) - \frac{4}{7}s^{-\frac{3}{7}} = \frac{2t^7}{s} - \frac{4}{7}s^{-\frac{3}{7}}$$
$$h_t(s,t) = \frac{\partial h}{\partial t} = 7t^6 \ln(s^2) - 27t^{-4}$$

Remember how to differentiate natural logarithms.

$$\frac{d}{dx}\left(\ln(g)(x)\right) = \frac{g'(x)}{g(x)}$$

(d) $f(x,y) = \cos\left(\dfrac{4}{x}\right)\mathbf{e}^{x^2y-5y^3}$

Now, we can't forget the product rule with derivatives. The product rule will work the same way here as it does with functions of one variable. We will just need to be careful to remember which variable we are differentiating with respect to.

Let's start out by differentiating with respect to x. In this case both the cosine and the exponential contain x's and so we've really got a product of two functions involving x's and so we'll need to product rule this up. Here is the derivative with respect to x.

$$\begin{aligned} f_x(x,y) &= -\sin\left(\frac{4}{x}\right)\left(-\frac{4}{x^2}\right)\mathbf{e}^{x^2y-5y^3} + \cos\left(\frac{4}{x}\right)\mathbf{e}^{x^2y-5y^3}(2xy) \\ &= \frac{4}{x^2}\sin\left(\frac{4}{x}\right)\mathbf{e}^{x^2y-5y^3} + 2xy\cos\left(\frac{4}{x}\right)\mathbf{e}^{x^2y-5y^3} \end{aligned}$$

Do not forget the chain rule for functions of one variable. We will be looking at the chain rule for some more complicated expressions for multivariable functions in a later section. However, at this point we're treating all the y's as constants and so the chain rule will continue to work as it did back in Calculus I.

Also, don't forget how to differentiate exponential functions,

$$\frac{d}{dx}\left(\mathbf{e}^{f(x)}\right) = f'(x)\,\mathbf{e}^{f(x)}$$

Now, let's differentiate with respect to y. In this case we don't have a product rule to worry about since the only place that the y shows up is in the exponential. Therefore, since x's are considered to be constants for this derivative, the cosine in the front will also be thought of as a multiplicative constant. Here is the derivative with respect to y.

$$f_y(x,y) = \left(x^2 - 15y^2\right)\cos\left(\frac{4}{x}\right)\mathbf{e}^{x^2y-5y^3}$$

Example 2

Find all of the first order partial derivatives for the following functions.

(a) $z = \dfrac{9u}{u^2+5v}$

(b) $g(x,y,z) = \dfrac{x\sin(y)}{z^2}$

(c) $z = \sqrt{x^2 + \ln\left(5x-3y^2\right)}$

Solution

(a) $z = \dfrac{9u}{u^2+5v}$

We also can't forget about the quotient rule. Since there isn't too much to this one, we will simply give the derivatives.

$$z_u = \frac{9\left(u^2+5v\right) - 9u(2u)}{\left(u^2+5v\right)^2} = \frac{-9u^2+45v}{\left(u^2+5v\right)^2}$$

$$z_v = \frac{(0)\left(u^2+5v\right) - 9u(5)}{\left(u^2+5v\right)^2} = \frac{-45u}{\left(u^2+5v\right)^2}$$

In the case of the derivative with respect to v recall that u's are constant and so when we differentiate the numerator we will get zero!

(b) $g(x,y,z) = \dfrac{x\sin(y)}{z^2}$

Now, we do need to be careful however to not use the quotient rule when it doesn't need to be used. In this case we do have a quotient, however, since the x's and y's only appear in the numerator and the z's only appear in the denominator this really isn't a quotient rule problem.

Let's do the derivatives with respect to x and y first. In both these cases the z's are constants and so the denominator in this is a constant and so we don't really need to worry too much about it. Here are the derivatives for these two cases.

$$g_x(x,y,z) = \frac{\sin(y)}{z^2} \qquad g_y(x,y,z) = \frac{x\cos(y)}{z^2}$$

Now, in the case of differentiation with respect to z we can avoid the quotient rule with a quick rewrite of the function. Here is the rewrite as well as the derivative with respect to z.

$$\begin{aligned} g(x,y,z) &= x\sin(y)\,z^{-2} \\ g_z(x,y,z) &= -2x\sin(y)\,z^{-3} = -\frac{2x\sin(y)}{z^3} \end{aligned}$$

We went ahead and put the derivative back into the "original" form just so we could say that we did. In practice you probably don't really need to do that.

(c) $z = \sqrt{x^2 + \ln(5x - 3y^2)}$

In this last part we are just going to do a somewhat messy chain rule problem. However, if you had a good background in Calculus I chain rule this shouldn't be all that difficult of a problem. Here are the two derivatives,

$$\begin{aligned} z_x &= \frac{1}{2}\left(x^2 + \ln(5x-3y^2)\right)^{-\frac{1}{2}} \frac{\partial}{\partial x}\left(x^2 + \ln(5x-3y^2)\right) \\ &= \frac{1}{2}\left(x^2 + \ln(5x-3y^2)\right)^{-\frac{1}{2}} \left(2x + \frac{5}{5x-3y^2}\right) \\ &= \left(x + \frac{5}{2(5x-3y^2)}\right)\left(x^2 + \ln(5x-3y^2)\right)^{-\frac{1}{2}} \end{aligned}$$

$$\begin{aligned} z_y &= \frac{1}{2}\left(x^2 + \ln(5x-3y^2)\right)^{-\frac{1}{2}} \frac{\partial}{\partial y}\left(x^2 + \ln(5x-3y^2)\right) \\ &= \frac{1}{2}\left(x^2 + \ln(5x-3y^2)\right)^{-\frac{1}{2}} \left(\frac{-6y}{5x-3y^2}\right) \\ &= -\frac{3y}{5x-3y^2}\left(x^2 + \ln(5x-3y^2)\right)^{-\frac{1}{2}} \end{aligned}$$

So, there are some examples of partial derivatives. Hopefully you will agree that as long as we can remember to treat the other variables as constants these work in exactly the same manner that derivatives of functions of one variable do. So, if you can do Calculus I derivatives you shouldn't have too much difficulty in doing basic partial derivatives.

There is one final topic that we need to take a quick look at in this section, implicit differentiation. Before getting into implicit differentiation for multiple variable functions let's first remember how implicit differentiation works for functions of one variable.

Example 3

Find $\dfrac{dy}{dx}$ for $3y^4 + x^7 = 5x$.

Solution

Remember that the key to this is to always think of y as a function of x, or $y = y(x)$ and so whenever we differentiate a term involving y's with respect to x we will really need to use the chain rule which will mean that we will add on a $\frac{dy}{dx}$ to that term.

The first step is to differentiate both sides with respect to x.

$$12y^3 \frac{dy}{dx} + 7x^6 = 5$$

The final step is to solve for $\frac{dy}{dx}$.

$$\frac{dy}{dx} = \frac{5 - 7x^6}{12y^3}$$

Now, we did this problem because implicit differentiation works in exactly the same manner with functions of multiple variables. If we have a function in terms of three variables x, y, and z we will assume that z is in fact a function of x and y. In other words, $z = z(x, y)$. Then whenever we differentiate z's with respect to x we will use the chain rule and add on a $\frac{\partial z}{\partial x}$. Likewise, whenever we differentiate z's with respect to y we will add on a $\frac{\partial z}{\partial y}$.

Let's take a quick look at a couple of implicit differentiation problems.

Example 4

Find $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$ for each of the following functions.

(a) $x^3 z^2 - 5xy^5 z = x^2 + y^3$

(b) $x^2 \sin(2y - 5z) = 1 + y\cos(6zx)$

Solution

(a) $x^3 z^2 - 5xy^5 z = x^2 + y^3$

Let's start with finding $\frac{\partial z}{\partial x}$. We first will differentiate both sides with respect to x and remember to add on a $\frac{\partial z}{\partial x}$ whenever we differentiate a z from the chain rule.

$$3x^2z^2 + 2x^3z\frac{\partial z}{\partial x} - 5y^5z - 5xy^5\frac{\partial z}{\partial x} = 2x$$

Remember that since we are assuming $z = z(x, y)$ then any product of x's and z's will be a product and so will need the product rule!

Now, solve for $\frac{\partial z}{\partial x}$.

$$\begin{aligned}\left(2x^3z - 5xy^5\right)\frac{\partial z}{\partial x} &= 2x - 3x^2z^2 + 5y^5z \\ \frac{\partial z}{\partial x} &= \frac{2x - 3x^2z^2 + 5y^5z}{2x^3z - 5xy^5}\end{aligned}$$

Now we'll do the same thing for $\frac{\partial z}{\partial y}$ except this time we'll need to remember to add on a $\frac{\partial z}{\partial y}$ whenever we differentiate a z from the chain rule.

$$\begin{aligned}2x^3z\frac{\partial z}{\partial y} - 25xy^4z - 5xy^5\frac{\partial z}{\partial y} &= 3y^2 \\ \left(2x^3z - 5xy^5\right)\frac{\partial z}{\partial y} &= 3y^2 + 25xy^4z \\ \frac{\partial z}{\partial y} &= \frac{3y^2 + 25xy^4z}{2x^3z - 5xy^5}\end{aligned}$$

(b) $x^2 \sin(2y - 5z) = 1 + y\cos(6zx)$

We'll do the same thing for this function as we did in the previous part. First let's find $\frac{\partial z}{\partial x}$.

$$2x\sin(2y - 5z) + x^2\cos(2y - 5z)\left(-5\frac{\partial z}{\partial x}\right) = -y\sin(6zx)\left(6z + 6x\frac{\partial z}{\partial x}\right)$$

Don't forget to do the chain rule on each of the trig functions and when we are differentiating the inside function on the cosine we will need to also use the product rule. Now let's solve for $\frac{\partial z}{\partial x}$.

$$\begin{aligned}2x\sin(2y - 5z) - 5\frac{\partial z}{\partial x}x^2\cos(2y - 5z) &= -6zy\sin(6zx) - 6yx\sin(6zx)\frac{\partial z}{\partial x} \\ 2x\sin(2y - 5z) + 6zy\sin(6zx) &= \left(5x^2\cos(2y - 5z) - 6yx\sin(6zx)\right)\frac{\partial z}{\partial x} \\ \frac{\partial z}{\partial x} &= \frac{2x\sin(2y - 5z) + 6zy\sin(6zx)}{5x^2\cos(2y - 5z) - 6yx\sin(6zx)}\end{aligned}$$

Now let's take care of $\frac{\partial z}{\partial y}$. This one will be slightly easier than the first one.

$$\begin{aligned}
x^2 \cos(2y - 5z)\left(2 - 5\frac{\partial z}{\partial y}\right) &= \cos(6zx) - y\sin(6zx)\left(6x\frac{\partial z}{\partial y}\right) \\
2x^2\cos(2y-5z) - 5x^2\cos(2y-5z)\frac{\partial z}{\partial y} &= \cos(6zx) - 6xy\sin(6zx)\frac{\partial z}{\partial y} \\
\left(6xy\sin(6zx) - 5x^2\cos(2y-5z)\right)\frac{\partial z}{\partial y} &= \cos(6zx) - 2x^2\cos(2y-5z) \\
\frac{\partial z}{\partial y} &= \frac{\cos(6zx) - 2x^2\cos(2y-5z)}{6xy\sin(6zx) - 5x^2\cos(2y-5z)}
\end{aligned}$$

There's quite a bit of work to these. We will see an easier way to do implicit differentiation in a later section.

13.3 Interpretations of Partial Derivatives

This is a fairly short section and is here so we can acknowledge that the two main interpretations of derivatives of functions of a single variable still hold for partial derivatives, with small modifications of course to account of the fact that we now have more than one variable.

The first interpretation we've already seen and is the more important of the two. As with functions of single variables partial derivatives represent the rates of change of the functions as the variables change. As we saw in the previous section, $f_x(x, y)$ represents the rate of change of the function $f(x, y)$ as we change x and hold y fixed while $f_y(x, y)$ represents the rate of change of $f(x, y)$ as we change y and hold x fixed.

Example 1

Determine if $f(x, y) = \dfrac{x^2}{y^3}$ is increasing or decreasing at $(2, 5)$,

(a) if we allow x to vary and hold y fixed.

(b) if we allow y to vary and hold x fixed.

Solution

(a) if we allow x to vary and hold y fixed.

In this case we will first need $f_x(x, y)$ and its value at the point.

$$f_x(x, y) = \frac{2x}{y^3} \qquad \Rightarrow \qquad f_x(2, 5) = \frac{4}{125} > 0$$

So, the partial derivative with respect to x is positive and so if we hold y fixed the function is increasing at $(2, 5)$ as we vary x.

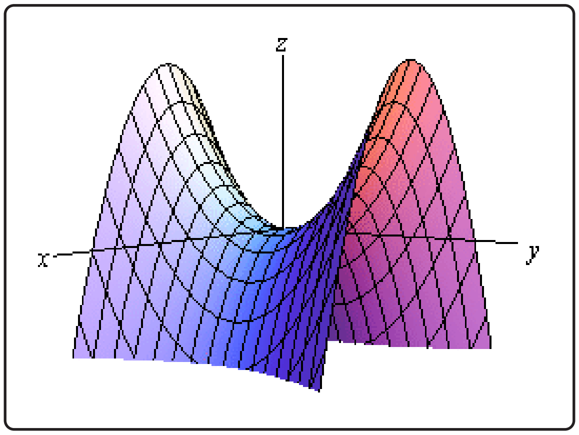
(b) if we allow y to vary and hold x fixed.

For this part we will need $f_y(x, y)$ and its value at the point.

$$f_y(x, y) = -\frac{3x^2}{y^4} \qquad \Rightarrow \qquad f_y(2, 5) = -\frac{12}{625} < 0$$

Here the partial derivative with respect to y is negative and so the function is decreasing at $(2, 5)$ as we vary y and hold x fixed.

Note that it is completely possible for a function to be increasing for a fixed y and decreasing for a fixed x at a point as this example has shown. To see a nice example of this take a look at the following graph.



This is a graph of a hyperbolic paraboloid and at the origin we can see that if we move in along the y-axis the graph is increasing and if we move along the x-axis the graph is decreasing. So it is completely possible to have a graph both increasing and decreasing at a point depending upon the direction that we move. We should never expect that the function will behave in exactly the same way at a point as each variable changes.

The next interpretation was one of the standard interpretations in a Calculus I class. We know from a Calculus I class that $f'(a)$ represents the slope of the tangent line to $y = f(x)$ at $x = a$. Well, $f_x(a,b)$ and $f_y(a,b)$ also represent the slopes of tangent lines. The difference here is the functions that they represent tangent lines to.

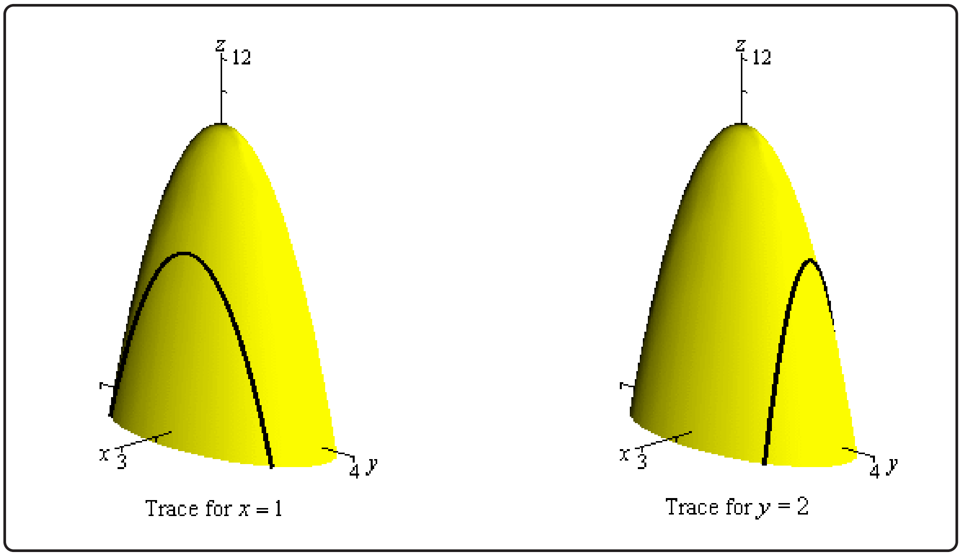
Partial derivatives are the slopes of traces. The partial derivative $f_x(a,b)$ is the slope of the trace of $f(x,y)$ for the plane $y = b$ at the point (a,b). Likewise the partial derivative $f_y(a,b)$ is the slope of the trace of $f(x,y)$ for the plane $x = a$ at the point (a,b).

Example 2

Find the slopes of the traces to $z = 10 - 4x^2 - y^2$ at the point $(1,2)$.

Solution

We sketched the traces for the planes $x = 1$ and $y = 2$ in a previous section and these are the two traces for this point. For reference purposes here are the graphs of the traces.



Next, we'll need the two partial derivatives so we can get the slopes.

$$f_x(x, y) = -8x \qquad f_y(x, y) = -2y$$

To get the slopes all we need to do is evaluate the partial derivatives at the point in question.

$$f_x(1, 2) = -8 \qquad f_y(1, 2) = -4$$

So, the tangent line at $(1, 2)$ for the trace to $z = 10 - 4x^2 - y^2$ for the plane $y = 2$ has a slope of -8. Also the tangent line at $(1, 2)$ for the trace to $z = 10 - 4x^2 - y^2$ for the plane $x = 1$ has a slope of -4.

Finally, let's briefly talk about getting the equations of the tangent line. Recall that the equation of a line in 3-D space is given by a vector equation. Also, to get the equation we need a point on the line and a vector that is parallel to the line.

The point is easy. Since we know the x-y coordinates of the point all we need to do is plug this into the equation to get the point. So, the point will be,

$$\left(a, b, f(a, b)\right)$$

The parallel (or tangent) vector is also just as easy. We can write the equation of the surface as a vector function as follows,

$$\vec{r}(x, y) = \langle x, y, z \rangle = \left\langle x, y, f(x, y) \right\rangle$$

We know that if we have a vector function of one variable we can get a tangent vector by differentiating the vector function. The same will hold true here. If we differentiate with respect to x we will get a tangent vector to traces for the plane $y = b$ (*i.e.* for fixed y) and if we differentiate with respect to y we will get a tangent vector to traces for the plane $x = a$ (or fixed x).

So, here is the tangent vector for traces with fixed y.

$$\vec{r}_x(x,y) = \left\langle 1, 0, f_x(x,y) \right\rangle$$

We differentiated each component with respect to x. Therefore, the first component becomes a 1 and the second becomes a zero because we are treating y as a constant when we differentiate with respect to x. The third component is just the partial derivative of the function with respect to x.

For traces with fixed x the tangent vector is,

$$\vec{r}_y(x,y) = \left\langle 0, 1, f_y(x,y) \right\rangle$$

The equation for the tangent line to traces with fixed y is then,

$$\vec{r}(t) = \left\langle a, b, f(a,b) \right\rangle + t\left\langle 1, 0, f_x(a,b) \right\rangle$$

and the tangent line to traces with fixed x is,

$$\vec{r}(t) = \left\langle a, b, f(a,b) \right\rangle + t\left\langle 0, 1, f_y(a,b) \right\rangle$$

Example 3

Write down the vector equations of the tangent lines to the traces to $z = 10 - 4x^2 - y^2$ at the point $(1, 2)$.

Solution

There really isn't all that much to do with these other than plugging the values and function into the formulas above. We've already computed the derivatives and their values at $(1, 2)$ in the previous example and the point on each trace is,

$$\left(1, 2, f(1,2)\right) = (1, 2, 2)$$

Here is the equation of the tangent line to the trace for the plane $y = 2$.

$$\vec{r}(t) = \langle 1, 2, 2 \rangle + t\,\langle 1, 0, -8 \rangle = \langle 1 + t, 2, 2 - 8t \rangle$$

Here is the equation of the tangent line to the trace for the plane $x = 1$.

$$\vec{r}(t) = \langle 1, 2, 2 \rangle + t\,\langle 0, 1, -4 \rangle = \langle 1, 2 + t, 2 - 4t \rangle$$

13.4 Higher Order Partial Derivatives

Just as we had higher order derivatives with functions of one variable we will also have higher order derivatives of functions of more than one variable. However, this time we will have more options since we do have more than one variable.

Consider the case of a function of two variables, $f(x, y)$ since both of the first order partial derivatives are also functions of x and y we could in turn differentiate each with respect to x or y. This means that for the case of a function of two variables there will be a total of four possible second order derivatives. Here they are and the notations that we'll use to denote them.

$$
\begin{aligned}
(f_x)_x &= f_{x\,x} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} \\
(f_x)_y &= f_{x\,y} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} \\
(f_y)_x &= f_{y\,x} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y} \\
(f_y)_y &= f_{y\,y} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2}
\end{aligned}
$$

The second and third second order partial derivatives are often called mixed partial derivatives since we are taking derivatives with respect to more than one variable. Note as well that the order that we take the derivatives in is given by the notation for each these. If we are using the subscripting notation, *e.g.* $f_{x\,y}$, then we will differentiate from left to right. In other words, in this case, we will differentiate first with respect to x and then with respect to y. With the fractional notation, *e.g.* $\frac{\partial^2 f}{\partial y \partial x}$, it is the opposite. In these cases we differentiate moving along the denominator from right to left. So, again, in this case we differentiate with respect to x first and then y.

Let's take a quick look at an example.

Example 1

Find all the second order derivatives for $f(x, y) = \cos(2x) - x^2 \mathbf{e}^{5y} + 3y^2$.

Solution

We'll first need the first order derivatives so here they are.

$$
\begin{aligned}
f_x(x, y) &= -2\sin(2x) - 2x\mathbf{e}^{5y} \\
f_y(x, y) &= -5x^2\mathbf{e}^{5y} + 6y
\end{aligned}
$$

Now, let's get the second order derivatives.

$$\begin{aligned} f_{xx} &= -4\cos(2x) - 2\mathbf{e}^{5y} \\ f_{xy} &= -10x\mathbf{e}^{5y} \\ f_{yx} &= -10x\mathbf{e}^{5y} \\ f_{yy} &= -25x^2\mathbf{e}^{5y} + 6 \end{aligned}$$

Notice that we dropped the (x, y) from the derivatives. This is fairly standard and we will be doing it most of the time from this point on. We will also be dropping it for the first order derivatives in most cases.

Now let's also notice that, in this case, $f_{xy} = f_{yx}$. This is not by coincidence. If the function is "nice enough" this will always be the case. So, what's "nice enough"? The following theorem tells us.

Clairaut's Theorem

Suppose that f is defined on a disk D that contains the point (a, b). If the functions f_{xy} and f_{yx} are continuous on this disk then,

$$f_{xy}(a, b) = f_{yx}(a, b)$$

Now, do not get too excited about the disk business and the fact that we gave the theorem for a specific point. In pretty much every example in this class if the two mixed second order partial derivatives are continuous then they will be equal.

Example 2

Verify Clairaut's Theorem for $f(x, y) = x\mathbf{e}^{-x^2y^2}$.

Solution

We'll first need the two first order derivatives.

$$\begin{aligned} f_x(x, y) &= \mathbf{e}^{-x^2y^2} - 2x^2y^2\mathbf{e}^{-x^2y^2} \\ f_y(x, y) &= -2yx^3\mathbf{e}^{-x^2y^2} \end{aligned}$$

Now, compute the two mixed second order partial derivatives.

$$\begin{aligned} f_{xy}(x, y) &= -2yx^2\mathbf{e}^{-x^2y^2} - 4x^2y\mathbf{e}^{-x^2y^2} + 4x^4y^3\mathbf{e}^{-x^2y^2} = -6x^2y\mathbf{e}^{-x^2y^2} + 4x^4y^3\mathbf{e}^{-x^2y^2} \\ f_{yx}(x, y) &= -6yx^2\mathbf{e}^{-x^2y^2} + 4y^3x^4\mathbf{e}^{-x^2y^2} \end{aligned}$$

Sure enough they are the same.

So far we have only looked at second order derivatives. There are, of course, higher order derivatives as well. Here are a couple of the third order partial derivatives of function of two variables.

$$f_{x\,y\,x} = (f_{xy})_x = \frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial y \partial x}\right) = \frac{\partial^3 f}{\partial x \partial y \partial x}$$
$$f_{y\,x\,x} = (f_{yx})_x = \frac{\partial}{\partial x}\left(\frac{\partial^2 f}{\partial x \partial y}\right) = \frac{\partial^3 f}{\partial x^2 \partial y}$$

Notice as well that for both of these we differentiate once with respect to y and twice with respect to x. There is also another third order partial derivative in which we can do this, $f_{x\,x\,y}$. There is an extension to Clairaut's Theorem that says if all three of these are continuous then they should all be equal,

$$f_{x\,x\,y} = f_{x\,y\,x} = f_{y\,x\,x}$$

To this point we've only looked at functions of two variables, but everything that we've done to this point will work regardless of the number of variables that we've got in the function and there are natural extensions to Clairaut's theorem to all of these cases as well. For instance,

$$f_{x\,z}\,(x, y, z) = f_{z\,x}\,(x, y, z)$$

provided both of the derivatives are continuous.

In general, we can extend Clairaut's theorem to any function and mixed partial derivatives. The only requirement is that in each derivative we differentiate with respect to each variable the same number of times. In other words, provided we meet the continuity condition, the following will be equal

$$f_{s\,s\,r\,t\,s\,r\,r} = f_{t\,r\,s\,r\,s\,s\,r}$$

because in each case we differentiate with respect to t once, s three times and r three times.

Let's do a couple of examples with higher (well higher order than two anyway) order derivatives and functions of more than two variables.

Example 3

Find the indicated derivative for each of the following functions.

(a) Find $f_{x\,x\,y\,z\,z}$ for $f(x,y,z) = z^3 y^2 \ln(x)$

(b) Find $\dfrac{\partial^3 f}{\partial y \partial x^2}$ for $f(x,y) = \mathbf{e}^{xy}$

Solution

(a) Find $f_{x\,x\,y\,z\,z}$ for $f(x,y,z) = z^3 y^2 \ln(x)$

In this case remember that we differentiate from left to right. Here are the derivatives for this part.

$$\begin{aligned} f_x &= \frac{z^3 y^2}{x} \\ f_{xx} &= -\frac{z^3 y^2}{x^2} \\ f_{xxy} &= -\frac{2z^3 y}{x^2} \\ f_{xxyz} &= -\frac{6z^2 y}{x^2} \\ f_{xxyzz} &= -\frac{12zy}{x^2} \end{aligned}$$

(b) Find $\dfrac{\partial^3 f}{\partial y \partial x^2}$ for $f(x,y) = \mathbf{e}^{xy}$

Here we differentiate from right to left. Here are the derivatives for this function.

$$\begin{aligned} \frac{\partial f}{\partial x} &= y\mathbf{e}^{xy} \\ \frac{\partial^2 f}{\partial x^2} &= y^2 \mathbf{e}^{xy} \\ \frac{\partial^3 f}{\partial y \partial x^2} &= 2y\mathbf{e}^{xy} + xy^2 \mathbf{e}^{xy} \end{aligned}$$

13.5 Differentials

This is a very short section and is here simply to acknowledge that just like we had differentials for functions of one variable we also have them for functions of more than one variable. Also, as we've already seen in previous sections, when we move up to more than one variable things work pretty much the same, but there are some small differences.

Given the function $z = f(x, y)$ the differential dz or df is given by,

$$dz = f_x\,dx + f_y\,dy \qquad \text{or} \qquad df = f_x\,dx + f_y\,dy$$

There is a natural extension to functions of three or more variables. For instance, given the function $w = g(x, y, z)$ the differential is given by,

$$dw = g_x\,dx + g_y\,dy + g_z\,dz$$

Let's do a couple of quick examples.

Example 1

Compute the differentials for each of the following functions.

(a) $z = \mathbf{e}^{x^2+y^2}\tan(2x)$

(b) $u = \dfrac{t^3 r^6}{s^2}$

Solution

(a) $z = \mathbf{e}^{x^2+y^2}\tan(2x)$

There really isn't a whole lot to these outside of some quick differentiation. Here is the differential for the function.

$$dz = \left(2x\mathbf{e}^{x^2+y^2}\tan(2x) + 2\mathbf{e}^{x^2+y^2}\sec^2(2x)\right)dx + 2y\mathbf{e}^{x^2+y^2}\tan(2x)\,dy$$

(b) $u = \dfrac{t^3 r^6}{s^2}$

Here is the differential for this function.

$$du = \frac{3t^2r^6}{s^2}dt + \frac{6t^3r^5}{s^2}dr - \frac{2t^3r^6}{s^3}ds$$

Note that sometimes these differentials are called the **total differentials**.

13.6 Chain Rule

We've been using the standard chain rule for functions of one variable throughout the last couple of sections. It's now time to extend the chain rule out to more complicated situations. Before we actually do that let's first review the notation for the chain rule for functions of one variable.

The notation that's probably familiar to most people is the following.

$$F(x) = f\big(g(x)\big) \qquad F'(x) = f'\big(g(x)\big)g'(x)$$

There is an alternate notation however that while probably not used much in Calculus I is more convenient at this point because it will match up with the notation that we are going to be using in this section. Here it is.

$$\text{If} \quad y = f(x) \quad \text{and} \quad x = g(t) \quad \text{then} \quad \frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}$$

Notice that the derivative $\frac{dy}{dt}$ really does make sense here since if we were to plug in for x then y really would be a function of t. One way to remember this form of the chain rule is to note that if we think of the two derivatives on the right side as fractions the dx's will cancel to get the same derivative on both sides.

Okay, now that we've got that out of the way let's move into the more complicated chain rules that we are liable to run across in this course.

As with many topics in multivariable calculus, there are in fact many different formulas depending upon the number of variables that we're dealing with. So, let's start this discussion off with a function of two variables, $z = f(x, y)$. From this point there are still many different possibilities that we can look at. We will be looking at two distinct cases prior to generalizing the whole idea out.

Case 1 : $z = f(x, y)$, $x = g(t)$, $y = h(t)$ and compute $\frac{dz}{dt}$.

This case is analogous to the standard chain rule from Calculus I that we looked at above. In this case we are going to compute an ordinary derivative since z really would be a function of t only if we were to substitute in for x and y.

The chain rule for this case is,

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

So, basically what we're doing here is differentiating f with respect to each variable in it and then multiplying each of these by the derivative of that variable with respect to t. The final step is to then add all this up.

Let's take a look at a couple of examples.

Example 1

Compute $\dfrac{dz}{dt}$ for each of the following.

(a) $z = x\mathbf{e}^{xy}$, $x = t^2$, $y = t^{-1}$

(b) $z = x^2y^3 + y\cos(x)$, $x = \ln\left(t^2\right)$, $y = \sin(4t)$

Solution

(a) $z = x\mathbf{e}^{xy}$, $x = t^2$, $y = t^{-1}$

There really isn't all that much to do here other than using the formula.

$$\begin{aligned}\frac{dz}{dt} &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \\ &= \left(\mathbf{e}^{xy} + yx\mathbf{e}^{xy}\right)(2t) + x^2\mathbf{e}^{xy}\left(-t^{-2}\right) \\ &= 2t\left(\mathbf{e}^{xy} + yx\mathbf{e}^{xy}\right) - t^{-2}x^2\mathbf{e}^{xy}\end{aligned}$$

So, technically we've computed the derivative. However, we should probably go ahead and substitute in for x and y as well at this point since we've already got t's in the derivative. Doing this gives,

$$\frac{dz}{dt} = 2t\left(\mathbf{e}^{t} + t\mathbf{e}^{t}\right) - t^{-2}t^4\mathbf{e}^{t} = 2t\mathbf{e}^{t} + t^2\mathbf{e}^{t}$$

Note that in this case it might actually have been easier to just substitute in for x and y in the original function and just compute the derivative as we normally would. For comparison's sake let's do that.

$$z = t^2\mathbf{e}^{t} \qquad \Rightarrow \qquad \frac{dz}{dt} = 2t\mathbf{e}^{t} + t^2\mathbf{e}^{t}$$

The same result for less work. Note however, that often it will actually be more work to do the substitution first.

(b) $z = x^2y^3 + y\cos(x)$, $x = \ln\left(t^2\right)$, $y = \sin(4t)$

Okay, in this case it would almost definitely be more work to do the substitution first so we'll use the chain rule first and then substitute.

$$\begin{aligned}\frac{dz}{dt} &= \left(2xy^3 - y\sin(x)\right)\left(\frac{2}{t}\right) + \left(3x^2y^2 + \cos(x)\right)\left(4\cos(4t)\right) \\ &= \frac{4\sin^3(4t)\ln t^2 - 2\sin(4t)\sin\left(\ln t^2\right)}{t} + 4\cos(4t)\left(3\sin^2(4t)\left[\ln t^2\right]^2 + \cos\left(\ln t^2\right)\right)\end{aligned}$$

Note that sometimes, because of the significant mess of the final answer, we will only simplify the first step a little and leave the answer in terms of x, y, and t. This is dependent upon the situation, class and instructor however so be careful about not substituting in for without first talking to your instructor.

Now, there is a special case that we should take a quick look at before moving on to the next case. Let's suppose that we have the following situation,

$$z = f(x, y) \qquad y = g(x)$$

In this case the chain rule for $\frac{dz}{dx}$ becomes,

$$\frac{dz}{dx} = \frac{\partial f}{\partial x}\frac{dx}{dx} + \frac{\partial f}{\partial y}\frac{dy}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx}$$

In the first term we are using the fact that,

$$\frac{dx}{dx} = \frac{d}{dx}(x) = 1$$

Let's take a quick look at an example.

Example 2

$\dfrac{dz}{dx}$ for $z = x\ln(xy) + y^3$, $y = \cos\left(x^2+1\right)$

Solution

We'll just plug into the formula.

$$\begin{aligned}\frac{dz}{dx} &= \left(\ln(xy) + x\frac{y}{xy}\right) + \left(x\frac{x}{xy} + 3y^2\right)\left(-2x\sin\left(x^2+1\right)\right)\\ &= \ln\left(x\cos\left(x^2+1\right)\right) + 1 - 2x\sin\left(x^2+1\right)\left(\frac{x}{\cos\left(x^2+1\right)} + 3\cos^2\left(x^2+1\right)\right)\\ &= \ln\left(x\cos\left(x^2+1\right)\right) + 1 - 2x^2\tan\left(x^2+1\right) - 6x\sin\left(x^2+1\right)\cos^2\left(x^2+1\right)\end{aligned}$$

Now let's take a look at the second case.

Case 2 : $\boldsymbol{z = f(x,y)}$, $\boldsymbol{x = g(s,t)}$, $\boldsymbol{y = h(s,t)}$ **and compute** $\boldsymbol{\dfrac{\partial z}{\partial s}}$ **and** $\boldsymbol{\dfrac{\partial z}{\partial t}}$.

In this case if we were to substitute in for x and y we would get that z is a function of s and t and so it makes sense that we would be computing partial derivatives here and that there would be two of them.

Here is the chain rule for both of these cases.

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} \qquad\qquad \frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}$$

So, not surprisingly, these are very similar to the first case that we looked at. Here is a quick example of this kind of chain rule.

Example 3

Find $\dfrac{\partial z}{\partial s}$ and $\dfrac{\partial z}{\partial t}$ for $z = \mathbf{e}^{2r}\sin(3\theta)$, $r = st - t^2$, $\theta = \sqrt{s^2+t^2}$.

Solution

Here is the chain rule for $\dfrac{\partial z}{\partial s}$.

$$\begin{aligned}\frac{\partial z}{\partial s} &= \left(2\mathbf{e}^{2r}\sin(3\theta)\right)(t) + \left(3\mathbf{e}^{2r}\cos(3\theta)\right)\frac{s}{\sqrt{s^2+t^2}} \\ &= t\left(2\mathbf{e}^{2(st-t^2)}\sin\left(3\sqrt{s^2+t^2}\right)\right) + \frac{3s\mathbf{e}^{2(st-t^2)}\cos\left(3\sqrt{s^2+t^2}\right)}{\sqrt{s^2+t^2}}\end{aligned}$$

Now the chain rule for $\dfrac{\partial z}{\partial t}$.

$$\begin{aligned}\frac{\partial z}{\partial t} &= \left(2\mathbf{e}^{2r}\sin(3\theta)\right)(s-2t) + \left(3\mathbf{e}^{2r}\cos(3\theta)\right)\frac{t}{\sqrt{s^2+t^2}} \\ &= (s-2t)\left(2\mathbf{e}^{2(st-t^2)}\sin\left(3\sqrt{s^2+t^2}\right)\right) + \frac{3t\mathbf{e}^{2(st-t^2)}\cos\left(3\sqrt{s^2+t^2}\right)}{\sqrt{s^2+t^2}}\end{aligned}$$

Okay, now that we've seen a couple of cases for the chain rule let's see the general version of the chain rule.

Chain Rule

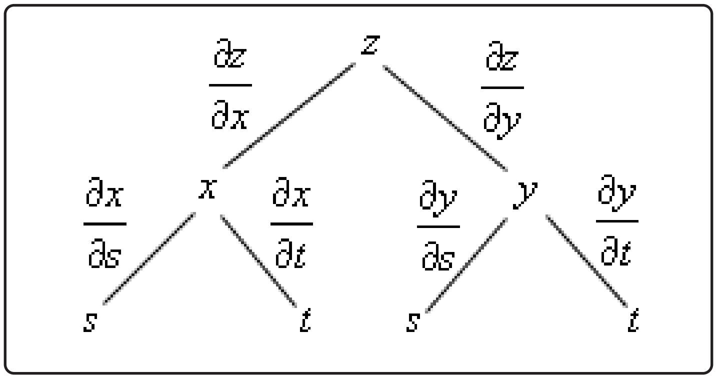
Suppose that z is a function of n variables, $x_1, x_2, \ldots, x_n$, and that each of these variables are in turn functions of m variables, $t_1, t_2, \ldots, t_m$. Then for any variable t_i, $i = 1, 2, \ldots, m$ we have the following,

$$\frac{\partial z}{\partial t_i} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial z}{\partial x_2}\frac{\partial x_2}{\partial t_i} + \cdots + \frac{\partial z}{\partial x_n}\frac{\partial x_n}{\partial t_i}$$

Wow. That's a lot to remember. There is actually an easier way to construct all the chain rules that we've discussed in the section or will look at in later examples. We can build up a **tree diagram** that will give us the chain rule for any situation. To see how these work let's go back and take a look at the chain rule for $\frac{\partial z}{\partial s}$ given that $z = f(x,y)$, $x = g(s,t)$, $y = h(s,t)$. We already know what this is, but it may help to illustrate the tree diagram if we already know the answer. For reference here is the chain rule for this case,

$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}$$

Here is the tree diagram for this case.



We start at the top with the function itself and the branch out from that point. The first set of branches is for the variables in the function. From each of these endpoints we put down a further set of branches that gives the variables that both x and y are a function of. We connect each letter with a line and each line represents a partial derivative as shown. Note that the letter in the numerator of the partial derivative is the upper "node" of the tree and the letter in the denominator of the partial derivative is the lower "node" of the tree.

To use this to get the chain rule we start at the bottom and for each branch that ends with the variable we want to take the derivative with respect to (s in this case) we move up the tree until we hit the top multiplying the derivatives that we see along that set of branches. Once we've done this for each branch that ends at s, we then add the results up to get the chain rule for that given situation.

Note that we don't always put the derivatives in the tree. Some of the trees get a little large/messy and so we won't put in the derivatives. Just remember what derivative should be on each branch and you'll be okay without actually writing them down.

Let's write down some chain rules.

Example 4

Use a tree diagram to write down the chain rule for the given derivatives.

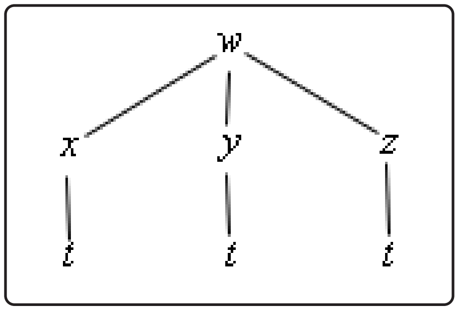
(a) $\dfrac{dw}{dt}$ for $w = f(x, y, z)$, $x = g_1(t)$, $y = g_2(t)$, and $z = g_3(t)$

(b) $\dfrac{\partial w}{\partial r}$ for $w = f(x, y, z)$, $x = g_1(s, t, r)$, $y = g_2(s, t, r)$, and $z = g_3(s, t, r)$

Solution

(a) $\dfrac{dw}{dt}$ for $w = f(x, y, z)$, $x = g_1(t)$, $y = g_2(t)$, and $z = g_3(t)$

So, we'll first need the tree diagram so let's get that.



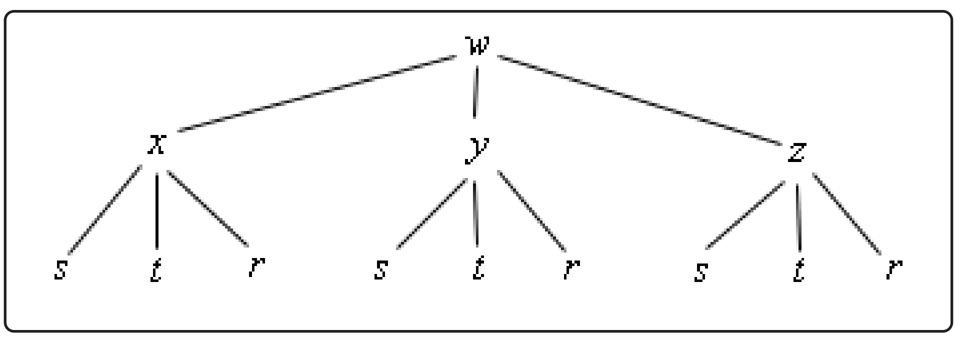
From this it looks like the chain rule for this case should be,

$$\frac{dw}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}$$

which is really just a natural extension to the two variable case that we saw above.

(b) $\dfrac{\partial w}{\partial r}$ for $w = f(x, y, z)$, $x = g_1(s, t, r)$, $y = g_2(s, t, r)$, and $z = g_3(s, t, r)$

Here is the tree diagram for this situation.



From this it looks like the derivative will be,

$$\frac{\partial w}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial r}$$

So, provided we can write down the tree diagram, and these aren't usually too bad to write down, we can do the chain rule for any set up that we might run across.

We've now seen how to take first derivatives of these more complicated situations, but what about higher order derivatives? How do we do those? It's probably easiest to see how to deal with these with an example.

Example 5

Compute $\dfrac{\partial^2 f}{\partial \theta^2}$ for $f(x, y)$ if $x = r\cos(\theta)$ and $y = r\sin(\theta)$.

Solution

We will need the first derivative before we can even think about finding the second derivative so let's get that. This situation falls into the second case that we looked at above so we don't need a new tree diagram. Here is the first derivative.

$$\begin{aligned}\frac{\partial f}{\partial \theta} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta}\\ &= -r\sin(\theta)\frac{\partial f}{\partial x} + r\cos(\theta)\frac{\partial f}{\partial y}\end{aligned}$$

Okay, now we know that the second derivative is,

$$\frac{\partial^2 f}{\partial \theta^2} = \frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial \theta}\right) = \frac{\partial}{\partial \theta}\left(-r\sin(\theta)\frac{\partial f}{\partial x} + r\cos(\theta)\frac{\partial f}{\partial y}\right)$$

The issue here is to correctly deal with this derivative. Since the two first order derivatives, $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$, are both functions of x and y which are in turn functions of r and θ both of these terms are products. So, the using the product rule gives the following,

$$\frac{\partial^2 f}{\partial \theta^2} = -r\cos(\theta)\frac{\partial f}{\partial x} - r\sin(\theta)\frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial x}\right) - r\sin(\theta)\frac{\partial f}{\partial y} + r\cos(\theta)\frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial y}\right)$$

We now need to determine what $\frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial x}\right)$ and $\frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial y}\right)$ will be. These are both chain rule problems again since both of the derivatives are functions of x and y and we want to take the derivative with respect to θ.

Before we do these let's rewrite the first chain rule that we did above a little.

$$\frac{\partial}{\partial \theta}\left(f\right) = -r\sin(\theta)\frac{\partial}{\partial x}\left(f\right) + r\cos(\theta)\frac{\partial}{\partial y}\left(f\right) \tag{13.1}$$

Note that all we've done is change the notation for the derivative a little. With the first chain rule written in this way we can think of Equation 13.1 as a formula for differentiating any function of x and y with respect to θ provided we have $x = r\cos(\theta)$ and $y = r\sin(\theta)$.

This however is exactly what we need to do the two new derivatives we need above. Both of the first order partial derivatives, $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$, are functions of x and y and $x = r\cos(\theta)$ and $y = r\sin(\theta)$ so we can use Equation 13.1 to compute these derivatives.

To do this we'll simply replace all the *f* 's in Equation 13.1 with the first order partial derivative that we want to differentiate. At that point all we need to do is a little notational work and we'll get the formula that we're after.

Here is the use of Equation 13.1 to compute $\frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial x}\right)$.

$$\begin{aligned}\frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial x}\right) &= -r\sin(\theta)\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) + r\cos(\theta)\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\\ &= -r\sin(\theta)\frac{\partial^2 f}{\partial x^2} + r\cos(\theta)\frac{\partial^2 f}{\partial y\partial x}\end{aligned}$$

Here is the computation for $\frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial y}\right)$.

$$\begin{aligned}\frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial y}\right) &= -r\sin(\theta)\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) + r\cos(\theta)\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)\\ &= -r\sin(\theta)\frac{\partial^2 f}{\partial x\partial y} + r\cos(\theta)\frac{\partial^2 f}{\partial y^2}\end{aligned}$$

The final step is to plug these back into the second derivative and do some simplifying.

$$\begin{aligned}\frac{\partial^2 f}{\partial \theta^2} &= -r\cos(\theta)\frac{\partial f}{\partial x} - r\sin(\theta)\left(-r\sin(\theta)\frac{\partial^2 f}{\partial x^2} + r\cos(\theta)\frac{\partial^2 f}{\partial y\partial x}\right) -\\ &\qquad r\sin(\theta)\frac{\partial f}{\partial y} + r\cos(\theta)\left(-r\sin(\theta)\frac{\partial^2 f}{\partial x\partial y} + r\cos(\theta)\frac{\partial^2 f}{\partial y^2}\right)\\ &= -r\cos(\theta)\frac{\partial f}{\partial x} + r^2\sin^2(\theta)\frac{\partial^2 f}{\partial x^2} - r^2\sin(\theta)\cos(\theta)\frac{\partial^2 f}{\partial y\partial x} -\\ &\qquad r\sin(\theta)\frac{\partial f}{\partial y} - r^2\sin(\theta)\cos(\theta)\frac{\partial^2 f}{\partial x\partial y} + r^2\cos^2(\theta)\frac{\partial^2 f}{\partial y^2}\\ &= -r\cos(\theta)\frac{\partial f}{\partial x} - r\sin(\theta)\frac{\partial f}{\partial y} + r^2\sin^2(\theta)\frac{\partial^2 f}{\partial x^2} -\\ &\qquad 2r^2\sin(\theta)\cos(\theta)\frac{\partial^2 f}{\partial y\partial x} + r^2\cos^2(\theta)\frac{\partial^2 f}{\partial y^2}\end{aligned}$$

It's long and fairly messy but there it is.

The final topic in this section is a revisiting of implicit differentiation. With these forms of the chain rule implicit differentiation actually becomes a fairly simple process. Let's start out with the implicit differentiation that we saw in a Calculus I course.

We will start with a function in the form $F(x, y) = 0$ (if it's not in this form simply move everything to one side of the equal sign to get it into this form) where $y = y(x)$. In a Calculus I course we were then asked to compute $\frac{dy}{dx}$ and this was often a fairly messy process. Using the chain rule from this section however we can get a nice simple formula for doing this. We'll start by differentiating both sides with respect to x. This will mean using the chain rule on the left side and the right side will, of course, differentiate to zero. Here are the results of that.

$$F_x + F_y \frac{dy}{dx} = 0 \qquad \Rightarrow \qquad \frac{dy}{dx} = -\frac{F_x}{F_y}$$

As shown, all we need to do next is solve for $\frac{dy}{dx}$ and we've now got a very nice formula to use for implicit differentiation. Note as well that in order to simplify the formula we switched back to using the subscript notation for the derivatives.

Let's check out a quick example.

Example 6

Find $\dfrac{dy}{dx}$ for $x\cos(3y) + x^3y^5 = 3x - \mathbf{e}^{xy}$.

Solution

The first step is to get a zero on one side of the equal sign and that's easy enough to do.

$$x\cos(3y) + x^3y^5 - 3x + \mathbf{e}^{xy} = 0$$

Now, the function on the left is $F(x, y)$ in our formula so all we need to do is use the formula to find the derivative.

$$\frac{dy}{dx} = -\frac{\cos(3y) + 3x^2y^5 - 3 + y\mathbf{e}^{xy}}{-3x\sin(3y) + 5x^3y^4 + x\mathbf{e}^{xy}}$$

There we go. It would have taken much longer to do this using the old Calculus I way of doing this.

We can also do something similar to handle the types of implicit differentiation problems involving partial derivatives like those we saw when we first introduced partial derivatives. In these cases we will start off with a function in the form $F(x, y, z) = 0$ and assume that $z = f(x, y)$ and we want to find $\frac{\partial z}{\partial x}$ and/or $\frac{\partial z}{\partial y}$.

Let's start by trying to find $\frac{\partial z}{\partial x}$. We will differentiate both sides with respect to x and we'll need to remember that we're going to be treating y as a constant. Also, the left side will require the chain rule. Here is this derivative.

$$\frac{\partial F}{\partial x}\frac{\partial x}{\partial x} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial x} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x} = 0$$

Now, we have the following,

$$\frac{\partial x}{\partial x} = 1 \qquad \text{and} \qquad \frac{\partial y}{\partial x} = 0$$

The first is because we are just differentiating x with respect to x and we know that is 1. The second is because we are treating the y as a constant and so it will differentiate to zero.

Plugging these in and solving for $\frac{\partial z}{\partial x}$ gives,

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z}$$

A similar argument can be used to show that,

$$\frac{\partial z}{\partial y} = -\frac{F_y}{F_z}$$

As with the one variable case we switched to the subscripting notation for derivatives to simplify the formulas. Let's take a quick look at an example of this.

Example 7

Find $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$ for $x^2 \sin(2y - 5z) = 1 + y\cos(6zx)$.

Solution

This was one of the functions that we used the old implicit differentiation on back in the Partial Derivatives section. You might want to go back and see the difference between the two.

First let's get everything on one side.

$$x^2 \sin(2y - 5z) - 1 - y\cos(6zx) = 0$$

Now, the function on the left is $F(x, y, z)$ and so all that we need to do is use the formulas developed above to find the derivatives.

$$\frac{\partial z}{\partial x} = -\frac{2x\sin(2y-5z) + 6yz\sin(6zx)}{-5x^2\cos(2y-5z) + 6yx\sin(6zx)}$$

$$\frac{\partial z}{\partial y} = -\frac{2x^2\cos(2y-5z) - \cos(6zx)}{-5x^2\cos(2y-5z) + 6yx\sin(6zx)}$$

If you go back and compare these answers to those that we found the first time around you will notice that they might appear to be different. However, if you take into account the minus sign that sits in the front of our answers here you will see that they are in fact the same.

13.7 Directional Derivatives

To this point we've only looked at the two partial derivatives $f_x(x, y)$ and $f_y(x, y)$. Recall that these derivatives represent the rate of change of f as we vary x (holding y fixed) and as we vary y (holding x fixed) respectively. We now need to discuss how to find the rate of change of f if we allow both x and y to change simultaneously. The problem here is that there are many ways to allow both x and y to change. For instance, one could be changing faster than the other and then there is also the issue of whether or not each is increasing or decreasing. So, before we get into finding the rate of change we need to get a couple of preliminary ideas taken care of first. The main idea that we need to look at is just how are we going to define the changing of x and/or y.

Let's start off by supposing that we wanted the rate of change of f at a particular point, say (x_0, y_0). Let's also suppose that both x and y are increasing and that, in this case, x is increasing twice as fast as y is increasing. So, as y increases one unit of measure x will increase two units of measure.

To help us see how we're going to define this change let's suppose that a particle is sitting at (x_0, y_0) and the particle will move in the direction given by the changing x and y. Therefore, the particle will move off in a direction of increasing x and y and the x coordinate of the point will increase twice as fast as the y coordinate. Now that we're thinking of this changing x and y as a direction of movement we can get a way of defining the change. We know from Calculus II that vectors can be used to define a direction and so the particle, at this point, can be said to be moving in the direction,

$$\vec{v} = \langle 2, 1 \rangle$$

Since this vector can be used to define how a particle at a point is changing we can also use it to describe how x and/or y is changing at a point. For our example we will say that we want the rate of change of f in the direction of $\vec{v} = \langle 2, 1 \rangle$. In this way we will know that x is increasing twice as fast as y is. There is still a small problem with this however. There are many vectors that point in the same direction. For instance, all of the following vectors point in the same direction as $\vec{v} = \langle 2, 1 \rangle$.

$$\vec{v} = \left\langle \frac{1}{5}, \frac{1}{10} \right\rangle \qquad \vec{v} = \langle 6, 3 \rangle \qquad \vec{v} = \left\langle \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right\rangle$$

We need a way to consistently find the rate of change of a function in a given direction. We will do this by insisting that the vector that defines the direction of change be a unit vector. Recall that a unit vector is a vector with length, or magnitude, of 1. This means that for the example that we started off thinking about we would want to use

$$\vec{v} = \left\langle \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right\rangle$$

since this is the unit vector that points in the direction of change.

For reference purposes recall that the magnitude or length of the vector $\vec{v} = \langle a, b, c \rangle$ is given by,

$$\|\vec{v}\| = \sqrt{a^2 + b^2 + c^2}$$

For two dimensional vectors we drop the c from the formula.

Sometimes we will give the direction of changing x and y as an angle. For instance, we may say that we want the rate of change of f in the direction of $\theta = \frac{\pi}{3}$. The unit vector that points in this direction is given by,

$$\vec{u} = \langle \cos(\theta), \sin(\theta) \rangle$$

Okay, now that we know how to define the direction of changing x and y its time to start talking about finding the rate of change of f in this direction. Let's start off with the official definition.

Definition

The rate of change of $f(x, y)$ in the direction of the unit vector $\vec{u} = \langle a, b \rangle$ is called the **directional derivative** and is denoted by $D_{\vec{u}} f(x, y)$. The definition of the directional derivative is,

$$D_{\vec{u}} f(x, y) = \lim_{h \to 0} \frac{f(x + ah, y + bh) - f(x, y)}{h}$$

So, the definition of the directional derivative is very similar to the definition of partial derivatives. However, in practice this can be a very difficult limit to compute so we need an easier way of taking directional derivatives. It's actually fairly simple to derive an equivalent formula for taking directional derivatives.

To see how we can do this let's define a new function of a single variable,

$$g(z) = f(x_0 + az, y_0 + bz)$$

where x_0, y_0, a, and b are some fixed numbers. Note that this really is a function of a single variable now since z is the only letter that is not representing a fixed number.

Then by the definition of the derivative for functions of a single variable we have,

$$g'(z) = \lim_{h \to 0} \frac{g(z + h) - g(z)}{h}$$

and the derivative at $z = 0$ is given by,

$$g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h}$$

If we now substitute in for $g(z)$ we get,

$$g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h} = \lim_{h \to 0} \frac{f(x_0 + ah, y_0 + bh) - f(x_0, y_0)}{h} = D_{\vec{u}} f(x_0, y_0)$$

So, it looks like we have the following relationship.

$$g'(0) = D_{\vec{u}} f(x_0, y_0) \tag{13.2}$$

Now, let's look at this from another perspective. Let's rewrite $g(z)$ as follows,

$$g(z) = f(x, y) \quad \text{where} \quad x = x_0 + az \text{ and } y = y_0 + bz$$

We can now use the chain rule from the previous section to compute,

$$g'(z) = \frac{dg}{dz} = \frac{\partial f}{\partial x}\frac{dx}{dz} + \frac{\partial f}{\partial y}\frac{dy}{dz} = f_x(x, y)\,a + f_y(x, y)\,b$$

So, from the chain rule we get the following relationship.

$$g'(z) = f_x(x, y)\,a + f_y(x, y)\,b \tag{13.3}$$

If we now take $z = 0$ we will get that $x = x_0$ and $y = y_0$ (from how we defined x and y above) and plug these into Equation 13.3 we get,

$$g'(0) = f_x(x_0, y_0)\,a + f_y(x_0, y_0)\,b \tag{13.4}$$

Now, simply equate Equation 13.2 and Equation 13.4 to get that,

$$D_{\vec{u}} f(x_0, y_0) = g'(0) = f_x(x_0, y_0)\,a + f_y(x_0, y_0)\,b$$

If we now go back to allowing x and y to be any number we get the following formula for computing directional derivatives.

2D Directional Derivative Formula

$$D_{\vec{u}} f(x, y) = f_x(x, y)\,a + f_y(x, y)\,b$$

This is much simpler than the limit definition. Also note that this definition assumed that we were working with functions of two variables. There are similar formulas that can be derived by the same type of argument for functions with more than two variables. For instance, the directional derivative of $f(x, y, z)$ in the direction of the unit vector $\vec{u} = \langle a, b, c \rangle$ is given by,

3D Directional Derivative Formula

$$D_{\vec{u}} f(x, y, z) = f_x(x, y, z)\,a + f_y(x, y, z)\,b + f_z(x, y, z)\,c$$

Let's work a couple of examples.

Example 1

Find each of the directional derivatives.

(a) $D_{\vec{u}}f(2,0)$ where $f(x,y) = x\mathbf{e}^{xy} + y$ and $\vec{u}$ is the unit vector in the direction of $\theta = \dfrac{2\pi}{3}$.

(b) $D_{\vec{u}}f(x,y,z)$ where $f(x,y,z) = x^2z + y^3z^2 - xyz$ in the direction of $\vec{v} = \langle -1, 0, 3\rangle$.

Solution

(a) $D_{\vec{u}}f(2,0)$ where $f(x,y) = x\mathbf{e}^{xy} + y$ and $\vec{u}$ is the unit vector in the direction of $\theta = \dfrac{2\pi}{3}$.

We'll first find $D_{\vec{u}}f(x,y)$ and then use this a formula for finding $D_{\vec{u}}f(2,0)$. The unit vector giving the direction is,

$$\vec{u} = \left\langle \cos\left(\frac{2\pi}{3}\right), \sin\left(\frac{2\pi}{3}\right)\right\rangle = \left\langle -\frac{1}{2}, \frac{\sqrt{3}}{2}\right\rangle$$

So, the directional derivative is,

$$D_{\vec{u}}f(x,y) = \left(-\frac{1}{2}\right)\left(\mathbf{e}^{xy} + xy\mathbf{e}^{xy}\right) + \left(\frac{\sqrt{3}}{2}\right)\left(x^2\mathbf{e}^{xy} + 1\right)$$

Now, plugging in the point in question gives,

$$D_{\vec{u}}f(2,0) = \left(-\frac{1}{2}\right)(1) + \left(\frac{\sqrt{3}}{2}\right)(5) = \frac{5\sqrt{3}-1}{2}$$

(b) $D_{\vec{u}}f(x,y,z)$ where $f(x,y,z) = x^2z + y^3z^2 - xyz$ in the direction of $\vec{v} = \langle -1, 0, 3\rangle$.

In this case let's first check to see if the direction vector is a unit vector or not and if it isn't convert it into one. To do this all we need to do is compute its magnitude.

$$\|\vec{v}\| = \sqrt{1+0+9} = \sqrt{10} \neq 1$$

So, it's not a unit vector. Recall that we can convert any vector into a unit vector that points in the same direction by dividing the vector by its magnitude. So, the unit vector that we need is,

$$\vec{u} = \frac{1}{\sqrt{10}}\langle -1, 0, 3\rangle = \left\langle -\frac{1}{\sqrt{10}}, 0, \frac{3}{\sqrt{10}}\right\rangle$$

The directional derivative is then,

$$\begin{aligned} D_{\vec{u}}f(x,y,z) &= \left(-\frac{1}{\sqrt{10}}\right)(2xz - yz) + (0)\left(3y^2z^2 - xz\right) + \left(\frac{3}{\sqrt{10}}\right)\left(x^2 + 2y^3z - xy\right) \\ &= \frac{1}{\sqrt{10}}\left(3x^2 + 6y^3z - 3xy - 2xz + yz\right) \end{aligned}$$

There is another form of the formula that we used to get the directional derivative that is a little nicer and somewhat more compact. It is also a much more general formula that will encompass both of the formulas above.

Let's start with the second one and notice that we can write it as follows,

$$\begin{aligned} D_{\vec{u}} f(x,y,z) &= f_x(x,y,z)\,a + f_y(x,y,z)\,b + f_z(x,y,z)\,c \\ &= \langle f_x, f_y, f_z \rangle \cdot \langle a, b, c \rangle \end{aligned}$$

In other words, we can write the directional derivative as a dot product and notice that the second vector is nothing more than the unit vector $\vec{u}$ that gives the direction of change. Also, if we had used the version for functions of two variables the third component wouldn't be there, but other than that the formula would be the same.

Now let's give a name and notation to the first vector in the dot product since this vector will show up fairly regularly throughout this course (and in other courses). The **gradient of** f or **gradient vector of** f is defined to be,

$$\nabla f = \langle f_x, f_y, f_z \rangle \qquad \text{or} \qquad \nabla f = \langle f_x, f_y \rangle$$

Or, if we want to use the standard basis vectors the gradient is,

$$\nabla f = f_x \vec{i} + f_y \vec{j} + f_z \vec{k} \qquad \text{or} \qquad \nabla f = f_x \vec{i} + f_y \vec{j}$$

The definition is only shown for functions of two or three variables, however there is a natural extension to functions of any number of variables that we'd like.

With the definition of the gradient we can now say that the directional derivative is given by,

Directional Derivative Gradient Formula

$$D_{\vec{u}} f = \nabla f \cdot \vec{u}$$

where we will no longer show the variable and use this formula for any number of variables. Note as well that we will sometimes use the following notation,

$$D_{\vec{u}} f(\vec{x}) = \nabla f \cdot \vec{u}$$

where $\vec{x} = \langle x, y, z \rangle$ or $\vec{x} = \langle x, y \rangle$ as needed. This notation will be used when we want to note the variables in some way, but don't really want to restrict ourselves to a particular number of variables. In other words, $\vec{x}$ will be used to represent as many variables as we need in the formula and we will most often use this notation when we are already using vectors or vector notation in the problem/formula.

Let's work a couple of examples using this formula of the directional derivative.

Example 2

Find each of the directional derivatives.

(a) $D_{\vec{u}} f(\vec{x})$ for $f(x,y) = x\cos(y)$ in the direction of $\vec{v} = \langle 2, 1\rangle$.

(b) $D_{\vec{u}} f(\vec{x})$ for $f(x,y,z) = \sin(yz) + \ln(x^2)$ at $(1,1,\pi)$ in the direction of $\vec{v} = \langle 1, 1, -1\rangle$.

Solution

(a) $D_{\vec{u}} f(\vec{x})$ for $f(x,y) = x\cos(y)$ in the direction of $\vec{v} = \langle 2, 1\rangle$.

Let's first compute the gradient for this function.

$$\nabla f = \langle \cos(y), -x\sin(y)\rangle$$

Also, as we saw earlier in this section the unit vector for this direction is,

$$\vec{u} = \left\langle \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right\rangle$$

The directional derivative is then,

$$\begin{aligned} D_{\vec{u}} f(\vec{x}) &= \langle \cos(y), -x\sin(y)\rangle \cdot \left\langle \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} \right\rangle \\ &= \frac{1}{\sqrt{5}}\left(2\cos(y) - x\sin(y)\right) \end{aligned}$$

(b) $D_{\vec{u}} f(\vec{x})$ for $f(x,y,z) = \sin(yz) + \ln(x^2)$ at $(1,1,\pi)$ in the direction of $\vec{v} = \langle 1, 1, -1\rangle$.

In this case are asking for the directional derivative at a particular point. To do this we will first compute the gradient, evaluate it at the point in question and then do the dot product. So, let's get the gradient.

$$\begin{aligned} \nabla f(x,y,z) &= \left\langle \frac{2}{x}, z\cos(yz), y\cos(yz) \right\rangle \\ \nabla f(1,1,\pi) &= \left\langle \frac{2}{1}, \pi\cos(\pi), \cos(\pi) \right\rangle = \langle 2, -\pi, -1\rangle \end{aligned}$$

Next, we need the unit vector for the direction,

$$\|\vec{v}\| = \sqrt{3} \qquad \vec{u} = \left\langle \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}} \right\rangle$$

Finally, the directional derivative at the point in question is,

$$\begin{aligned} D_{\vec{u}} f(1,1,\pi) &= \langle 2, -\pi, -1 \rangle \cdot \left\langle \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}} \right\rangle \\ &= \frac{1}{\sqrt{3}}(2 - \pi + 1) \\ &= \frac{3-\pi}{\sqrt{3}} \end{aligned}$$

Before proceeding let's note that the first order partial derivatives that we were looking at in the majority of the section can be thought of as special cases of the directional derivatives. For instance, f_x can be thought of as the directional derivative of f in the direction of $\vec{u} = \langle 1, 0 \rangle$ or $\vec{u} = \langle 1, 0, 0 \rangle$, depending on the number of variables that we're working with. The same can be done for f_y and f_z

We will close out this section with a couple of nice facts about the gradient vector. The first tells us how to determine the maximum rate of change of a function at a point and the direction that we need to move in order to achieve that maximum rate of change.

Maximum Rate of Change

The maximum value of $D_{\vec{u}} f(\vec{x})$ (and hence then the maximum rate of change of the function $f(\vec{x})$) is given by $\|\nabla f(\vec{x})\|$ and will occur in the direction given by $\nabla f(\vec{x})$.

Proof

This is a really simple proof. First, if we start with the dot product form $D_{\vec{u}} f(\vec{x})$ and use a nice fact about dot products as well as the fact that $\vec{u}$ is a unit vector we get,

$$D_{\vec{u}} f = \nabla f \cdot \vec{u} = \|\nabla f\| \; \|\vec{u}\| \cos(\theta) = \|\nabla f\| \cos(\theta)$$

where θ is the angle between the gradient and $\vec{u}$.

Now the largest possible value of $\cos(\theta)$ is 1 which occurs at $\theta = 0$. Therefore the maximum value of $D_{\vec{u}} f(\vec{x})$ is $\|\nabla f(\vec{x})\|$ Also, the maximum value occurs when the angle between the gradient and $\vec{u}$ is zero, or in other words when $\vec{u}$ is pointing in the same direction as the gradient, $\nabla f(\vec{x})$.

Let's take a quick look at an example.

Example 3

Suppose that the height of a hill above sea level is given by $z = 1000 - 0.01x^2 - 0.02y^2$. If you are at the point $(60, 100)$ in what direction is the elevation changing fastest? What is the maximum rate of change of the elevation at this point?

Solution

First, you will hopefully recall from the Quadric Surfaces section that this is an elliptic paraboloid that opens downward. So even though most hills aren't this symmetrical it will at least be vaguely hill shaped and so the question makes at least a little sense.

Now on to the problem. There are a couple of questions to answer here, but using the theorem makes answering them very simple. We'll first need the gradient vector.

$$\nabla f\left(\vec{x}\right) = \langle -0.02x, -0.04y \rangle$$

The maximum rate of change of the elevation will then occur in the direction of

$$\nabla f\left(60, 100\right) = \langle -1.2, -4 \rangle$$

The maximum rate of change of the elevation at this point is,

$$\left\|\nabla f\left(60, 100\right)\right\| = \sqrt{\left(-1.2\right)^2 + \left(-4\right)^2} = \sqrt{17.44} = 4.176$$

Before leaving this example let's note that we're at the point $(60, 100)$ and the direction of greatest rate of change of the elevation at this point is given by the vector $\langle -1.2, -4 \rangle$. Since both of the components are negative it looks like the direction of maximum rate of change points up the hill towards the center rather than away from the hill.

The second fact about the gradient vector that we need to give in this section will be very convenient in some later sections.

Fact

The gradient vector $\nabla f\left(x_0, y_0\right)$ is orthogonal (or perpendicular) to the level curve $f\left(x, y\right) = k$ at the point (x_0, y_0). Likewise, the gradient vector $\nabla f\left(x_0, y_0, z_0\right)$ is orthogonal to the level surface $f\left(x, y, z\right) = k$ at the point (x_0, y_0, z_0).

Proof

We're going to do the proof for the $\mathbb{R}^3$ case. The proof for the $\mathbb{R}^2$ case is identical. We'll also need some notation out of the way to make life easier for us let's let S be the level surface given by $f(x, y, z) = k$ and let $P = (x_0, y_0, z_0)$. Note as well that P will be on S.

Now, let C be any curve on S that contains P. Let $\vec{r}(t) = \langle x(t), y(t), z(t)\rangle$ be the vector equation for C and suppose that t_0 be the value of t such that $\vec{r}(t_0) = \langle x_0, y_0, z_0\rangle$. In other words, t_0 be the value of t that gives P.

Because C lies on S we know that points on C must satisfy the equation for S. Or,

$$f(x(t), y(t), z(t)) = k$$

Next, let's use the Chain Rule on this to get,

$$\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = 0$$

Notice that $\nabla f = \langle f_x, f_y, f_z\rangle$ and $\vec{r}\,'(t) = \langle x'(t), y'(t), z'(t)\rangle$ so this becomes,

$$\nabla f \cdot \vec{r}\,'(t) = 0$$

At, $t = t_0$ this is,

$$\nabla f(x_0, y_0, z_0) \cdot \vec{r}\,'(t_0) = 0$$

This then tells us that the gradient vector at P , $\nabla f(x_0, y_0, z_0)$, is orthogonal to the tangent vector, $\vec{r}\,'(t_0)$, to any curve C that passes through P and on the surface S and so must also be orthogonal to the surface S.

As we will be seeing in later sections we are often going to be needing vectors that are orthogonal to a surface or curve and using this fact we will know that all we need to do is compute a gradient vector and we will get the orthogonal vector that we need. We will see the first application of this in the next chapter.

14 Applications of Partial Derivatives

In this chapter we'll take a look at a couple of applications of partial derivatives. The applications here are either very similar to applications we saw for derivatives of single variable functions or extensions of those applications.

For example we will be looking at the tangent plane to a surface rather than tangent lines to curves as we did with single variable functions.

In addition we be finding relative and absolute extrema of multi-variable functions. The difference in this chapter compared to the last time we saw these applications is that they will often involve a lot more work. Because of the increased difficulty of the problems we'll be restricting ourselves to finding the relative and absolute extrema of functions of two variables only.

We will also be looking at Lagrange Multipliers. This is a method that will allow us to optimize a function that is subject to some constraint. That is to say optimizing a function of two or three variables where the variables must also satisfy some constraint (usually in the form on an equation involving the variables).

14.1 Tangent Planes and Linear Approximations

Earlier we saw how the two partial derivatives f_x and f_y can be thought of as the slopes of traces. We want to extend this idea out a little in this section. The graph of a function $z = f(x, y)$ is a surface in $\mathbb{R}^3$ (three dimensional space) and so we can now start thinking of the plane that is "tangent" to the surface as a point.

Let's start out with a point (x_0, y_0) and let's let C_1 represent the trace to $f(x, y)$ for the plane $y = y_0$ (*i.e.* allowing x to vary with y held fixed) and we'll let C_2 represent the trace to $f(x, y)$ for the plane $x = x_0$ (*i.e.* allowing y to vary with x held fixed). Now, we know that $f_x(x_0, y_0)$ is the slope of the tangent line to the trace C_1 and $f_y(x_0, y_0)$ is the slope of the tangent line to the trace C_2. So, let L_1 be the tangent line to the trace C_1 and let L_2 be the tangent line to the trace C_2.

The tangent plane will then be the plane that contains the two lines L_1 and L_2. Geometrically this plane will serve the same purpose that a tangent line did in Calculus I. A tangent line to a curve was a line that just touched the curve at that point and was "parallel" to the curve at the point in question. Well tangent planes to a surface are planes that just touch the surface at the point and are "parallel" to the surface at the point. Note that this gives us a point that is on the plane. Since the tangent plane and the surface touch at (x_0, y_0) the following point will be on both the surface and the plane.

$$(x_0, y_0, z_0) = \left(x_0, y_0, f(x_0, y_0)\right)$$

What we need to do now is determine the equation of the tangent plane. We know that the general equation of a plane is given by,

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

where (x_0, y_0, z_0) is a point that is on the plane, which we have. Let's rewrite this a little. We'll move the x terms and y terms to the other side and divide both sides by c. Doing this gives,

$$z - z_0 = -\frac{a}{c}(x - x_0) - \frac{b}{c}(y - y_0)$$

Now, let's rename the constants to simplify up the notation a little. Let's rename them as follows,

$$A = -\frac{a}{c} \qquad B = -\frac{b}{c}$$

With this renaming the equation of the tangent plane becomes,

$$z - z_0 = A(x - x_0) + B(y - y_0)$$

and we need to determine values for A and B.

Let's first think about what happens if we hold y fixed, *i.e.* if we assume that $y = y_0$. In this case the equation of the tangent plane becomes,

$$z - z_0 = A(x - x_0)$$

This is the equation of a line and this line must be tangent to the surface at (x_0, y_0) (since it's part of the tangent plane). In addition, this line assumes that $y = y_0$ (*i.e.* fixed) and A is the slope of this line. But if we think about it this is exactly what the tangent to C_1 is, a line tangent to the surface at (x_0, y_0) assuming that $y = y_0$. In other words,

$$z - z_0 = A(x - x_0)$$

is the equation for L_1 and we know that the slope of L_1 is given by $f_x(x_0, y_0)$. Therefore, we have the following,

$$A = f_x(x_0, y_0)$$

If we hold x fixed at $x = x_0$ the equation of the tangent plane becomes,

$$z - z_0 = B(y - y_0)$$

However, by a similar argument to the one above we can see that this is nothing more than the equation for L_2 and that it's slope is B or $f_y(x_0, y_0)$. So,

$$B = f_y(x_0, y_0)$$

The equation of the tangent plane to the surface given by $z = f(x, y)$ at (x_0, y_0) is then,

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

Also, if we use the fact that $z_0 = f(x_0, y_0)$ we can rewrite the equation of the tangent plane as,

$$\begin{aligned} z - f(x_0, y_0) &= f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ z &= f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \end{aligned}$$

We will see an easier derivation of this formula (actually a more general formula) in the next section so if you didn't quite follow this argument hold off until then to see a better derivation.

Example 1

Find the equation of the tangent plane to $z = \ln(2x + y)$ at $(-1, 3)$.

Solution

There really isn't too much to do here other than taking a couple of derivatives and doing some quick evaluations.

$$f(x,y) = \ln(2x+y) \qquad z_0 = f(-1,3) = \ln(1) = 0$$
$$f_x(x,y) = \frac{2}{2x+y} \qquad f_x(-1,3) = 2$$
$$f_y(x,y) = \frac{1}{2x+y} \qquad f_y(-1,3) = 1$$

The equation of the plane is then,

$$\begin{aligned} z - 0 &= 2(x+1) + (1)(y-3) \\ z &= 2x + y - 1 \end{aligned}$$

One nice use of tangent planes is they give us a way to approximate a surface near a point. As long as we are near to the point (x_0, y_0) then the tangent plane should nearly approximate the function at that point. Because of this we define the **linear approximation** to be,

$$L(x,y) = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$

and as long as we are "near" (x_0, y_0) then we should have that,

$$f(x,y) \approx L(x,y) = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$

Example 2

Find the linear approximation to $z = 3 + \frac{x^2}{16} + \frac{y^2}{9}$ at $(-4, 3)$.

Solution

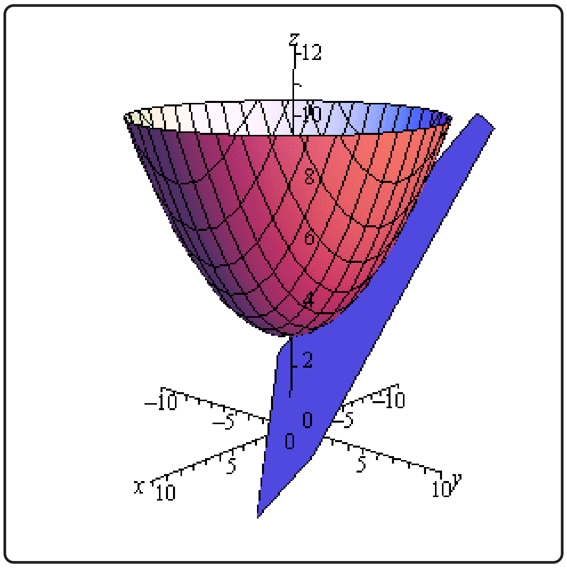
So, we're really asking for the tangent plane so let's find that.

$$f(x,y) = 3 + \frac{x^2}{16} + \frac{y^2}{9} \qquad f(-4,3) = 3 + 1 + 1 = 5$$
$$f_x(x,y) = \frac{x}{8} \qquad f_x(-4,3) = -\frac{1}{2}$$
$$f_y(x,y) = \frac{2y}{9} \qquad f_y(-4,3) = \frac{2}{3}$$

The tangent plane, or linear approximation, is then,

$$L(x,y) = 5 - \frac{1}{2}(x+4) + \frac{2}{3}(y-3)$$

For reference purposes here is a sketch of the surface and the tangent plane/linear approximation.


z
12
10
8
6
4
2
0
−10
−5
0
5
x 10
−10
−5
5
10 y

14.2 Gradient Vector, Tangent Planes and Normal Lines

In this section we want to revisit tangent planes only this time we'll look at them in light of the gradient vector. In the process we will also take a look at a normal line to a surface.

Let's first recall the equation of a plane that contains the point (x_0, y_0, z_0) with normal vector $\vec{n} = \langle a, b, c \rangle$ is given by,

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

When we introduced the gradient vector in the section on directional derivatives we gave the following fact.

Fact

The gradient vector $\nabla f(x_0, y_0)$ is orthogonal (or perpendicular) to the level curve $f(x, y) = k$ at the point (x_0, y_0). Likewise, the gradient vector $\nabla f(x_0, y_0, z_0)$ is orthogonal to the level surface $f(x, y, z) = k$ at the point (x_0, y_0, z_0).

Actually, all we need here is the last part of this fact. This says that the gradient vector is always orthogonal, or *normal*, to the surface at a point.

Also recall that the gradient vector is,

$$\nabla f = \langle f_x, f_y, f_z \rangle$$

So, the tangent plane to the surface given by $f(x, y, z) = k$ at (x_0, y_0, z_0) has the equation,

$$f_x(x_0, y_0, z_0)(x - x_0) + f_y(x_0, y_0, z_0)(y - y_0) + f_z(x_0, y_0, z_0)(z - z_0) = 0$$

This is a much more general form of the equation of a tangent plane than the one that we derived in the previous section.

Note however, that we can also get the equation from the previous section using this more general formula. To see this let's start with the equation $z = f(x, y)$ and we want to find the tangent plane to the surface given by $z = f(x, y)$ at the point (x_0, y_0, z_0) where $z_0 = f(x_0, y_0)$. In order to use the formula above we need to have all the variables on one side. This is easy enough to do. All we need to do is subtract a z from both sides to get,

$$f(x, y) - z = 0$$

Now, if we define a new function

$$F(x, y, z) = f(x, y) - z$$

we can see that the surface given by $z = f(x, y)$ is identical to the surface given by $F(x, y, z) = 0$ and this new equivalent equation is in the correct form for the equation of the tangent plane that we derived in this section.

So, the first thing that we need to do is find the gradient vector for F.

$$\nabla F = \langle F_x, F_y, F_z \rangle = \langle f_x, f_y, -1 \rangle$$

Notice that

$$F_x = \frac{\partial}{\partial x}\Big(f(x,y) - z\Big) = f_x \qquad\qquad F_y = \frac{\partial}{\partial y}\Big(f(x,y) - z\Big) = f_y$$
$$F_z = \frac{\partial}{\partial z}\Big(f(x,y) - z\Big) = -1$$

The equation of the tangent plane is then,

$$f_x(x_0,y_0)(x - x_0) + f_y(x_0,y_0)(y - y_0) - (z - z_0) = 0$$

Or, upon solving for z, we get,

$$z = f(x_0,y_0) + f_x(x_0,y_0)(x - x_0) + f_y(x_0,y_0)(y - y_0)$$

which is identical to the equation that we derived in the previous section.

We can get another nice piece of information out of the gradient vector as well. We might on occasion want a line that is orthogonal to a surface at a point, sometimes called the **normal line**. This is easy enough to get if we recall that the equation of a line only requires that we have a point and a parallel vector. Since we want a line that is at the point (x_0, y_0, z_0) we know that this point must also be on the line and we know that $\nabla f(x_0, y_0, z_0)$ is a vector that is normal to the surface and hence will be parallel to the line. Therefore, the equation of the normal line is,

$$\vec{r}(t) = \langle x_0, y_0, z_0 \rangle + t\,\nabla f(x_0, y_0, z_0)$$

Example 1

Find the tangent plane and normal line to $x^2 + y^2 + z^2 = 30$ at the point $(1, -2, 5)$.

Solution

For this case the function that we're going to be working with is,

$$F(x,y,z) = x^2 + y^2 + z^2$$

and note that we don't have to have a zero on one side of the equal sign. All that we need is a constant. To finish this problem out we simply need the gradient evaluated at the point.

$$\nabla F = \langle 2x, 2y, 2z \rangle$$
$$\nabla F(1,-2,5) = \langle 2, -4, 10 \rangle$$

The tangent plane is then,

$$2\left(x-1\right)-4\left(y+2\right)+10\left(z-5\right)=0$$

The normal line is,

$$\vec{r}\left(t\right)=\langle 1,-2,5\rangle+t\,\langle 2,-4,10\rangle=\langle 1+2t,-2-4t,5+10t\rangle$$

14.3 Relative Minimums and Maximums

In this section we are going to extend one of the more important ideas from Calculus I into functions of two variables. We are going to start looking at trying to find minimums and maximums of functions. This in fact will be the topic of the following two sections as well.

In this section we are going to be looking at identifying relative minimums and relative maximums. Recall as well that we will often use the word extrema to refer to both minimums and maximums.

The definition of relative extrema for functions of two variables is identical to that for functions of one variable we just need to remember now that we are working with functions of two variables. So, for the sake of completeness here is the definition of relative minimums and relative maximums for functions of two variables.

Definition

1. A function $f(x,y)$ has a **relative minimum** at the point (a,b) if $f(x,y) \ge f(a,b)$ for all points (x,y) in some region around (a,b).
2. A function $f(x,y)$ has a **relative maximum** at the point (a,b) if $f(x,y) \le f(a,b)$ for all points (x,y) in some region around (a,b).

Note that this definition does not say that a relative minimum is the smallest value that the function will ever take. It only says that in some region around the point (a,b) the function will always be larger than $f(a,b)$. Outside of that region it is completely possible for the function to be smaller. Likewise, a relative maximum only says that around (a,b) the function will always be smaller than $f(a,b)$. Again, outside of the region it is completely possible that the function will be larger.

Next, we need to extend the idea of **critical points** up to functions of two variables. Recall that a critical point of the function $f(x)$ was a number $x = c$ so that either $f'(c) = 0$ or $f'(c)$ doesn't exist. We have a similar definition for critical points of functions of two variables.

Definition

The point (a,b) is a **critical point** (or a **stationary point**) of $f(x,y)$ provided one of the following is true,

1. $\nabla f(a,b) = \vec{0}$ (this is equivalent to saying that $f_x(a,b) = 0$ and $f_y(a,b) = 0$),
2. $f_x(a,b)$ and/or $f_y(a,b)$ doesn't exist.

To see the equivalence in the first part let's start off with $\nabla f = \vec{0}$ and put in the definition of each part.

$$\nabla f\left(a,b\right) = \vec{0}$$
$$\left\langle f_x\left(a,b\right), f_y\left(a,b\right)\right\rangle = \left\langle 0,0\right\rangle$$

The only way that these two vectors can be equal is to have $f_x\left(a,b\right) = 0$ and $f_y\left(a,b\right) = 0$. In fact, we will use this definition of the critical point more than the gradient definition since it will be easier to find the critical points if we start with the partial derivative definition.

Note as well that BOTH of the first order partial derivatives must be zero at (a,b). If only one of the first order partial derivatives are zero at the point then the point will NOT be a critical point.

We now have the following fact that, at least partially, relates critical points to relative extrema.

Fact

If the point (a,b) is a relative extrema of the function $f\left(x,y\right)$ and the first order derivatives of $f\left(x,y\right)$ exist at (a,b) then (a,b) is also a critical point of $f\left(x,y\right)$ and in fact we'll have $\nabla f\left(a,b\right) = \vec{0}$.

Proof

This is a really simple proof that relies on the single variable version that we saw in Calculus I version, often called Fermat's Theorem.

Let's start off by defining $g\left(x\right) = f\left(x,b\right)$ and suppose that $f\left(x,y\right)$ has a relative extrema at (a,b). However, this also means that $g\left(x\right)$ also has a relative extrema (of the same kind as $f\left(x,y\right)$) at $x = a$. By Fermat's Theorem we then know that $g'\left(a\right) = 0$. But we also know that $g'\left(a\right) = f_x\left(a,b\right)$ and so we have that $f_x\left(a,b\right) = 0$.

If we now define $h\left(y\right) = f\left(a,y\right)$ and going through exactly the same process as above we will see that $f_y\left(a,b\right) = 0$.

So, putting all this together means that $\nabla f\left(a,b\right) = \vec{0}$ and so $f\left(x,y\right)$ has a critical point at (a,b).

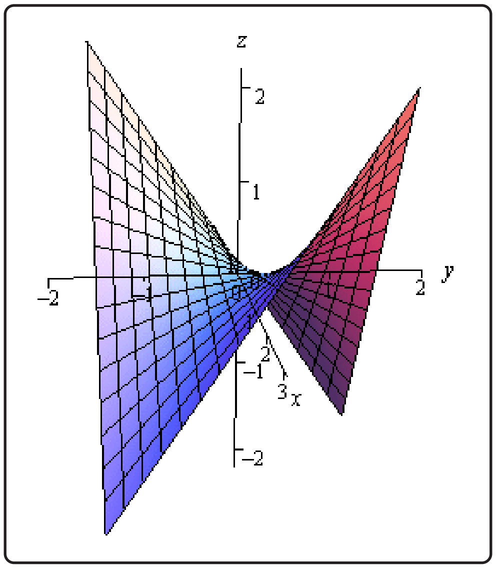
Note that this does NOT say that all critical points are relative extrema. It only says that relative extrema will be critical points of the function. To see this let's consider the function

$$f\left(x,y\right) = xy$$

The two first order partial derivatives are,

$$f_x\left(x,y\right) = y \qquad\qquad f_y\left(x,y\right) = x$$

The only point that will make both of these derivatives zero at the same time is $(0,0)$ and so $(0,0)$ is a critical point for the function. Here is a graph of the function.



Note that the axes are not in the standard orientation here so that we can see more clearly what is happening at the origin, *i.e.* at $(0,0)$. If we start at the origin and move into either of the quadrants where both x and y are the same sign the function increases. However, if we start at the origin and move into either of the quadrants where x and y have the opposite sign then the function decreases. In other words, no matter what region you take about the origin there will be points larger than $f(0,0) = 0$ and points smaller than $f(0,0) = 0$. Therefore, there is no way that $(0,0)$ can be a relative extrema.

Critical points that exhibit this kind of behavior are called **saddle points** The point (a,b) is a **critical point** (or a **stationary point**) of $f(x,y)$ provided one of the following is true, .

While we have to be careful to not misinterpret the results of this fact it is very useful in helping us to identify relative extrema. Because of this fact we know that if we have all the critical points of a function then we also have every possible relative extrema for the function. The fact tells us that all relative extrema must be critical points so we know that if the function does have relative extrema then they must be in the collection of all the critical points. Remember however, that it will be completely possible that at least one of the critical points won't be a relative extrema.

So, once we have all the critical points in hand all we will need to do is test these points to see if they are relative extrema or not. To determine if a critical point is a relative extrema (and in fact to determine if it is a minimum or a maximum) we can use the following fact.

Fact

Suppose that (a, b) is a critical point of $f(x, y)$ and that the second order partial derivatives are continuous in some region that contains (a, b). Next define,

$$D = D(a,b) = f_{xx}(a,b)\, f_{yy}(a,b) - \left[f_{xy}(a,b)\right]^2$$

We then have the following classifications of the critical point.

1. If $D > 0$ and $f_{xx}(a,b) > 0$ then there is a relative minimum at (a, b).
2. If $D > 0$ and $f_{xx}(a,b) < 0$ then there is a relative maximum at (a, b).
3. If $D < 0$ then the point (a, b) is a saddle point.
4. If $D = 0$ then the point (a, b) may be a relative minimum, relative maximum or a saddle point. Other techniques would need to be used to classify the critical point.

Note that if $D > 0$ then both $f_{xx}(a,b)$ and $f_{yy}(a,b)$ will have the same sign and so in the first two cases above we could just as easily replace $f_{xx}(a,b)$ with $f_{yy}(a,b)$. Also note that we aren't going to be seeing any cases in this class where $D = 0$ as these can often be quite difficult to classify. We will be able to classify all the critical points that we find.

Let's see a couple of examples.

Example 1

Find and classify all the critical points of $f(x,y) = 4 + x^3 + y^3 - 3xy$.

Solution

We first need all the first order (to find the critical points) and second order (to classify the critical points) partial derivatives so let's get those.

$$f_x = 3x^2 - 3y \qquad f_y = 3y^2 - 3x$$
$$f_{xx} = 6x \qquad f_{yy} = 6y \qquad f_{xy} = -3$$

Let's first find the critical points. Critical points will be solutions to the system of equations,

$$\begin{aligned} f_x &= 3x^2 - 3y = 0 \\ f_y &= 3y^2 - 3x = 0 \end{aligned}$$

This is a non-linear system of equations and these can, on occasion, be difficult to solve. However, in this case it's not too bad. We can solve the first equation for y as follows,

$$3x^2 - 3y = 0 \qquad \Rightarrow \qquad y = x^2$$

Plugging this into the second equation gives,

$$3\left(x^2\right)^2 - 3x = 3x\left(x^3 - 1\right) = 0$$

From this we can see that we must have $x = 0$ or $x = 1$. Now use the fact that $y = x^2$ to get the critical points.

$$\begin{aligned} x = 0: &\quad y = 0^2 = 0 \quad \Rightarrow &\quad (0,0) \\ x = 1: &\quad y = 1^2 = 1 \quad \Rightarrow &\quad (1,1) \end{aligned}$$

So, we get two critical points. All we need to do now is classify them. To do this we will need D. Here is the general formula for D.

$$\begin{aligned} D(x,y) &= f_{x\,x}(x,y)\, f_{y\,y}(x,y) - \left[f_{x\,y}(x,y)\right]^2 \\ &= (6x)(6y) - (-3)^2 \\ &= 36xy - 9 \end{aligned}$$

To classify the critical points all that we need to do is plug in the critical points and use the fact above to classify them.

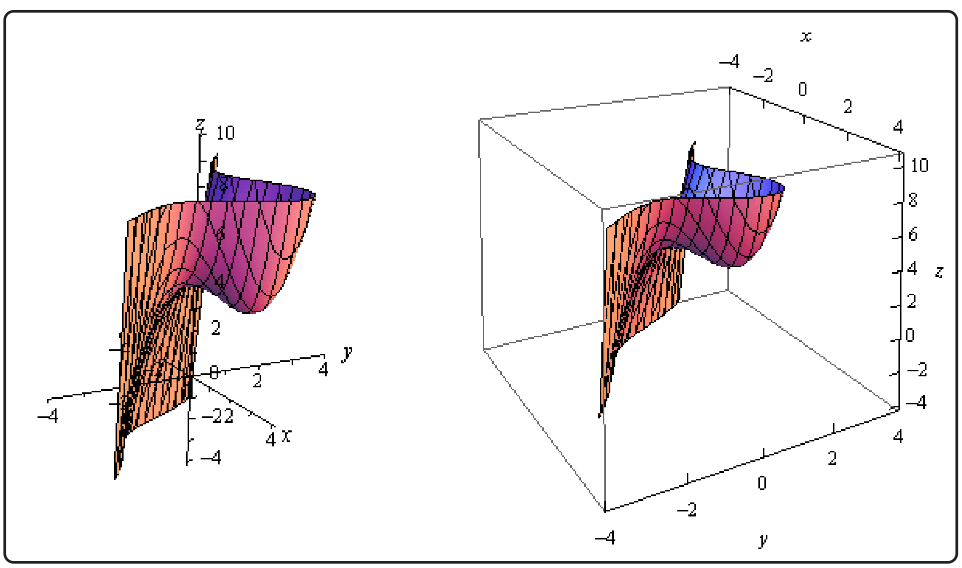
$(0,0):\ D = D(0,0) = -9 < 0$

So, for $(0,0)$ D is negative and so this must be a saddle point.

$(1,1):\ D = D(1,1) = 36 - 9 = 27 > 0 \qquad f_{x\,x}(1,1) = 6 > 0$

For $(1,1)$ D is positive and $f_{x\,x}$ is positive and so we must have a relative minimum.

For the sake of completeness here is a graph of this function.



Notice that in order to get a better visual we used a somewhat nonstandard orientation. We can see that there is a relative minimum at $(1,1)$ and (hopefully) it's clear that at $(0,0)$ we do get a saddle point.

Example 2

Find and classify all the critical points for $f(x,y) = 3x^2y + y^3 - 3x^2 - 3y^2 + 2$

Solution

As with the first example we will first need to get all the first and second order derivatives.

$$f_x = 6xy - 6x \qquad f_y = 3x^2 + 3y^2 - 6y$$
$$f_{x\,x} = 6y - 6 \qquad f_{y\,y} = 6y - 6 \qquad f_{x\,y} = 6x$$

We'll first need the critical points. The equations that we'll need to solve this time are,

$$\begin{aligned} 6xy - 6x &= 0 \\ 3x^2 + 3y^2 - 6y &= 0 \end{aligned}$$

These equations are a little trickier to solve than the first set, but once you see what to do they really aren't terribly bad.

First, let's notice that we can factor out a 6 x from the first equation to get,

$$6x(y-1) = 0$$

So, we can see that the first equation will be zero if $x = 0$ or $y = 1$. Be careful to not just cancel the x from both sides. If we had done that we would have missed $x = 0$.

To find the critical points we can plug these (individually) into the second equation and solve for the remaining variable.

$x = 0$:

$$3y^2 - 6y = 3y(y-2) = 0 \qquad \Rightarrow \qquad y = 0,\ y = 2$$

$y = 1$:

$$3x^2 - 3 = 3\left(x^2 - 1\right) = 0 \qquad \Rightarrow \qquad x = -1,\ x = 1$$

So, if $x = 0$ we have the following critical points,

$$(0,0) \qquad (0,2)$$

and if $y = 1$ the critical points are,

$$(1,1) \qquad (-1,1)$$

Now all we need to do is classify the critical points. To do this we'll need the general formula for D.

$$D\left(x,y\right) = \left(6y-6\right)\left(6y-6\right) - \left(6x\right)^2 = \left(6y-6\right)^2 - 36x^2$$

$\left(0,0\right):\ D = D\left(0,0\right) = 36 > 0 \qquad f_{x\,x}\left(0,0\right) = -6 < 0$ $\left(0,2\right):\ D = D\left(0,2\right) = 36 > 0 \qquad f_{x\,x}\left(0,2\right) = 6 > 0$ $\left(1,1\right):\ D = D\left(1,1\right) = -36 < 0$ $\left(-1,1\right):\ D = D\left(-1,1\right) = -36 < 0$

So, it looks like we have the following classification of each of these critical points.

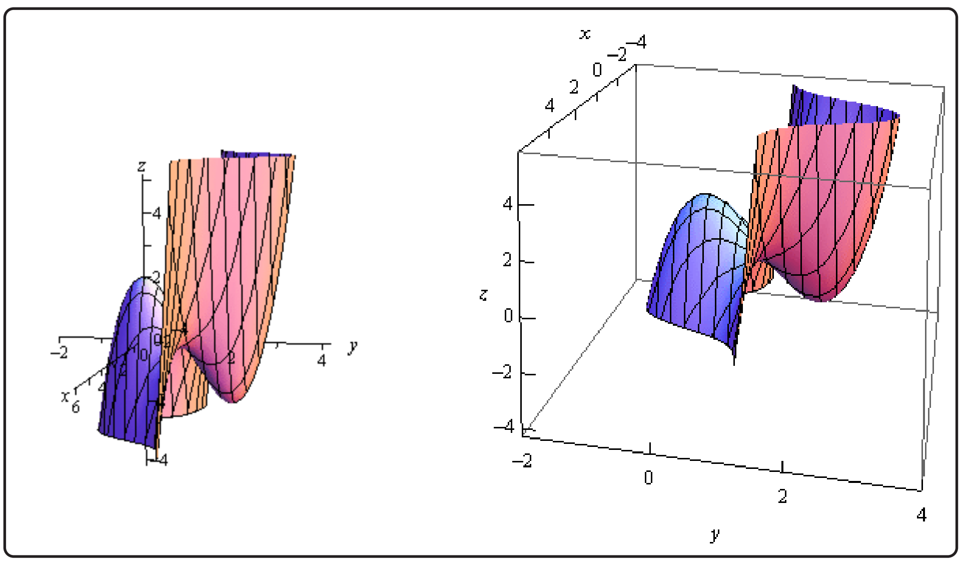
$\left(0,0\right)$: Relative Maximum

$\left(0,2\right)$: Relative Minimum

$\left(1,1\right)$: Saddle Point

$\left(-1,1\right)$: Saddle Point

Here is a graph of the surface for the sake of completeness.



Let's do one more example that is a little different from the first two.

Example 3

Determine the point on the plane $4x-2y+z=1$ that is closest to the point $(-2,-1,5)$.

Solution

Note that we are NOT asking for the critical points of the plane. In order to do this example we are going to need to first come up with the equation that we are going to have to work with.

First, let's suppose that (x,y,z) is any point on the plane. The distance between this point and the point in question, $(-2,-1,5)$, is given by the formula,

$$d=\sqrt{(x+2)^2+(y+1)^2+(z-5)^2}$$

What we are then asked to find is the minimum value of this equation. The point (x,y,z) that gives the minimum value of this equation will be the point on the plane that is closest to $(-2,-1,5)$.

There are a couple of issues with this equation. First, it is a function of x, y and z and we can only deal with functions of x and y at this point. However, this is easy to fix. We can solve the equation of the plane to see that,

$$z=1-4x+2y$$

Plugging this into the distance formula gives,

$$\begin{aligned} d &= \sqrt{(x+2)^2+(y+1)^2+(1-4x+2y-5)^2} \\ &= \sqrt{(x+2)^2+(y+1)^2+(-4-4x+2y)^2} \end{aligned}$$

Now, the next issue is that there is a square root in this formula and we know that we're going to be differentiating this eventually. So, in order to make our life a little easier let's notice that finding the minimum value of d will be equivalent to finding the minimum value of d^2.

So, let's instead find the minimum value of

$$f(x,y)=d^2=(x+2)^2+(y+1)^2+(-4-4x+2y)^2$$

Now, we need to be a little careful here. We are being asked to find the closest point on the plane to $(-2,-1,5)$ and that is not really the same thing as what we've been doing in this section. In this section we've been finding and classifying critical points as relative minimums or maximums and what we are really asking is to find the smallest value the function will take, or the absolute minimum. Hopefully, it does make sense from a physical standpoint that there will be a closest point on the plane to $(-2, -1, 5)$. This point should also be a relative minimum in addition to being an absolute minimum.

So, let's go through the process from the first and second example and see what we get as far as relative minimums go. If we only get a single relative minimum then we will be done since that point will also need to be the absolute minimum of the function and hence the point on the plane that is closest to $(-2, -1, 5)$.

We'll need the derivatives first.

$$\begin{aligned} f_x &= 2(x+2) + 2(-4)(-4-4x+2y) = 36 + 34x - 16y \\ f_y &= 2(y+1) + 2(2)(-4-4x+2y) = -14 - 16x + 10y \\ f_{xx} &= 34 \\ f_{yy} &= 10 \\ f_{xy} &= -16 \end{aligned}$$

Now, before we get into finding the critical point let's compute D quickly.

$$D = 34(10) - (-16)^2 = 84 > 0$$

So, in this case D will always be positive and also notice that $f_{xx} = 34 > 0$ is always positive and so any critical points that we get will be guaranteed to be relative minimums.

Now let's find the critical point(s). This will mean solving the system.

$$\begin{aligned} 36 + 34x - 16y &= 0 \\ -14 - 16x + 10y &= 0 \end{aligned}$$

To do this we can solve the first equation for x.

$$x = \frac{1}{34}(16y - 36) = \frac{1}{17}(8y - 18)$$

Now, plug this into the second equation and solve for y.

$$-14 - \frac{16}{17}(8y - 18) + 10y = 0 \qquad \Rightarrow \qquad y = -\frac{25}{21}$$

Back substituting this into the equation for x gives $x = -\frac{34}{21}$.

So, it looks like we get a single critical point : $\left(-\frac{34}{21}, -\frac{25}{21}\right)$. Also, since we know this will be a relative minimum and it is the only critical point we know that this is also the x and y coordinates of the point on the plane that we're after. We can find the z coordinate by plugging into the equation of the plane as follows,

$$z = 1 - 4\left(-\frac{34}{21}\right) + 2\left(-\frac{25}{21}\right) = \frac{107}{21}$$

So, the point on the plane that is closest to $(-2, -1, 5)$ is $\left(-\frac{34}{21}, -\frac{25}{21}, \frac{107}{21}\right)$.

14.4 Absolute Extrema

In this section we are going to extend the work from the previous section. In the previous section we were asked to find and classify all critical points as relative minimums, relative maximums and/or saddle points. In this section we want to optimize a function, that is identify the absolute minimum and/or the absolute maximum of the function, on a given region in $\mathbb{R}^2$. Note that when we say we are going to be working on a region in $\mathbb{R}^2$ we mean that we're going to be looking at some region in the *xy*-plane.

In order to optimize a function in a region we are going to need to get a couple of definitions out of the way and a fact. Let's first get the definitions out of the way.

Definitions

1. A region in $\mathbb{R}^2$ is called **closed** if it includes its boundary. A region is called **open** if it doesn't include any of its boundary points.
2. A region in $\mathbb{R}^2$ is called **bounded** if it can be completely contained in a disk. In other words, a region will be bounded if it is finite.

Let's think a little more about the definition of closed. We said a region is closed if it includes its boundary. Just what does this mean? Let's think of a rectangle. Below are two definitions of a rectangle, one is closed and the other is open.

Open	Closed
$-5 < x < 3$	$-5 \le x \le 3$
$1 < y < 6$	$1 \le y \le 6$

In this first case we don't allow the ranges to include the endpoints (*i.e.* we aren't including the edges of the rectangle) and so we aren't allowing the region to include any points on the edge of the rectangle. In other words, we aren't allowing the region to include its boundary and so it's open.

In the second case we are allowing the region to contain points on the edges and so will contain its entire boundary and hence will be closed.

This is an important idea because of the following fact.

Extreme Value Theorem

If $f(x, y)$ is continuous in some closed, bounded set D in $\mathbb{R}^2$ then there are points in D, (x_1, y_1) and (x_2, y_2) so that $f(x_1, y_1)$ is the absolute maximum and $f(x_2, y_2)$ is the absolute minimum of the function in D.

Note that this theorem does NOT tell us where the absolute minimum or absolute maximum will occur. It only tells us that they will exist. Note as well that the absolute minimum and/or absolute maximum may occur in the interior of the region or it may occur on the boundary of the region.

The basic process for finding absolute maximums is pretty much identical to the process that we used in Calculus I when we looked at finding absolute extrema of functions of single variables. There will however, be some procedural changes to account for the fact that we now are dealing with functions of two variables. Here is the process.

Finding Absolute Extrema

1. Find all the critical points of the function that lie in the region D and determine the function value at each of these points.
2. Find all extrema of the function on the boundary. This usually involves the Calculus I approach for this work.
3. The largest and smallest values found in the first two steps are the absolute minimum and the absolute maximum of the function.

The main difference between this process and the process that we used in Calculus I is that the "boundary" in Calculus I was just two points and so there really wasn't a lot to do in the second step. For these problems the majority of the work is often in the second step as we will often end up doing a Calculus I absolute extrema problem one or more times.

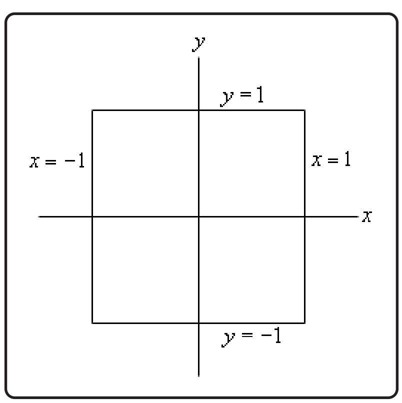
Let's take a look at an example or two.

Example 1

Find the absolute minimum and absolute maximum of $f(x, y) = x^2 + 4y^2 - 2x^2y + 4$ on the rectangle given by $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$.

Solution

Let's first get a quick picture of the rectangle for reference purposes.



The boundary of this rectangle is given by the following conditions.

$$\begin{aligned} &\text{right side :} \quad x = 1, \ -1 \le y \le 1 \\ &\text{left side :} \quad x = -1, \ -1 \le y \le 1 \\ &\text{upper side :} \quad y = 1, \ -1 \le x \le 1 \\ &\text{lower side :} \quad y = -1, \ -1 \le x \le 1 \end{aligned}$$

These will be important in the second step of our process.

We'll start this off by finding all the critical points that lie inside the given rectangle. To do this we'll need the two first order derivatives.

$$f_x = 2x - 4xy \qquad f_y = 8y - 2x^2$$

Note that since we aren't going to be classifying the critical points we don't need the second order derivatives. To find the critical points we will need to solve the system,

$$\begin{aligned} 2x - 4xy &= 0 \\ 8y - 2x^2 &= 0 \end{aligned}$$

We can solve the second equation for y to get,

$$y = \frac{x^2}{4}$$

Plugging this into the first equation gives us,

$$2x - 4x\left(\frac{x^2}{4}\right) = 2x - x^3 = x\left(2 - x^2\right) = 0$$

This tells us that we must have $x = 0$ or $x = \pm\sqrt{2} = \pm 1.414....$ Now, recall that we only want critical points in the region that we're given. That means that we only want critical points for which $-1 \le x \le 1$. The only value of x that will satisfy this is the first one so we can ignore the last two for this problem. Note however that a simple change to the boundary would include these two so don't forget to always check if the critical points are in the region (or on the boundary since that can also happen).

Plugging $x = 0$ into the equation for y gives us,

$$y = \frac{0^2}{4} = 0$$

The single critical point, in the region (and again, that's important), is $(0, 0)$. We now need to get the value of the function at the critical point.

$$f(0, 0) = 4$$

Eventually we will compare this to values of the function found in the next step and take the largest and smallest as the absolute extrema of the function in the rectangle.

Now we have reached the long part of this problem. We need to find the absolute extrema of the function along the boundary of the rectangle. What this means is that we're going to need to look at what the function is doing along each of the sides of the rectangle listed above.

Let's first take a look at the right side. As noted above the right side is defined by

$$x = 1, \ -1 \le y \le 1$$

Notice that along the right side we know that $x = 1$. Let's take advantage of this by defining a new function as follows,

$$g(y) = f(1, y) = 1^2 + 4y^2 - 2\left(1^2\right)y + 4 = 5 + 4y^2 - 2y$$

Now, finding the absolute extrema of $f(x, y)$ along the right side will be equivalent to finding the absolute extrema of $g(y)$ in the range $-1 \le y \le 1$. Hopefully you recall how to do this from Calculus I. We find the critical points of $g(y)$ in the range $-1 \le y \le 1$ and then evaluate $g(y)$ at the critical points and the end points of the range of y's.

Let's do that for this problem.

$$g'(y) = 8y - 2 \qquad \Rightarrow \qquad y = \frac{1}{4}$$

This is in the range and so we will need the following function evaluations.

$$g(-1) = 11 \qquad g(1) = 7 \qquad g\left(\frac{1}{4}\right) = \frac{19}{4} = 4.75$$

Notice that, using the definition of $g(y)$ these are also function values for $f(x, y)$.

$$\begin{aligned} g(-1) &= f(1, -1) = 11 \\ g(1) &= f(1, 1) = 7 \\ g\left(\frac{1}{4}\right) &= f\left(1, \frac{1}{4}\right) = \frac{19}{4} = 4.75 \end{aligned}$$

We can now do the left side of the rectangle which is defined by,

$$x = -1, \ -1 \le y \le 1$$

Again, we'll define a new function as follows,

$$g(y) = f(-1, y) = (-1)^2 + 4y^2 - 2(-1)^2 y + 4 = 5 + 4y^2 - 2y$$

Notice however that, for this boundary, this is the same function as we looked at for the right side. This will not always happen, but since it has let's take advantage of the fact that we've already done the work for this function. We know that the critical point is $y = \frac{1}{4}$ and we know that the function value at the critical point and the end points are,

$$g(-1) = 11 \qquad g(1) = 7 \qquad g\left(\frac{1}{4}\right) = \frac{19}{4} = 4.75$$

The only real difference here is that these will correspond to values of $f(x, y)$ at different points than for the right side. In this case these will correspond to the following function values for $f(x, y)$.

$$\begin{aligned} g(-1) &= f(-1, -1) = 11 \\ g(1) &= f(-1, 1) = 7 \\ g\left(\frac{1}{4}\right) &= f\left(-1, \frac{1}{4}\right) = \frac{19}{4} = 4.75 \end{aligned}$$

We can now look at the upper side defined by,

$$y = 1, \ -1 \le x \le 1$$

We'll again define a new function except this time it will be a function of x.

$$h(x) = f(x, 1) = x^2 + 4\left(1^2\right) - 2x^2(1) + 4 = 8 - x^2$$

We need to find the absolute extrema of $h(x)$ on the range $-1 \le x \le 1$. First find the critical point(s).

$$h'(x) = -2x \qquad \Rightarrow \qquad x = 0$$

The value of this function at the critical point and the end points is,

$$h(-1) = 7 \qquad h(1) = 7 \qquad h(0) = 8$$

and these in turn correspond to the following function values for $f(x, y)$

$$\begin{aligned} h(-1) &= f(-1, 1) = 7 \\ h(1) &= f(1, 1) = 7 \\ h(0) &= f(0, 1) = 8 \end{aligned}$$

Note that there are several "repeats" here. The first two function values have already been computed when we looked at the right and left side. This will often happen.

Finally, we need to take care of the lower side. This side is defined by,

$$y = -1, \ -1 \le x \le 1$$

The new function we'll define in this case is,

$$h(x) = f(x,-1) = x^2 + 4(-1)^2 - 2x^2(-1) + 4 = 8 + 3x^2$$

The critical point for this function is,

$$h'(x) = 6x \qquad \Rightarrow \qquad x = 0$$

The function values at the critical point and the endpoint are,

$$h(-1) = 11 \qquad h(1) = 11 \qquad h(0) = 8$$

and the corresponding values for $f(x,y)$ are,

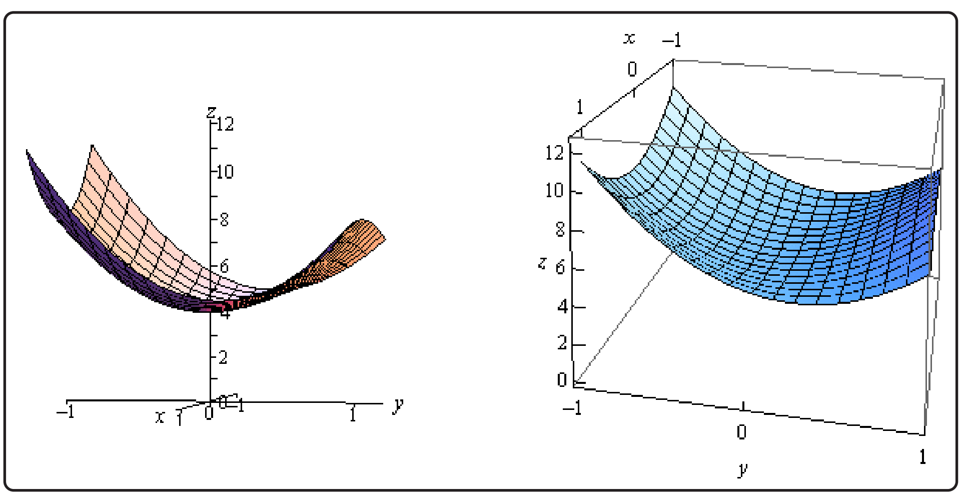
$$\begin{aligned} h(-1) &= f(-1,-1) = 11 \\ h(1) &= f(1,-1) = 11 \\ h(0) &= f(0,-1) = 8 \end{aligned}$$

The final step to this (long...) process is to collect up all the function values for $f(x,y)$ that we've computed in this problem. Here they are,

$$\begin{array}{ccc} f(0,0) = 4 & f(1,-1) = 11 & f(1,1) = 7 \\ f\left(1,\frac{1}{4}\right) = 4.75 & f(-1,1) = 7 & f(-1,-1) = 11 \\ f\left(-1,\frac{1}{4}\right) = 4.75 & f(0,1) = 8 & f(0,-1) = 8 \end{array}$$

The absolute minimum is at $(0,0)$ since gives the smallest function value and the absolute maximum occurs at $(1,-1)$ and $(-1,-1)$ since these two points give the largest function value.

Here is a sketch of the function on the rectangle for reference purposes.



As this example has shown these can be very long problems on occasion. Let's take a look at an easier, well shorter anyway, problem with a different kind of boundary.

Example 2

Find the absolute minimum and absolute maximum of $f(x,y) = 2x^2 - y^2 + 6y$ on the disk of radius 4, $x^2 + y^2 \le 16$

Solution

First note that a disk of radius 4 is given by the inequality in the problem statement. The "less than" inequality is included to get the interior of the disk and the equal sign is included to get the boundary. Of course, this also means that the boundary of the disk is a circle of radius 4.

Let's first find the critical points of the function that lies inside the disk. This will require the following two first order partial derivatives.

$$f_x = 4x \qquad f_y = -2y + 6$$

To find the critical points we'll need to solve the following system.

$$\begin{aligned} 4x &= 0 \\ -2y + 6 &= 0 \end{aligned}$$

This is actually a fairly simple system to solve however. The first equation tells us that $x = 0$ and the second tells us that $y = 3$. So, the only critical point for this function is $(0,3)$ and this is inside the disk of radius 4. The function value at this critical point is,

$$f(0,3) = 9$$

Now we need to look at the boundary. This one will be somewhat different from the previous example. In this case we don't have fixed values of x and y on the boundary. Instead we have,

$$x^2 + y^2 = 16$$

We can solve this for x^2 and plug this into the x^2 in $f(x,y)$ to get a function of y as follows.

$$x^2 = 16 - y^2$$
$$g(y) = 2\left(16 - y^2\right) - y^2 + 6y = 32 - 3y^2 + 6y$$

We will need to find the absolute extrema of this function on the range $-4 \le y \le 4$ (this is the range of y's for the disk....). We'll first need the critical points of this function.

$$g'(y) = -6y + 6 \qquad \Rightarrow \qquad y = 1$$

The value of this function at the critical point and the endpoints are,

$$g(-4) = -40 \qquad g(4) = 8 \qquad g(1) = 35$$

Unlike the first example we will still need to find the values of x that correspond to these. We can do this by plugging the value of y into our equation for the circle and solving for x.

$$\begin{aligned} y &= -4: & x^2 &= 16 - 16 = 0 & &\Rightarrow & x &= 0 \\ y &= 4: & x^2 &= 16 - 16 = 0 & &\Rightarrow & x &= 0 \\ y &= 1: & x^2 &= 16 - 1 = 15 & &\Rightarrow & x &= \pm\sqrt{15} = \pm 3.87 \end{aligned}$$

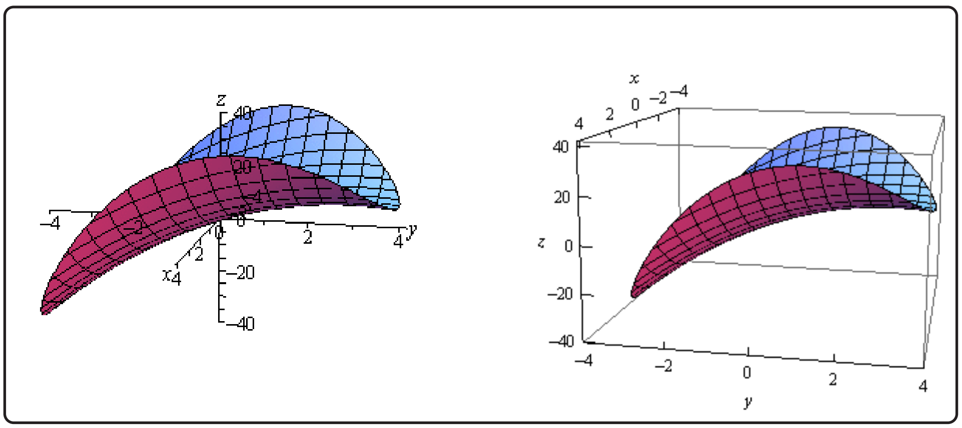
The function values for $g(y)$ then correspond to the following function values for $f(x,y)$.

$$\begin{aligned} g(-4) &= -40 & &\Rightarrow & & f(0,-4) = -40 \\ g(4) &= 8 & &\Rightarrow & & f(0,4) = 8 \\ g(1) &= 35 & &\Rightarrow & & f\left(-\sqrt{15},1\right) = 35 \text{ and } f\left(\sqrt{15},1\right) = 35 \end{aligned}$$

Note that the third one actually corresponded to two different values for $f(x,y)$ since that y also produced two different values of x.

So, comparing these values to the value of the function at the critical point of $f(x,y)$ that we found earlier we can see that the absolute minimum occurs at $(0,-4)$ while the absolute maximum occurs twice at $\left(-\sqrt{15},1\right)$ and $\left(\sqrt{15},1\right)$.

Here is a sketch of the region for reference purposes.



In both of these examples one of the absolute extrema actually occurred at more than one place. Sometimes this will happen and sometimes it won't so don't read too much into the fact that it happened in both examples given here.

Also note that, as we've seen, absolute extrema will often occur on the boundaries of these regions, although they don't have to occur at the boundaries. Had we given much more complicated examples with multiple critical points we would have seen examples where the absolute extrema occurred interior to the region and not on the boundary.

14.5 Lagrange Multipliers

In the previous section we optimized (*i.e.* found the absolute extrema) a function on a region that contained its boundary. Finding potential optimal points in the interior of the region isn't too bad in general, all that we needed to do was find the critical points and plug them into the function. However, as we saw in the examples finding potential optimal points on the boundary was often a fairly long and messy process.

In this section we are going to take a look at another way of optimizing a function subject to given constraint(s). The constraint(s) may be the equation(s) that describe the boundary of a region although in this section we won't concentrate on those types of problems since this method just requires a general constraint and doesn't really care where the constraint came from.

So, let's get things set up. We want to optimize (*i.e.* find the minimum and maximum value of) a function, $f(x,y,z)$, subject to the constraint $g(x,y,z) = k$. Again, the constraint may be the equation that describes the boundary of a region or it may not be. The process is actually fairly simple, although the work can still be a little overwhelming at times.

Method of Lagrange Multipliers

1. Solve the following system of equations.

$$\begin{aligned} \nabla f(x,y,z) &= \lambda \, \nabla g(x,y,z) \\ g(x,y,z) &= k \end{aligned}$$

2. Plug in all solutions, (x,y,z), from the first step into $f(x,y,z)$ and identify the minimum and maximum values, provided they exist and $\nabla g \neq \vec{0}$ at the point.

The constant, λ, is called the **Lagrange Multiplier**.

Notice that the system of equations from the method actually has four equations, we just wrote the system in a simpler form. To see this let's take the first equation and put in the definition of the gradient vector to see what we get.

$$\langle f_x, f_y, f_z \rangle = \lambda \langle g_x, g_y, g_z \rangle = \langle \lambda g_x, \lambda g_y, \lambda g_z \rangle$$

In order for these two vectors to be equal the individual components must also be equal. So, we actually have three equations here.

$$f_x = \lambda g_x \qquad f_y = \lambda g_y \qquad f_z = \lambda g_z$$

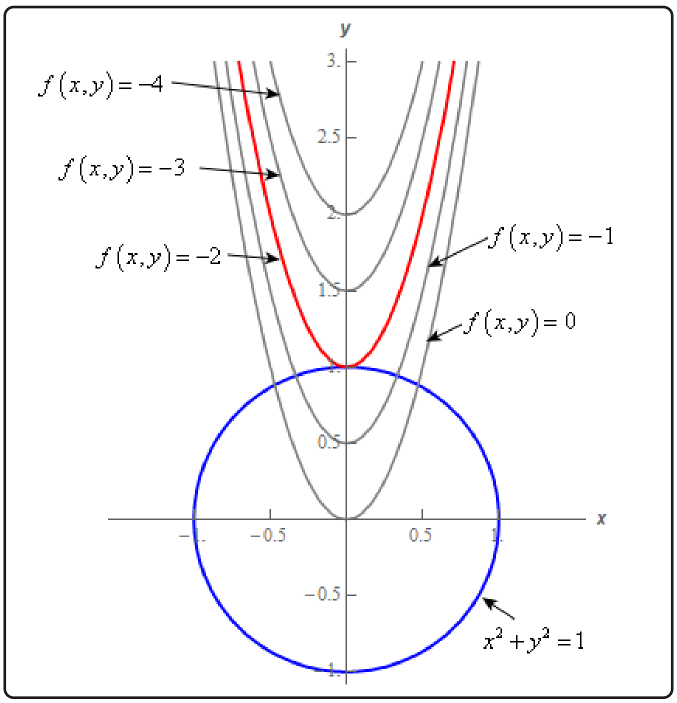
These three equations along with the constraint, $g(x,y,z) = c$, give four equations with four unknowns x, y, z, and λ.

Note as well that if we only have functions of two variables then we won't have the third component of the gradient and so will only have three equations in three unknowns x, y, and λ.

As a final note we also need to be careful with the fact that in some cases minimums and maximums won't exist even though the method will seem to imply that they do. In every problem we'll need to make sure that minimums and maximums will exist before we start the problem.

To see a physical justification for the formulas above. Let's consider the minimum and maximum value of $f(x,y) = 8x^2 - 2y$ subject to the constraint $x^2 + y^2 = 1$. In the practice problems for this section (problem #2 to be exact) we will show that minimum value of $f(x,y)$ is -2 which occurs at $(0,1)$ and the maximum value of $f(x,y)$ is 8.125 which occurs at $\left(-\frac{3\sqrt{7}}{8}, -\frac{1}{8}\right)$ and $\left(\frac{3\sqrt{7}}{8}, -\frac{1}{8}\right)$.

Here is a sketch of the constraint as well as $f(x.y) = k$ for various values of k.



First remember that solutions to the system must be somewhere on the graph of the constraint, $x^2+y^2 = 1$ in this case. Because we are looking for the minimum/maximum value of $f(x,y)$ this, in turn, means that the location of the minimum/maximum value of $f(x,y)$, *i.e.* the point (x,y), must occur where the graph of $f(x,y) = k$ intersects the graph of the constraint when k is either the minimum or maximum value of $f(x,y)$.

Now, we can see that the graph of $f(x,y) = -2$, *i.e.* the graph of the minimum value of $f(x,y)$, just touches the graph of the constraint at $(0,1)$. In fact, the two graphs at that point are tangent.

If the two graphs are tangent at that point then their normal vectors must be parallel, *i.e.* the two normal vectors must be scalar multiples of each other. Mathematically, this means,

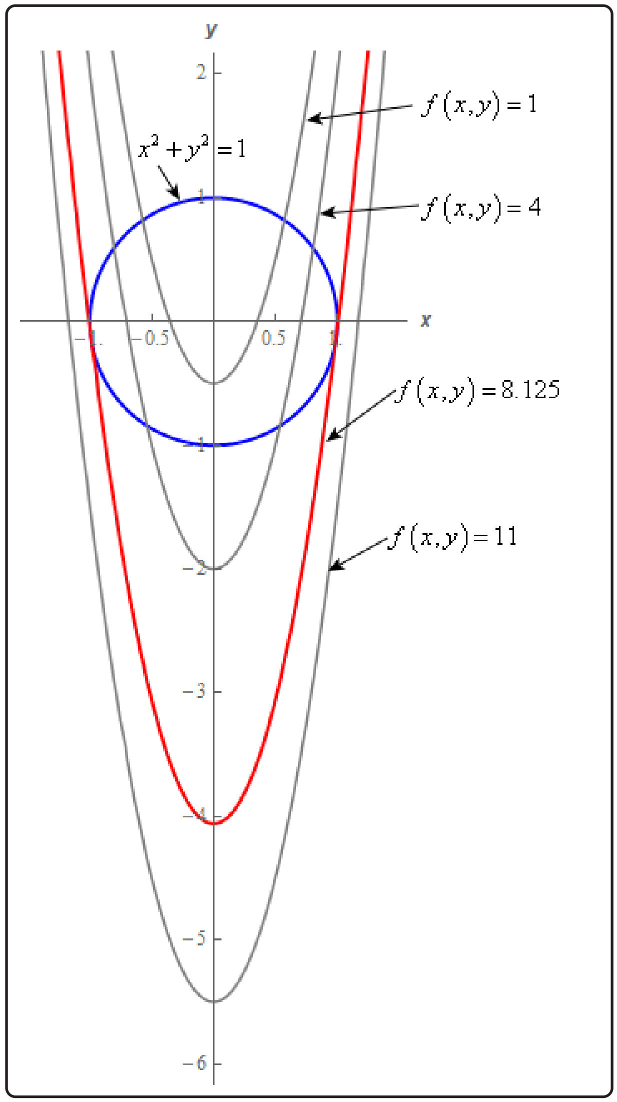
$$\nabla f(x,y,z) = \lambda \, \nabla g(x,y,z)$$

for some scalar λ and this is exactly the first equation in the system we need to solve in the method.

Note as well that if k is smaller than the minimum value of $f(x,y)$ the graph of $f(x,y) = k$ doesn't intersect the graph of the constraint and so it is not possible for the function to take that value of k at a point that will satisfy the constraint.

Likewise, if k is larger than the minimum value of $f(x,y)$ the graph of $f(x,y) = k$ will intersect the graph of the constraint but the two graphs are not tangent at the intersection point(s). This means that the method will not find those intersection points as we solve the system of equations.

Next, the graph to the right shows a different set of values of k. In this case, the values of k include the maximum value of $f(x,y)$ as well as a few values on either side of the maximum value.



Again, we can see that the graph of $f(x,y) = 8.125$ will just touch the graph of the constraint at two points. This is a good thing as we know the solution does say that it should occur at two points. Also note that at those points again the graph of $f(x,y) = 8.125$ and the constraint are tangent and so, just as with the minimum values, the normal vectors must be parallel at these points.

Likewise, for value of k greater than 8.125 the graph of $f(x,y) = k$ does not intersect the graph of the constraint and so it will not be possible for $f(x,y)$ to take on those larger values at points that are on the constraint.

Also, for values of k less than 8.125 the graph of $f(x,y) = k$ does intersect the graph of the constraint but will not be tangent at the intersection points and so again the method will not produce these intersection points as we solve the system of equations.

So, with these graphs we've seen that the minimum/maximum values of $f(x,y)$ will come where the graph of $f(x,y) = k$ and the graph of the constraint are tangent and so their normal vectors are parallel. Also, because the point must occur on the constraint itself. In other words, the system of equations we need to solve to determine the minimum/maximum value of $f(x,y)$ are exactly those given in the above when we introduced the method.

Note that the physical justification above was done for a two dimensional system but the same justification can be done in higher dimensions. The difference is that in higher dimensions we won't be working with curves. For example, in three dimensions we would be working with surfaces. However, the same ideas will still hold. At the points that give minimum and maximum value(s) of the surfaces would be parallel and so the normal vectors would also be parallel.

Let's work a couple of examples.

Example 1

Find the dimensions of the box with largest volume if the total surface area is 64 cm^2.

Solution

Before we start the process here note that we also saw a way to solve this kind of problem in Calculus I, except in those problems we required a condition that related one of the sides of the box to the other sides so that we could get down to a volume and surface area function that only involved two variables. We no longer need this condition for these problems.

Now, let's get on to solving the problem. We first need to identify the function that we're going to optimize as well as the constraint. Let's set the length of the box to be x, the width of the box to be y and the height of the box to be z. Let's also note that because we're dealing with the dimensions of a box it is safe to assume that x, y, and z are all positive quantities.

We want to find the largest volume and so the function that we want to optimize is given by,

$$f\left(x, y, z\right) = xyz$$

Next, we know that the surface area of the box must be a constant 64. So this is the constraint. The surface area of a box is simply the sum of the areas of each of the sides so the constraint is given by,

$$2xy + 2xz + 2yz = 64 \qquad \Rightarrow \qquad xy + xz + yz = 32$$

Note that we divided the constraint by 2 to simplify the equation a little. Also, we get the function $g\left(x, y, z\right)$ from this.

$$g\left(x, y, z\right) = xy + xz + yz$$

The function itself, $f\left(x, y, z\right) = xyz$ will clearly have neither minimums or maximums unless we put some restrictions on the variables. The only real restriction that we've got is that all the variables must be positive. This, of course, instantly means that the function does have a minimum, zero, even though this is a silly value as it also means we pretty much don't have a box. It does however mean that we know the minimum of $f\left(x, y, z\right)$ does exist.

So, let's now see if $f(x, y, z)$ will have a maximum. Clearly, hopefully, $f(x, y, z)$ will not have a maximum if all the variables are allowed to increase without bound. That however, can't happen because of the constraint,

$$xy + xz + yz = 32$$

Here we've got the sum of three positive numbers (remember that we x, y, and z are positive because we are working with a box) and the sum must equal 32. So, if one of the variables gets very large, say x, then because each of the products must be less than 32 both y and z must be very small to make sure the first two terms are less than 32. So, there is no way for all the variables to increase without bound and so it should make some sense that the function, $f(x, y, z) = xyz$, will have a maximum.

This is not an exact proof that $f(x, y, z)$ will have a maximum but it should help to visualize that $f(x, y, z)$ should have a maximum value as long as it is subject to the constraint.

Here are the four equations that we need to solve.

$$yz = \lambda(y + z) \qquad (f_x = \lambda g_x) \tag{14.1}$$

$$xz = \lambda(x + z) \qquad (f_y = \lambda g_y) \tag{14.2}$$

$$xy = \lambda(x + y) \qquad (f_z = \lambda g_z) \tag{14.3}$$

$$xy + xz + yz = 32 \qquad (g(x, y, z) = 32) \tag{14.4}$$

There are many ways to solve this system. We'll solve it in the following way. Let's multiply equation Equation 14.1 by x, equation Equation 14.2 by y and equation Equation 14.3 by z. This gives,

$$xyz = \lambda x(y + z) \tag{14.5}$$

$$xyz = \lambda y(x + z) \tag{14.6}$$

$$xyz = \lambda z(x + y) \tag{14.7}$$

Now notice that we can set equations Equation 14.5 and Equation 14.6 equal. Doing this gives,

$$\begin{aligned} \lambda x(y + z) &= \lambda y(x + z) \\ \lambda(xy + xz) - \lambda(yx + yz) &= 0 \\ \lambda(xz - yz) &= 0 \qquad \Rightarrow \qquad \lambda = 0 \quad \text{or} \quad xz = yz \end{aligned}$$

This gave two possibilities. The first, $\lambda = 0$ is not possible since if this was the case equation Equation 14.1 would reduce to

$$yz = 0 \qquad \Rightarrow \qquad y = 0 \ \text{ or } \ z = 0$$

Since we are talking about the dimensions of a box neither of these are possible so we can discount $\lambda = 0$. This leaves the second possibility.

$$xz = yz$$

Since we know that $z \neq 0$ (again since we are talking about the dimensions of a box) we can cancel the z from both sides. This gives,

$$x = y \tag{14.8}$$

Next, let's set equations Equation 14.6 and Equation 14.7 equal. Doing this gives,

$$\begin{aligned} \lambda y\left(x+z\right) &= \lambda z\left(x+y\right) \\ \lambda\left(yx+yz-zx-zy\right) &= 0 \\ \lambda\left(yx-zx\right) &= 0 \qquad \Rightarrow \qquad \lambda = 0 \text{ or } yx = zx \end{aligned}$$

As already discussed we know that $\lambda = 0$ won't work and so this leaves,

$$yx = zx$$

We can also say that $x \neq 0$ since we are dealing with the dimensions of a box so we must have,

$$z = y \tag{14.9}$$

Plugging equations Equation 14.8 and Equation 14.9 into equation Equation 14.4 we get,

$$y^2 + y^2 + y^2 = 3y^2 = 32 \qquad y = \pm\sqrt{\frac{32}{3}} = \pm 3.266$$

However, we know that y must be positive since we are talking about the dimensions of a box. Therefore, the only solution that makes physical sense here is

$$x = y = z = 3.266$$

So, it looks like we've got a cube.

We should be a little careful here. Since we've only got one solution we might be tempted to assume that these are the dimensions that will give the largest volume. Anytime we get a single solution we really need to verify that it is a maximum (or minimum if that is what we are looking for).

This is actually pretty simple to do. First, let's note that the volume at our solution above is,

$$V = f\left(\sqrt{\frac{32}{3}}, \sqrt{\frac{32}{3}}, \sqrt{\frac{32}{3}}\right) = \left(\sqrt{\frac{32}{3}}\right)^3 = 34.8376$$

Now, we know that a maximum of $f(x, y, z)$ will exist ("proved" that earlier in the solution) and so to verify that that this really is a maximum all we need to do if find another set of dimensions that satisfy our constraint and check the volume. If the volume of this new set of dimensions is smaller that the volume above then we know that our solution does give a maximum.

If, on the other hand, the new set of dimensions give a larger volume we have a problem. We only have a single solution and we know that a maximum exists and the method should generate that maximum. So, in this case, the likely issue is that we will have made a mistake somewhere and we'll need to go back and find it.

So, let's find a new set of dimensions for the box. The only thing we need to worry about is that they will satisfy the constraint. Outside of that there aren't other constraints on the size of the dimensions. So, we can freely pick two values and then use the constraint to determine the third value.

Let's choose $x = y = 1$. No reason for these values other than they are "easy" to work with. Plugging these into the constraint gives,

$$1 + z + z = 32 \qquad \rightarrow \qquad 2z = 31 \qquad \rightarrow \qquad z = \frac{31}{2}$$

So, this is a set of dimensions that satisfy the constraint and the volume for this set of dimensions is,

$$V = f\left(1, 1, \frac{31}{2}\right) = \frac{31}{2} = 15.5 < 34.8376$$

So, the new dimensions give a smaller volume and so our solution above is, in fact, the dimensions that will give a maximum volume of the box are $x = y = z = 3.266$

Notice that we never actually found values for λ in the above example. This is fairly standard for these kinds of problems. The value of λ isn't really important to determining if the point is a maximum or a minimum so often we will not bother with finding a value for it. On occasion we will need its value to help solve the system, but even in those cases we won't use it past finding the point.

Example 2

Find the maximum and minimum of $f(x, y) = 5x - 3y$ subject to the constraint $x^2 + y^2 = 136$.

Solution

This one is going to be a little easier than the previous one since it only has two variables.

Also, note that it's clear from the constraint that region of possible solutions lies on a disk of radius $\sqrt{136}$ which is a closed and bounded region, $-\sqrt{136} \leq x, y \leq \sqrt{136}$, and hence by the Extreme Value Theorem we know that a minimum and maximum value must exist.

Here is the system that we need to solve.

$$\begin{aligned} 5 &= 2\lambda x \\ -3 &= 2\lambda y \\ x^2 + y^2 &= 136 \end{aligned}$$

Notice that, as with the last example, we can't have $\lambda = 0$ since that would not satisfy the first two equations. So, since we know that $\lambda \neq 0$ we can solve the first two equations for x and y respectively. This gives,

$$x = \frac{5}{2\lambda} \qquad\qquad y = -\frac{3}{2\lambda}$$

Plugging these into the constraint gives,

$$\frac{25}{4\lambda^2} + \frac{9}{4\lambda^2} = \frac{17}{2\lambda^2} = 136$$

We can solve this for λ.

$$\lambda^2 = \frac{1}{16} \qquad \Rightarrow \qquad \lambda = \pm\frac{1}{4}$$

Now, that we know λ we can find the points that will be potential maximums and/or minimums.

If $\lambda = -\frac{1}{4}$ we get,

$$x = -10 \qquad\qquad y = 6$$

and if $\lambda = \frac{1}{4}$ we get,

$$x = 10 \qquad\qquad y = -6$$

To determine if we have maximums or minimums we just need to plug these into the function. Also recall from the discussion at the start of this solution that we know these will be the minimum and maximums because the Extreme Value Theorem tells us that minimums and maximums will exist for this problem.

Here are the minimum and maximum values of the function.

$$\begin{aligned} &f(-10, 6) = -68 && \text{Minimum at } (-10, 6) \\ &f(10, -6) = 68 && \text{Maximum at } (10, -6) \end{aligned}$$

In the first two examples we've excluded $\lambda = 0$ either for physical reasons or because it wouldn't solve one or more of the equations. Do not always expect this to happen. Sometimes we will be able to automatically exclude a value of λ and sometimes we won't.

Let's take a look at another example.

Example 3

Find the maximum and minimum values of $f(x, y, z) = xyz$ subject to the constraint $x + y + z = 1$. Assume that $x, y, z \geq 0$.

Solution

First note that our constraint is a sum of three positive or zero number and it must be 1. Therefore, it is clear that our solution will fall in the range $0 \leq x, y, z \leq 1$ and so the solution must lie in a closed and bounded region and so by the Extreme Value Theorem we know that a minimum and maximum value must exist.

Here is the system of equation that we need to solve.

$$yz = \lambda \tag{14.10}$$

$$xz = \lambda \tag{14.11}$$

$$xy = \lambda \tag{14.12}$$

$$x + y + z = 1 \tag{14.13}$$

Let's start this solution process off by noticing that since the first three equations all have λ they are all equal. So, let's start off by setting equations Equation 14.10 and Equation 14.11 equal.

$$yz = xz \quad \Rightarrow \quad z(y - x) = 0 \quad \Rightarrow \quad z = 0 \text{ or } y = x$$

So, we've got two possibilities here. Let's start off with by assuming that $z = 0$. In this case we can see from either equation Equation 14.10 or Equation 14.11 that we must then have $\lambda = 0$. From equation Equation 14.12 we see that this means that $xy = 0$. This in turn means that either $x = 0$ or $y = 0$.

So, we've got two possible cases to deal with there. In each case two of the variables must be zero. Once we know this we can plug into the constraint, equation Equation 14.13, to find the remaining value.

$$\begin{aligned} z = 0,\ x = 0 \quad &\Rightarrow \quad y = 1 \\ z = 0,\ y = 0 \quad &\Rightarrow \quad x = 1 \end{aligned}$$

So, we've got two possible solutions $(0, 1, 0)$ and $(1, 0, 0)$.

Now let's go back and take a look at the other possibility, $y = x$. We also have two possible cases to look at here as well.

This first case is $x = y = 0$. In this case we can see from the constraint that we must have $z = 1$ and so we now have a third solution $(0, 0, 1)$.

The second case is $x = y \neq 0$. Let's set equations Equation 14.11 and Equation 14.12 equal.

$$xz = xy \qquad \Rightarrow \qquad x(z - y) = 0 \qquad \Rightarrow \qquad x = 0 \text{ or } z = y$$

Now, we've already assumed that $x \neq 0$ and so the only possibility is that $z = y$. However, this also means that,

$$x = y = z$$

Using this in the constraint gives,

$$3x = 1 \qquad \Rightarrow \qquad x = \frac{1}{3}$$

So, the next solution is $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

We got four solutions by setting the first two equations equal.

To completely finish this problem out we should probably set equations Equation 14.10 and Equation 14.12 equal as well as setting equations Equation 14.11 and Equation 14.12 equal to see what we get. Doing this gives,

$$\begin{aligned} yz = xy \qquad &\Rightarrow \qquad y(z - x) = 0 \qquad \Rightarrow \qquad y = 0 \text{ or } z = x \\ xz = xy \qquad &\Rightarrow \qquad x(z - y) = 0 \qquad \Rightarrow \qquad x = 0 \text{ or } z = y \end{aligned}$$

Both of these are very similar to the first situation that we looked at and we'll leave it up to you to show that in each of these cases we arrive back at the four solutions that we already found.

So, we have four solutions that we need to check in the function to see whether we have minimums or maximums.

$$\begin{aligned} &f(0,0,1) = 0 \qquad f(0,1,0) = 0 \qquad f(1,0,0) = 0 \qquad && \text{All Minimums} \\ &f\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = \frac{1}{27} && \text{Maximum} \end{aligned}$$

So, in this case the maximum occurs only once while the minimum occurs three times.

Note as well that we never really used the assumption that $x, y, z \geq 0$ in the actual solution to the problem. We used it to make sure that we had a closed and bounded region to guarantee we would have absolute extrema. To see why this is important let's take a look at what might happen without this assumption Without this assumption it wouldn't be too difficult to find points that give both larger and smaller values of the functions. For example.

$$\begin{aligned} x = -100,\ y = 100,\ z = 1 \ &: \ -100 + 100 + 1 = 1 \qquad f(-100, 100, 1) = -10000 \\ x = -50,\ y = -50,\ z = 101 \ &: \ -50 - 50 + 101 = 1 \qquad f(-50, -50, 101) = 252500 \end{aligned}$$

With these examples you can clearly see that it's not too hard to find points that will give larger and smaller function values. However, all of these examples required negative values of x, y and/or z to make sure we satisfy the constraint. By eliminating these we will know that we've got minimum and maximum values by the Extreme Value Theorem.

Before we proceed we need to address a quick issue that the last example illustrates about the method of Lagrange Multipliers. We found the absolute minimum and maximum to the function. However, what we did not find is all the locations for the absolute minimum. For example, assuming $x, y, z \geq 0$, consider the following sets of points.

$$\begin{array}{lll} (0, y, z) & \text{where} & y + z = 1 \\ (x, 0, z) & \text{where} & x + z = 1 \\ (x, y, 0) & \text{where} & x + y = 1 \end{array}$$

Every point in this set of points will satisfy the constraint from the problem and in every case the function will evaluate to zero and so also give the absolute minimum.

So, what is going on? Recall from the previous section that we had to check both the critical points and the boundaries to make sure we had the absolute extrema. The same was true in Calculus I. We had to check both critical points and end points of the interval to make sure we had the absolute extrema.

It turns out that we really need to do the same thing here if we want to know that we've found all the locations of the absolute extrema. The method of Lagrange multipliers will find the absolute extrema, it just might not find all the locations of them as the method does not take the end points of variables ranges into account (note that we might luck into some of these points but we can't guarantee that).

So, after going through the Lagrange Multiplier method we should then ask what happens at the end points of our variable ranges. For the example that means looking at what happens if $x = 0$, $y = 0$, $z = 0$, $x = 1$, $y = 1$, and $z = 1$. In the first three cases we get the points listed above that do happen to also give the absolute minimum. For the later three cases we can see that if one of the variables are 1 the other two must be zero (to meet the constraint) and those were actually found in the example. Sometimes that will happen and sometimes it won't.

In the case of this example the end points of each of the variable ranges gave absolute extrema but there is no reason to expect that to happen every time. In Example 2 above, for example, the end points of the ranges for the variables do not give absolute extrema (we'll let you verify this).

The moral of this is that if we want to know that we have every location of the absolute extrema for a particular problem we should also check the end points of any variable ranges that we might have. If all we are interested in is the value of the absolute extrema then there is no reason to do this.

Okay, it's time to move on to a slightly different topic. To this point we've only looked at constraints that were equations. We can also have constraints that are inequalities. The process for these types of problems is nearly identical to what we've been doing in this section to this point. The main difference between the two types of problems is that we will also need to find all the critical points that satisfy the inequality in the constraint and check these in the function when we check the values we found using Lagrange Multipliers.

Let's work an example to see how these kinds of problems work.

Example 4

Find the maximum and minimum values of $f(x,y) = 4x^2 + 10y^2$ on the disk $x^2 + y^2 \le 4$.

Solution

Note that the constraint here is the inequality for the disk. Because this is a closed and bounded region the Extreme Value Theorem tells us that a minimum and maximum value must exist.

The first step is to find all the critical points that are in the disk (*i.e.* satisfy the constraint). This is easy enough to do for this problem. Here are the two first order partial derivatives.

$$\begin{aligned} f_x &= 8x & &\Rightarrow & 8x &= 0 & &\Rightarrow & x &= 0 \\ f_y &= 20y & &\Rightarrow & 20y &= 0 & &\Rightarrow & y &= 0 \end{aligned}$$

So, the only critical point is $(0,0)$ and it does satisfy the inequality.

At this point we proceed with Lagrange Multipliers and we treat the constraint as an equality instead of the inequality. We only need to deal with the inequality when finding the critical points.

So, here is the system of equations that we need to solve.

$$\begin{aligned} 8x &= 2\lambda x \\ 20y &= 2\lambda y \\ x^2 + y^2 &= 4 \end{aligned}$$

From the first equation we get,

$$2x(4-\lambda) = 0 \qquad \Rightarrow \qquad x = 0 \;\; \text{or} \;\; \lambda = 4$$

If we have $x = 0$ then the constraint gives us $y = \pm 2$.

If we have $\lambda = 4$ the second equation gives us,

$$20y = 8y \qquad \Rightarrow \qquad y = 0$$

The constraint then tells us that $x = \pm 2$.

If we'd performed a similar analysis on the second equation we would arrive at the same points.

So, Lagrange Multipliers gives us four points to check : $(0, 2)$, $(0, -2)$, $(2, 0)$, and $(-2, 0)$.

To find the maximum and minimum we need to simply plug these four points along with the critical point in the function.

$$\begin{aligned} f(0,0) &= 0 && \text{Minimum} \\ f(2,0) &= f(-2,0) = 16 && \\ f(0,2) &= f(0,-2) = 40 && \text{Maximum} \end{aligned}$$

In this case, the minimum was interior to the disk and the maximum was on the boundary of the disk.

The final topic that we need to discuss in this section is what to do if we have more than one constraint. We will look only at two constraints, but we can naturally extend the work here to more than two constraints.

We want to optimize $f(x, y, z)$ subject to the constraints $g(x, y, z) = c$ and $h(x, y, z) = k$. The system that we need to solve in this case is,

$$\begin{aligned} \nabla f(x,y,z) &= \lambda \nabla g(x,y,z) + \mu \nabla h(x,y,z) \\ g(x,y,z) &= c \\ h(x,y,z) &= k \end{aligned}$$

So, in this case we get two Lagrange Multipliers. Also, note that the first equation really is three equations as we saw in the previous examples. Let's see an example of this kind of optimization problem.

Example 5

Find the maximum and minimum of $f(x, y, z) = 4y - 2z$ subject to the constraints $2x - y - z = 2$ and $x^2 + y^2 = 1$.

Solution

Verifying that we will have a minimum and maximum value here is a little trickier. Clearly, because of the second constraint we've got to have $-1 \le x, y \le 1$. With this in mind there must also be a set of limits on z in order to make sure that the first constraint is met. If one really wanted to determine that range you could find the minimum and maximum values of $2x - y$ subject to $x^2 + y^2 = 1$ and you could then use this to determine the minimum and maximum values of z. We won't do that here. The point is only to acknowledge that once again the possible solutions must lie in a closed and bounded region and so minimum and maximum values must exist by the Extreme Value Theorem.

Here is the system of equations that we need to solve.

$$0 = 2\lambda + 2\mu x \qquad (f_x = \lambda g_x + \mu h_x) \tag{14.14}$$

$$4 = -\lambda + 2\mu y \qquad (f_y = \lambda g_y + \mu h_y) \tag{14.15}$$

$$-2 = -\lambda \qquad (f_z = \lambda g_z + \mu h_z) \tag{14.16}$$

$$2x - y - z = 2 \tag{14.17}$$

$$x^2 + y^2 = 1 \tag{14.18}$$

First, let's notice that from equation Equation 14.16 we get $\lambda = 2$. Plugging this into equation Equation 14.14 and equation Equation 14.15 and solving for x and y respectively gives,

$$0 = 4 + 2\mu x \qquad \Rightarrow \qquad x = -\frac{2}{\mu} \tag{14.19}$$

$$4 = -2 + 2\mu y \qquad \Rightarrow \qquad y = \frac{3}{\mu} \tag{14.20}$$

Now, plug these into equation Equation 14.18.

$$\frac{4}{\mu^2} + \frac{9}{\mu^2} = \frac{13}{\mu^2} = 1 \qquad \Rightarrow \qquad \mu = \pm\sqrt{13}$$

So, we have two cases to look at here. First, let's see what we get when $\mu = \sqrt{13}$. In this case we know that,

$$x = -\frac{2}{\sqrt{13}} \qquad y = \frac{3}{\sqrt{13}}$$

Plugging these into equation Equation 14.17 gives,

$$-\frac{4}{\sqrt{13}} - \frac{3}{\sqrt{13}} - z = 2 \qquad \Rightarrow \qquad z = -2 - \frac{7}{\sqrt{13}}$$

So, we've got one solution.

Let's now see what we get if we take $\mu = -\sqrt{13}$. Here we have,

$$x = \frac{2}{\sqrt{13}} \qquad y = -\frac{3}{\sqrt{13}}$$

Plugging these into equation Equation 14.17 gives,

$$\frac{4}{\sqrt{13}} + \frac{3}{\sqrt{13}} - z = 2 \qquad \Rightarrow \qquad z = -2 + \frac{7}{\sqrt{13}}$$

and there's a second solution.

Now all that we need to is check the two solutions in the function to see which is the maximum and which is the minimum.

$$f\left(-\frac{2}{\sqrt{13}}, \frac{3}{\sqrt{13}}, -2-\frac{7}{\sqrt{13}}\right) = 4+\frac{26}{\sqrt{13}} = 11.2111$$
$$f\left(\frac{2}{\sqrt{13}}, -\frac{3}{\sqrt{13}}, -2+\frac{7}{\sqrt{13}}\right) = 4-\frac{26}{\sqrt{13}} = -3.2111$$

So, we have a maximum at $\left(-\frac{2}{\sqrt{13}}, \frac{3}{\sqrt{13}}, -2-\frac{7}{\sqrt{13}}\right)$ and a minimum at $\left(\frac{2}{\sqrt{13}}, -\frac{3}{\sqrt{13}}, -2+\frac{7}{\sqrt{13}}\right)$.

15 Multiple Integrals

We now need to start discussing integration of multi-variable functions.

When we looked at definite integrals of single variable functions the values of the independent variable were in some interval $[a, b]$. For functions of multiple variables the values of the independent variables will not just come from intervals anymore. For functions of two variables, for example, the values of the independent variables will come from a two dimensional region. Likewise, for functions of three variables the values of the independent variables will come from a three dimensional region.

We will be discussing Double Integrals (for integrating functions of two variables) and Triple Integrals (for integrating functions of three variables). While most of the integration will be done in terms of Cartesian coordinates we will also discuss converting integrals from Cartesian coordinates into Polar coordinates (for functions of two variables) and Cylindrical or Spherical coordinates (for functions of three variables).

We will also formalize the process for converting an integral from one coordinate system into another. In the process we will derive some of the formulas that were using to convert integrals from Cartesian into Polar, Cylindrical or Spherical coordinates.

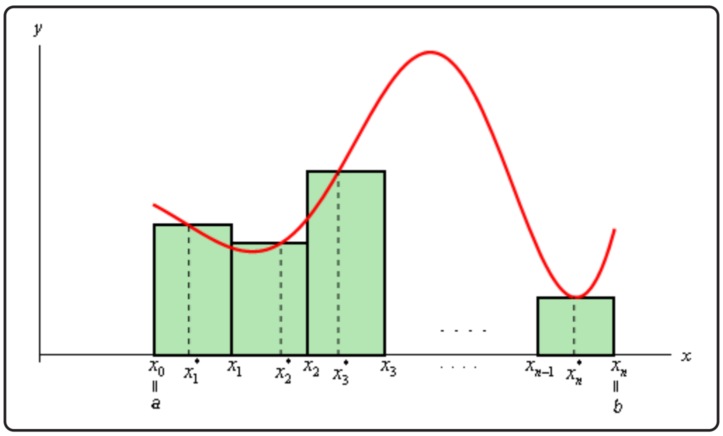
15.1 Double Integrals

Before starting on double integrals let's do a quick review of the definition of definite integrals for functions of single variables. First, when working with the integral,

$$\int_a^b f(x)\, dx$$

we think of x's as coming from the interval $a \le x \le b$. For these integrals we can say that we are integrating over the interval $a \le x \le b$. Note that this does assume that $a < b$, however, if we have $b < a$ then we can just use the interval $b \le x \le a$.

Now, when we derived the definition of the definite integral we first thought of this as an area problem. We first asked what the area under the curve was and to do this we broke up the interval $a \le x \le b$ into n subintervals of width Δx and choose a point, x_i^*, from each interval as shown below,



Each of the rectangles has height of $f(x_i^*)$ and we could then use the area of each of these rectangles to approximate the area as follows.

$$A \approx f(x_1^*)\,\Delta x + f(x_2^*)\,\Delta x + \cdots + f(x_i^*)\,\Delta x + \cdots + f(x_n^*)\,\Delta x$$

To get the exact area we then took the limit as n goes to infinity and this was also the definition of the definite integral.

$$\int_a^b f(x)\, dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\,\Delta x$$

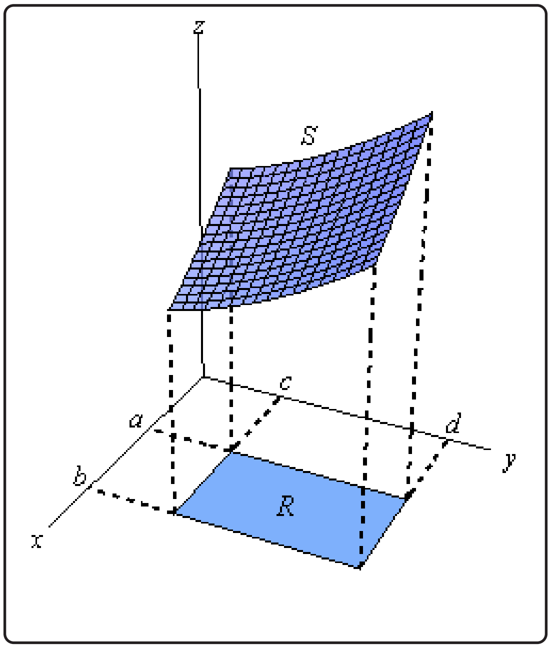
In this section we want to integrate a function of two variables, $f(x, y)$. With functions of one variable we integrated over an interval (*i.e.* a one-dimensional space) and so it makes some sense then that when integrating a function of two variables we will integrate over a region of $\mathbb{R}^2$ (two-dimensional space).

We will start out by assuming that the region in $\mathbb{R}^2$ is a rectangle which we will denote as follows,

$$R = [a, b] \times [c, d]$$

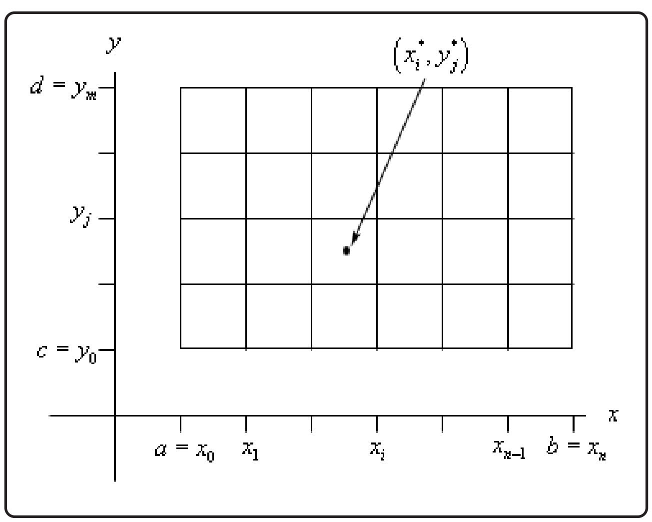
This means that the ranges for x and y are $a \leq x \leq b$ and $c \leq y \leq d$.

Also, we will initially assume that $f(x, y) \geq 0$ although this doesn't really have to be the case. Let's start out with the graph of the surface S given by graphing $f(x, y)$ over the rectangle R.

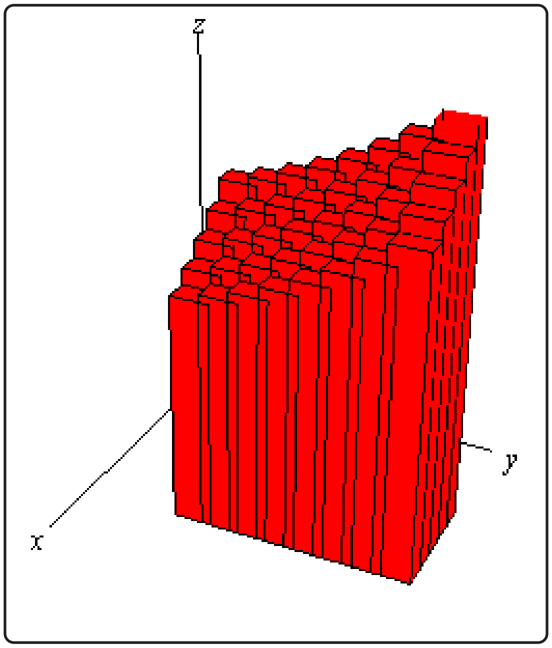


Now, just like with functions of one variable let's not worry about integrals quite yet. Let's first ask what the volume of the region under S (and above the *xy*-plane of course) is.

We will approximate the volume much as we approximated the area above. We will first divide up $a \leq x \leq b$ into n subintervals and divide up $c \leq y \leq d$ into m subintervals. This will divide up R into a series of smaller rectangles and from each of these we will choose a point $\left(x_i^*, y_j^*\right)$. Here is a sketch of this set up.



Now, over each of these smaller rectangles we will construct a box whose height is given by $f\left(x_i^*, y_j^*\right)$. Here is a sketch of that.



Each of the rectangles has a base area of ΔA and a height of $f\left(x_i^*, y_j^*\right)$ so the volume of each of these boxes is $f\left(x_i^*, y_j^*\right)\,\Delta A$. The volume under the surface S is then approximately,

$$V \approx \sum_{i=1}^{n}\sum_{j=1}^{m} f\left(x_i^*, y_j^*\right)\,\Delta A$$

We will have a *double* sum since we will need to add up volumes in both the x and y directions.

To get a better estimation of the volume we will take n and m larger and larger and to get the exact volume we will need to take the limit as both n and m go to infinity. In other words,

$$V = \lim_{n,\, m \to \infty} \sum_{i=1}^{n}\sum_{j=1}^{m} f\left(x_i^*, y_j^*\right)\,\Delta A$$

Now, this should look familiar. This looks a lot like the definition of the integral of a function of single variable. In fact, this is also the definition of a double integral, or more exactly an integral of a function of two variables over a rectangle.

Here is the official definition of a double integral of a function of two variables over a rectangular region R as well as the notation that we'll use for it.

$$\iint\limits_R f\left(x,y\right)\,dA = \lim_{n,\, m \to \infty} \sum_{i=1}^{n}\sum_{j=1}^{m} f\left(x_i^*, y_j^*\right)\,\Delta A$$

Note the similarities and differences in the notation to single integrals. We have two integrals to denote the fact that we are dealing with a two dimensional region and we have a differential here as well. Note that the differential is dA instead of the dx and dy that we're used to seeing. Note as well that we don't have limits on the integrals in this notation. Instead we have the R written below the two integrals to denote the region that we are integrating over.

As indicated above one interpretation of the double integral of $f(x, y)$ over the rectangle R is the volume under the function $f(x, y)$ (and above the *xy*-plane). Or,

Fact

$$\text{Volume} = \iint\limits_R f(x, y)\, dA$$

We can use this double sum in the definition to estimate the value of a double integral if we need to. We can do this by choosing $\left(x_i^*, y_j^*\right)$ to be the midpoint of each rectangle. When we do this we usually denote the point as $\left(\overline{x}_i, \overline{y}_j\right)$. This leads to the **Midpoint Rule**,

$$\iint\limits_R f(x, y)\, dA \approx \sum_{i=1}^{n} \sum_{j=1}^{m} f\left(\overline{x}_i, \overline{y}_j\right) \Delta A$$

In the next section we start looking at how to actually compute double integrals.

15.2 Iterated Integrals

In the previous section we gave the definition of the double integral. However, just like with the definition of a single integral the definition is very difficult to use in practice and so we need to start looking into how we actually compute double integrals. We will continue to assume that we are integrating over the rectangle

$$R = [a, b] \times [c, d]$$

We will look at more general regions in the next section.

The following theorem tells us how to compute a double integral over a rectangle.

Fubini's Theorem

If $f(x, y)$ is continuous on $R = [a, b] \times [c, d]$ then,

$$\iint\limits_R f(x,y)\,dA = \int_a^b \int_c^d f(x,y)\,dy\,dx = \int_c^d \int_a^b f(x,y)\,dx\,dy$$

These integrals are called **iterated integrals**.

Note that there are in fact two ways of computing a double integral over a rectangle and also notice that the inner differential matches up with the limits on the inner integral and similarly for the outer differential and limits. In other words, if the inner differential is dy then the limits on the inner integral must be y limits of integration and if the outer differential is dy then the limits on the outer integral must be y limits of integration.

Now, on some level this is just notation and doesn't really tell us how to compute the double integral. Let's just take the first possibility above and change the notation a little.

$$\iint\limits_R f(x,y)\,dA = \int_a^b \left[\int_c^d f(x,y)\,dy\right] dx$$

We will compute the double integral by first computing

$$\int_c^d f(x,y)\,dy$$

and we compute this by holding x constant and integrating with respect to y as if this were a single integral. This will give a function involving only x's which we can in turn integrate.

We've done a similar process with partial derivatives. To take the derivative of a function with respect to y we treated the x's as constants and differentiated with respect to y as if it was a function of a single variable.

Double integrals work in the same manner. We think of all the x's as constants and integrate with respect to y or we think of all y's as constants and integrate with respect to x.

Let's take a look at some examples.

Example 1

Compute each of the following double integrals over the indicated rectangles.

(a) $\displaystyle\iint\limits_R 6xy^2\,dA,\ R = [2,4] \times [1,2]$

(b) $\displaystyle\iint\limits_R 2x - 4y^3\,dA,\ R = [-5,4] \times [0,3]$

(c) $\displaystyle\iint\limits_R x^2y^2 + \cos(\pi x) + \sin(\pi y)\,dA,\ R = [-2,-1] \times [0,1]$

(d) $\displaystyle\iint\limits_R \frac{1}{(2x+3y)^2}\,dA,\ R = [0,1] \times [1,2]$

(e) $\displaystyle\iint\limits_R x\mathbf{e}^{xy}\,dA,\ R = [-1,2] \times [0,1]$

Solution

(a) $\displaystyle\iint\limits_R 6xy^2\,dA,\ R = [2,4] \times [1,2]$

It doesn't matter which variable we integrate with respect to first, we will get the same answer regardless of the order of integration. To prove that let's work this one with each order to make sure that we do get the same answer.

Solution 1

In this case we will integrate with respect to y first. So, the iterated integral that we need to compute is,

$$\iint\limits_R 6xy^2\,dA = \int_2^4 \int_1^2 6xy^2\,dy\,dx$$

When setting these up make sure the limits match up to the differentials. Since the dy is the inner differential (*i.e.* we are integrating with respect to y first) the inner integral needs to have y limits.

To compute this we will do the inner integral first and we typically keep the outer integral around as follows,

$$\begin{aligned}\iint\limits_R 6xy^2\,dA &= \int_2^4 \left(2xy^3\right)\Big|_1^2\,dx\\ &= \int_2^4 16x - 2x\,dx\\ &= \int_2^4 14x\,dx\end{aligned}$$

Remember that we treat the x as a constant when doing the first integral and we don't do any integration with it yet. Now, we have a normal single integral so let's finish the integral by computing this.

$$\iint\limits_R 6xy^2\,dA = 7x^2\Big|_2^4 = 84$$

Solution 2

In this case we'll integrate with respect to x first and then y. Here is the work for this solution.

$$\begin{aligned}\iint\limits_R 6xy^2\,dA &= \int_1^2 \int_2^4 6xy^2\,dx\,dy\\ &= \int_1^2 \left(3x^2y^2\right)\Big|_2^4\,dy\\ &= \int_1^2 36y^2\,dy\\ &= 12y^3\Big|_1^2\\ &= 84\end{aligned}$$

Sure enough the same answer as the first solution.

So, remember that we can do the integration in any order.

(b) $\displaystyle\iint\limits_R 2x - 4y^3\,dA,\ R = [-5,4] \times [0,3]$

For this integral we'll integrate with respect to y first.

$$\begin{aligned}\iint\limits_R 2x - 4y^3\,dA &= \int_{-5}^{4}\int_{0}^{3} 2x - 4y^3\,dy\,dx \\ &= \int_{-5}^{4} \left(2xy - y^4\right)\Big|_0^3\,dx \\ &= \int_{-5}^{4} 6x - 81\,dx \\ &= \left(3x^2 - 81x\right)\Big|_{-5}^{4} \\ &= -756\end{aligned}$$

Remember that when integrating with respect to y all x's are treated as constants and so as far as the inner integral is concerned the 2 x is a constant and we know that when we integrate constants with respect to y we just tack on a y and so we get $2xy$ from the first term.

(c) $\displaystyle\iint\limits_R x^2y^2 + \cos(\pi x) + \sin(\pi y)\ dA,\ R = [-2,-1] \times [0,1]$

In this case we'll integrate with respect to x first.

$$\begin{aligned}\iint\limits_R x^2y^2 + \cos(\pi x) + \sin(\pi y)\ dA &= \int_{0}^{1}\int_{-2}^{-1} x^2y^2 + \cos(\pi x) + \sin(\pi y)\ dx\,dy \\ &= \int_{0}^{1} \left(\frac{1}{3}x^3y^2 + \frac{1}{\pi}\sin(\pi x) + x\sin(\pi y)\right)\Bigg|_{-2}^{-1} dy \\ &= \int_{0}^{1} \frac{7}{3}y^2 + \sin(\pi y)\ dy \\ &= \frac{7}{9}y^3 - \frac{1}{\pi}\cos(\pi y)\Big|_0^1 \\ &= \frac{7}{9} + \frac{2}{\pi}\end{aligned}$$

Don't forget your basic Calculus I substitutions!

(d) $\displaystyle\iint\limits_R \frac{1}{(2x+3y)^2}\,dA,\ R=[0,1]\times[1,2]$

In this case because the limits for x are kind of nice (*i.e.* they are zero and one which are often nice for evaluation) let's integrate with respect to x first. We'll also rewrite the integrand to help with the first integration.

$$\begin{aligned}\iint\limits_R (2x+3y)^{-2}\,dA &= \int_1^2\int_0^1 (2x+3y)^{-2}\,dx\,dy\\ &= \int_1^2 \left(-\frac{1}{2}(2x+3y)^{-1}\right)\Big|_0^1\,dy\\ &= -\frac{1}{2}\int_1^2 \frac{1}{2+3y}-\frac{1}{3y}\,dy\\ &= -\frac{1}{2}\left(\frac{1}{3}\ln|2+3y|-\frac{1}{3}\ln|y|\right)\Big|_1^2\\ &= -\frac{1}{6}\left(\ln(8)-\ln(2)-\ln(5)\right)\end{aligned}$$

(e) $\displaystyle\iint\limits_R x\mathbf{e}^{xy}\,dA,\ R=[-1,2]\times[0,1]$

Now, while we can technically integrate with respect to either variable first sometimes one way is significantly easier than the other way. In this case it will be significantly easier to integrate with respect to y first as we will see.

$$\iint\limits_R x\mathbf{e}^{xy}\,dA = \int_{-1}^2\int_0^1 x\mathbf{e}^{xy}\,dy\,dx$$

The y integration can be done with the quick substitution,

$$u = xy \qquad\qquad du = x\,dy$$

which gives

$$\begin{aligned}\iint\limits_R x\mathbf{e}^{xy}\,dA &= \int_{-1}^2 \mathbf{e}^{xy}\Big|_0^1\,dx\\ &= \int_{-1}^2 \mathbf{e}^x - 1\,dx\\ &= \left(\mathbf{e}^x - x\right)\Big|_{-1}^2\\ &= \mathbf{e}^2-2-\left(\mathbf{e}^{-1}+1\right)\\ &= \mathbf{e}^2-\mathbf{e}^{-1}-3\end{aligned}$$

So, not too bad of an integral there provided you get the substitution. Now let's see what would happen if we had integrated with respect to x first.

$$\iint\limits_R x\mathbf{e}^{xy}\,dA = \int_0^1 \int_{-1}^2 x\mathbf{e}^{xy}\,dx\,dy$$

In order to do this we would have to use integration by parts as follows,

$$\begin{aligned} u &= x & dv &= \mathbf{e}^{xy}\,dx \\ du &= dx & v &= \frac{1}{y}\mathbf{e}^{xy} \end{aligned}$$

The integral is then,

$$\begin{aligned} \iint\limits_R x\mathbf{e}^{xy}\,dA &= \int_0^1 \left(\frac{x}{y}\mathbf{e}^{xy} - \int \frac{1}{y}\mathbf{e}^{xy}\,dx\right)\Bigg|_{-1}^2\,dy \\ &= \int_0^1 \left(\frac{x}{y}\mathbf{e}^{xy} - \frac{1}{y^2}\mathbf{e}^{xy}\right)\Bigg|_{-1}^2\,dy \\ &= \int_0^1 \left(\frac{2}{y}\mathbf{e}^{2y} - \frac{1}{y^2}\mathbf{e}^{2y}\right) - \left(-\frac{1}{y}\mathbf{e}^{-y} - \frac{1}{y^2}\mathbf{e}^{-y}\right)\,dy \end{aligned}$$

We're not even going to continue here as these are very difficult, if not impossible, integrals to do.

As we saw in the previous set of examples we can do the integral in either direction. However, sometimes one direction of integration is significantly easier than the other so make sure that you think about which one you should do first before actually doing the integral.

The next topic of this section is a quick fact that can be used to make some iterated integrals somewhat easier to compute on occasion.

There is a nice special case of this kind of integral. First, let's assume that $f(x,y) = g(x)\,h(y)$ and let's also assume we are integrating over a rectangle given by $R = [a,b] \times [c,d]$. Then, the integral becomes,

$$\iint\limits_R f(x,y)\,dA = \iint\limits_R g(x)\,h(y)\,dA = \int_c^d \int_a^b g(x)\,h(y)\,dx\,dy$$

Note that it doesn't matter in this case which variable we integrate first as either order will arrive at the same result with the same work.

Next, notice that because the inner integral is with respect to x and $h(y)$ is a function only of y it can be considered a "constant" as far as the x integration is concerned (changing x will not affect the value of y!) and because it is also times $g(x)$ we can factor the $h(y)$ out of the inner integral.

Doing this gives,

$$\iint\limits_R f(x,y)\ dA = \iint\limits_R g(x)\,h(y)\ dA = \int_c^d h(y)\int_a^b g(x)\ dx\,dy$$

Now, $\int_a^b g(x)\ dx$ is a standard Calculus I definite integral and we know that its value is just a constant. Therefore, it can be factored out of the y integration to get,

$$\iint\limits_R f(x,y)\ dA = \iint\limits_R g(x)\,h(y)\ dA = \int_a^b g(x)\ dx\int_c^d h(y)\ dy$$

In other words, if we can break up the function into a function only of x times a function of only y then we can do the two integrals individually and multiply them together.

Here is a quick summary of this idea.

Fact

If $f(x,y) = g(x)\,h(y)$ and we are integrating over the rectangle $R = [a,b]\times[c,d]$ then,

$$\iint\limits_R f(x,y)\ dA = \iint\limits_R g(x)\,h(y)\ dA = \left(\int_a^b g(x)\ dx\right)\left(\int_c^d h(y)\ dy\right)$$

Let's do a quick example using this integral.

Example 2

Evaluate $\displaystyle\iint\limits_R x\cos^2(y)\ dA$, $R = [-2,3]\times\left[0,\frac{\pi}{2}\right]$.

Solution

Since the integrand is a function of x times a function of y we can use the fact.

$$\begin{aligned}\iint\limits_R x\cos^2(y)\ dA &= \left(\int_{-2}^{3} x\,dx\right)\left(\int_0^{\frac{\pi}{2}} \cos^2(y)\ dy\right)\\ &= \left(\frac{1}{2}x^2\right)\Bigg|_{-2}^{3}\left(\frac{1}{2}\int_0^{\frac{\pi}{2}} 1+\cos(2y)\ dy\right)\\ &= \left(\frac{5}{2}\right)\left(\frac{1}{2}\left(y+\frac{1}{2}\sin(2y)\right)\Bigg|_0^{\frac{\pi}{2}}\right)\\ &= \frac{5\pi}{8}\end{aligned}$$

We have one more topic to discuss in this section. This topic really doesn't have anything to do with iterated integrals, but this is as good a place as any to put it and there are liable to be some questions about it at this point as well so this is as good a place as any.

What we want to do is discuss single indefinite integrals of a function of two variables. In other words, we want to look at integrals like the following.

$$\int x \sec^2(2y) + 4xy\,dy$$
$$\int x^3 - \mathbf{e}^{-\frac{x}{y}}\,dx$$

From Calculus I we know that these integrals are asking what function that we differentiated to get the integrand. However, in this case we need to pay attention to the differential (dy or dx) in the integral, because that will change things a little.

In the case of the first integral we are asking what function we differentiated with respect to y to get the integrand while in the second integral we're asking what function differentiated with respect to x to get the integrand. For the most part answering these questions isn't that difficult. The important issue is how we deal with the constant of integration.

Here are the integrals.

$$\int x \sec^2(2y) + 4xy\,dy = \frac{x}{2}\tan(2y) + 2xy^2 + g(x)$$
$$\int x^3 - \mathbf{e}^{-\frac{x}{y}}\,dx = \frac{1}{4}x^4 + y\,\mathbf{e}^{-\frac{x}{y}} + h(y)$$

Notice that the "constants" of integration are now functions of the opposite variable. In the first integral we are differentiating with respect to y and we know that any function involving only x's will differentiate to zero and so when integrating with respect to y we need to acknowledge that there may have been a function of only x's in the function and so the "constant" of integration is a function of x.

Likewise, in the second integral, the "constant" of integration must be a function of y since we are integrating with respect to x. Again, remember if we differentiate the answer with respect to x then any function of only y's will differentiate to zero.

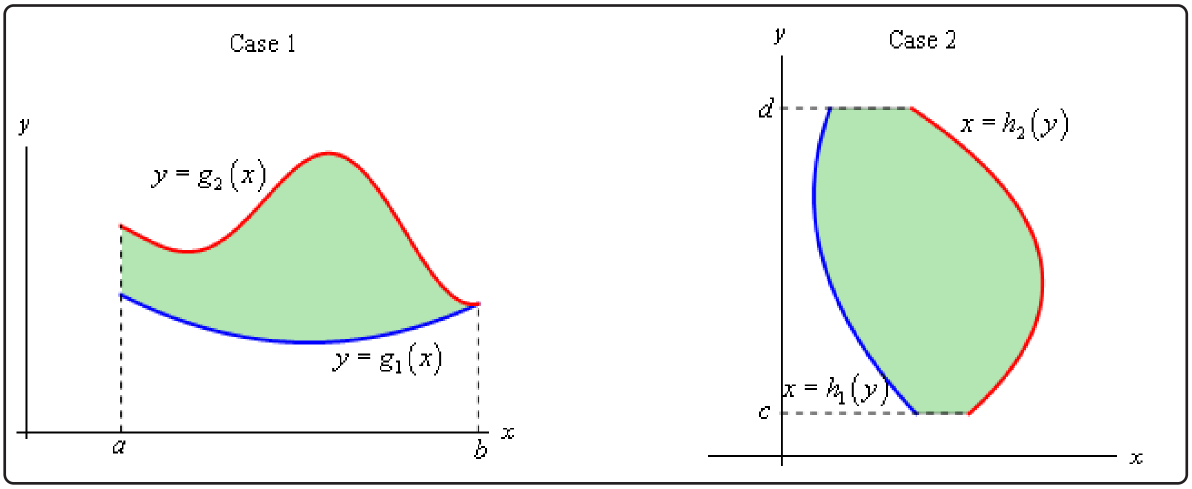
15.3 Double Integrals over General Regions

In the previous section we looked at double integrals over rectangular regions. The problem with this is that most of the regions are not rectangular so we need to now look at the following double integral,

$$\iint\limits_D f(x,y)\, dA$$

where D is any region.

There are two types of regions that we need to look at. Here is a sketch of both of them.



We will often use *set builder notation* to describe these regions. Here is the definition for the region in Case 1

$$D = \{(x,y)\,|a \le x \le b,\ g_1(x) \le y \le g_2(x)\}$$

and here is the definition for the region in Case 2.

$$D = \{(x,y)\,|h_1(y) \le x \le h_2(y),\ c \le y \le d\}$$

This notation is really just a fancy way of saying we are going to use all the points, (x, y), in which both of the coordinates satisfy the two given inequalities.

The double integral for both of these cases are defined in terms of iterated integrals as follows.

In Case 1 where $D = \{(x,y)\,|a \le x \le b,\ g_1(x) \le y \le g_2(x)\}$ the integral is defined to be,

$$\iint\limits_D f(x,y)\, dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\, dy\, dx$$

In Case 2 where $D = \{(x,y)\,|h_1(y) \le x \le h_2(y),\ c \le y \le d\}$ the integral is defined to be,

$$\iint\limits_D f(x,y)\, dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x,y)\, dx\, dy$$

Here are some properties of the double integral that we should go over before we actually do some examples. Note that all three of these properties are really just extensions of properties of single integrals that have been extended to double integrals.

Properties

1. $\displaystyle\iint\limits_D f(x,y) + g(x,y)\, dA = \iint\limits_D f(x,y)\, dA + \iint\limits_D g(x,y)\, dA$
2. $\displaystyle\iint\limits_D cf(x,y)\, dA = c\iint\limits_D f(x,y)\, dA$, where c is any constant.
3. If the region D can be split into two separate regions D_1 and D_2 then the integral can be written as
$$\iint\limits_D f(x,y)\, dA = \iint\limits_{D_1} f(x,y)\, dA + \iint\limits_{D_2} f(x,y)\, dA$$

Let's take a look at some examples of double integrals over general regions.

Example 1

Evaluate each of the following integrals over the given region D.

(a) $\displaystyle\iint\limits_D \mathbf{e}^{\frac{x}{y}}\, dA$, $D = \left\{(x,y)\,|1 \le y \le 2,\ y \le x \le y^3\right\}$

(b) $\displaystyle\iint\limits_D 4xy - y^3\, dA$, D is the region bounded by $y = \sqrt{x}$ and $y = x^3$.

(c) $\displaystyle\iint\limits_D 6x^2 - 40y\, dA$, D is the triangle with vertices $(0,3)$, $(1,1)$, and $(5,3)$.

Solution

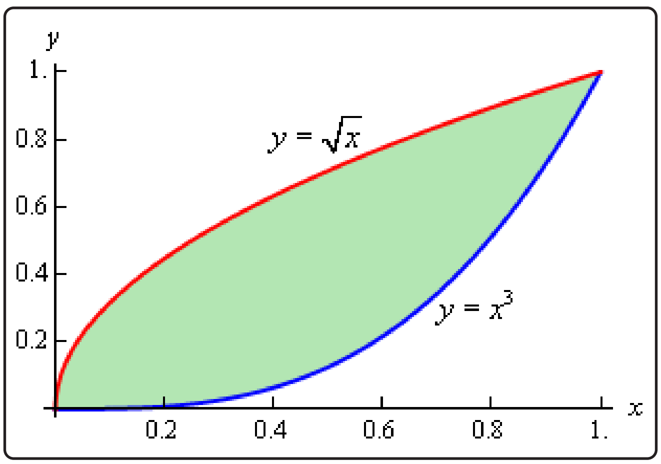
(a) $\displaystyle\iint\limits_D \mathbf{e}^{\frac{x}{y}}\, dA$, $D = \left\{(x,y)\,|1 \le y \le 2,\ y \le x \le y^3\right\}$

Okay, this first one is set up to just use the formula above so let's do that.

$$\begin{aligned}\iint\limits_D \mathbf{e}^{\frac{x}{y}}\,dA &= \int_1^2 \int_y^{y^3} \mathbf{e}^{\frac{x}{y}}\,dx\,dy = \int_1^2 y\,\mathbf{e}^{\frac{x}{y}}\Big|_y^{y^3}\,dy \\ &= \int_1^2 y\,\mathbf{e}^{y^2} - y\mathbf{e}^1\,dy \\ &= \left(\frac{1}{2}\mathbf{e}^{y^2} - \frac{1}{2}y^2\mathbf{e}^1\right)\Big|_1^2 = \frac{1}{2}\mathbf{e}^4 - 2\mathbf{e}^1\end{aligned}$$

(b) $\displaystyle\iint\limits_D 4xy - y^3\,dA$, D is the region bounded by $y = \sqrt{x}$ and $y = x^3$.

In this case we need to determine the two inequalities for x and y that we need to do the integral. The best way to do this is the graph the two curves. Here is a sketch.



So, from the sketch we can see that that two inequalities are,

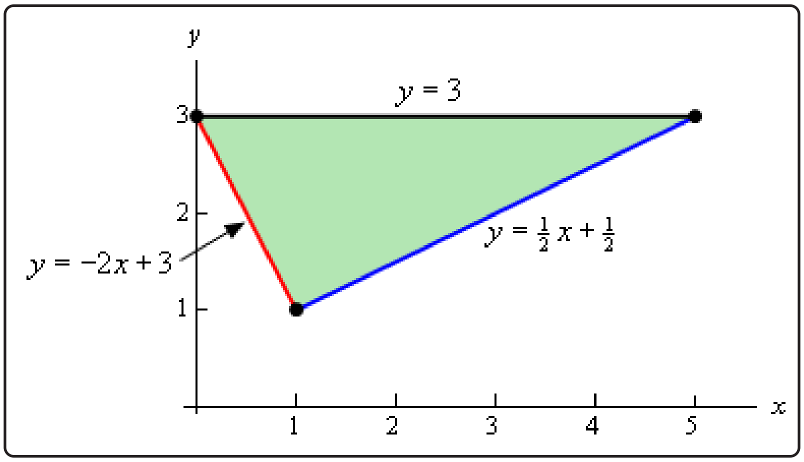
$$0 \le x \le 1 \qquad x^3 \le y \le \sqrt{x}$$

We can now do the integral,

$$\begin{aligned}\iint\limits_D 4xy - y^3\,dA &= \int_0^1 \int_{x^3}^{\sqrt{x}} 4xy - y^3\,dy\,dx \\ &= \int_0^1 \left(2xy^2 - \frac{1}{4}y^4\right)\Big|_{x^3}^{\sqrt{x}}\,dx \\ &= \int_0^1 \frac{7}{4}x^2 - 2x^7 + \frac{1}{4}x^{12}\,dx \\ &= \left(\frac{7}{12}x^3 - \frac{1}{4}x^8 + \frac{1}{52}x^{13}\right)\Big|_0^1 = \frac{55}{156}\end{aligned}$$

(c) $\displaystyle\iint\limits_D 6x^2 - 40y\,dA$, D is the triangle with vertices $(0,3)$, $(1,1)$, and $(5,3)$.

We got even less information about the region this time. Let's start this off by sketching the triangle.



Since we have two points on each edge it is easy to get the equations for each edge and so we'll leave it to you to verify the equations.

Now, there are two ways to describe this region. If we use functions of x, as shown in the image we will have to break the region up into two different pieces since the lower function is different depending upon the value of x. In this case the region would be given by $D = D_1 \cup D_2$ where,

$$D_1 = \{(x,y)\,|0 \le x \le 1,\ \ -2x+3 \le y \le 3\}$$
$$D_2 = \left\{(x,y)\,|1 \le x \le 5,\ \ \frac{1}{2}x + \frac{1}{2} \le y \le 3\right\}$$

Note the $\cup$ is the "union" symbol and just means that D is the region we get by combing the two regions. If we do this then we'll need to do two separate integrals, one for each of the regions.

To avoid this we could turn things around and solve the two equations for x to get,

$$y = -2x + 3 \qquad \Rightarrow \qquad x = -\frac{1}{2}y + \frac{3}{2}$$
$$y = \frac{1}{2}x + \frac{1}{2} \qquad \Rightarrow \qquad x = 2y - 1$$

If we do this we can notice that the same function is always on the right and the same function is always on the left and so the region is,

$$D = \left\{(x,y)\,|\ -\frac{1}{2}y + \frac{3}{2} \le x \le 2y - 1,\ \ 1 \le y \le 3\right\}$$

Writing the region in this form means doing a single integral instead of the two integrals we'd have to do otherwise.

Either way should give the same answer and so we can get an example in the notes of splitting a region up let's do both integrals.

Solution 1

$$\begin{aligned}
\iint\limits_D 6x^2 - 40y\,dA &= \iint\limits_{D_1} 6x^2 - 40y\,dA + \iint\limits_{D_2} 6x^2 - 40y\,dA \\
&= \int_0^1 \int_{-2x+3}^{3} 6x^2 - 40y\,dy\,dx + \int_1^5 \int_{\frac{1}{2}x+\frac{1}{2}}^{3} 6x^2 - 40y\,dy\,dx \\
&= \int_0^1 \left(6x^2y - 20y^2\right)\Big|_{-2x+3}^{3} dx + \int_1^5 \left(6x^2y - 20y^2\right)\Big|_{\frac{1}{2}x+\frac{1}{2}}^{3} dx \\
&= \int_0^1 12x^3 - 180 + 20(3-2x)^2\,dx + \\
&\qquad \int_1^5 -3x^3 + 15x^2 - 180 + 20\left(\frac{1}{2}x + \frac{1}{2}\right)^2 dx \\
&= \left(3x^4 - 180x - \frac{10}{3}(3-2x)^3\right)\Big|_0^1 + \\
&\qquad \left(-\frac{3}{4}x^4 + 5x^3 - 180x + \frac{40}{3}\left(\frac{1}{2}x + \frac{1}{2}\right)^3\right)\Big|_1^5 \\
&= -\frac{935}{3}
\end{aligned}$$

That was a lot of work. Notice however, that after we did the first substitution that we didn't multiply everything out. The two quadratic terms can be easily integrated with a basic Calc I substitution and so we didn't bother to multiply them out. We'll do that on occasion to make some of these integrals a little easier.

Solution 2

This solution will be a lot less work since we are only going to do a single integral.

$$\begin{aligned}
\iint\limits_D 6x^2 - 40y\,dA &= \int_1^3 \int_{-\frac{1}{2}y+\frac{3}{2}}^{2y-1} 6x^2 - 40y\,dx\,dy \\
&= \int_1^3 \left(2x^3 - 40xy\right)\Big|_{-\frac{1}{2}y+\frac{3}{2}}^{2y-1} dy \\
&= \int_1^3 100y - 100y^2 + 2(2y-1)^3 - 2\left(-\frac{1}{2}y + \frac{3}{2}\right)^3 dy \\
&= \left(50y^2 - \frac{100}{3}y^3 + \frac{1}{4}(2y-1)^4 + \left(-\frac{1}{2}y + \frac{3}{2}\right)^4\right)\Big|_1^3 \\
&= -\frac{935}{3}
\end{aligned}$$

So, the numbers were a little messier, but other than that there was much less work for the same result. Also notice that again we didn't cube out the two terms as they are easier to deal with using a Calc I substitution.

As the last part of the previous example has shown us we can integrate these integrals in either order (*i.e.* x followed by y or y followed by x), although often one order will be easier than the other. In fact, there will be times when it will not even be possible to do the integral in one order while it will be possible to do the integral in the other order.

Also, do not forget about Calculus I substitutions. Students often just get in a hurry and multiply everything out after doing the integral evaluation and end up missing a really simple Calculus I substitution that avoids the hassle of multiplying everything out. Calculus I substitutions don't always show up, but when they do they almost always simplify the work for the rest of the problem.

Let's see a couple of examples of these kinds of integrals.

Example 2

Evaluate the following integrals by first reversing the order of integration.

(a) $\displaystyle\int_0^3 \int_{x^2}^9 x^3 \mathbf{e}^{y^3}\, dy\, dx$

(b) $\displaystyle\int_0^8 \int_{\sqrt[3]{y}}^2 \sqrt{x^4+1}\, dx\, dy$

Solution

(a) $\displaystyle\int_0^3 \int_{x^2}^9 x^3 \mathbf{e}^{y^3}\, dy\, dx$

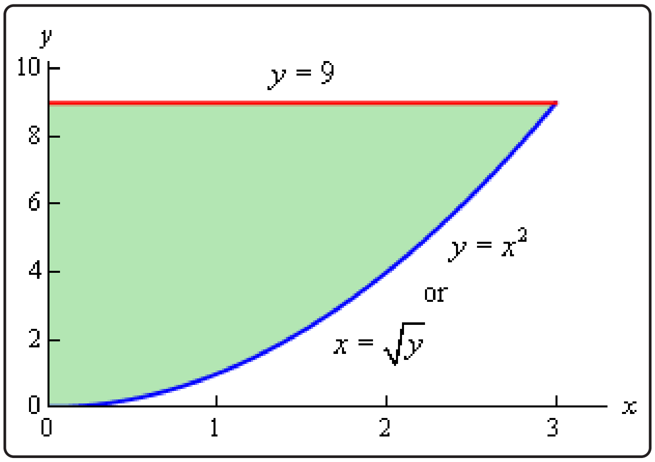
First, notice that if we try to integrate with respect to y we can't do the integral because we would need a y^2 in front of the exponential in order to do the y integration. We are going to hope that if we reverse the order of integration we will get an integral that we can do.

Now, when we say that we're going to reverse the order of integration this means that we want to integrate with respect to x first and then y. Note as well that we can't just interchange the integrals, keeping the original limits, and be done with it. This would not fix our original problem and in order to integrate with respect to x we can't have x's in the limits of the integrals. Even if we ignored that the answer would not be a constant as it should be.

So, let's see how we reverse the order of integration. The best way to reverse the order of integration is to first sketch the region given by the original limits of integration. From the integral we see that the inequalities that define this region are,

$$\begin{gathered} 0 \le x \le 3 \\ x^2 \le y \le 9 \end{gathered}$$

These inequalities tell us that we want the region with $y = x^2$ on the lower boundary and $y = 9$ on the upper boundary that lies between $x = 0$ and $x = 3$. Here is a sketch of that region.



Since we want to integrate with respect to x first we will need to determine limits of x (probably in terms of y) and then get the limits on the y's. Here they are for this region.

$$\begin{gathered} 0 \le x \le \sqrt{y} \\ 0 \le y \le 9 \end{gathered}$$

Any horizontal line drawn in this region will start at $x = 0$ and end at $x = \sqrt{y}$ and so these are the limits on the x's and the range of y's for the regions is 0 to 9.

The integral, with the order reversed, is now,

$$\int_0^3 \int_{x^2}^9 x^3 \mathbf{e}^{y^3}\, dy\, dx = \int_0^9 \int_0^{\sqrt{y}} x^3 \mathbf{e}^{y^3}\, dx\, dy$$

and notice that we can do the first integration with this order. We'll also hope that this will give us a second integral that we can do. Here is the work for this integral.

$$\begin{aligned}\int_0^3 \int_{x^2}^9 x^3 \mathbf{e}^{y^3}\, dy\, dx &= \int_0^9 \int_0^{\sqrt{y}} x^3 \mathbf{e}^{y^3}\, dx\, dy \\ &= \int_0^9 \left. \frac{1}{4} x^4 \mathbf{e}^{y^3} \right|_0^{\sqrt{y}} dy \\ &= \int_0^9 \frac{1}{4} y^2 \mathbf{e}^{y^3}\, dy \\ &= \left. \frac{1}{12} \mathbf{e}^{y^3} \right|_0^9 \\ &= \frac{1}{12}\left(\mathbf{e}^{729} - 1\right)\end{aligned}$$

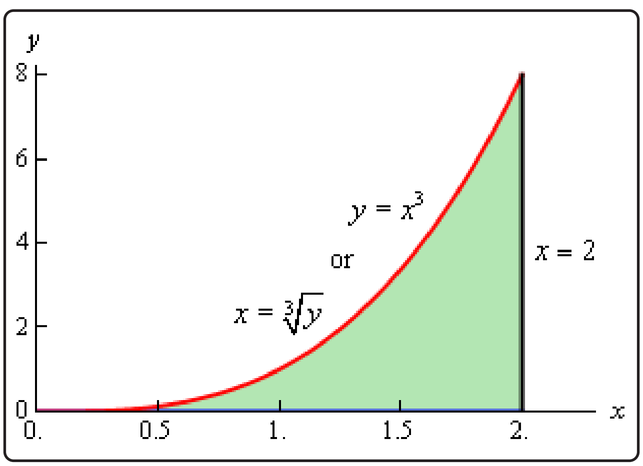
So, as we hoped, we were able to do the integral once we interchanged the order of integration.

(b) $\displaystyle\int_0^8 \int_{\sqrt[3]{y}}^2 \sqrt{x^4+1}\, dx\, dy$

As with the first integral we cannot do this integral by integrating with respect to x first so we'll hope that by reversing the order of integration we will get something that we can integrate. Here are the limits for the variables that we get from this integral.

$$\begin{aligned}\sqrt[3]{y} \le x \le 2 \\ 0 \le y \le 8\end{aligned}$$

and here is a sketch of this region.



So, if we reverse the order of integration we get the following limits.

$$\begin{aligned}0 \le x \le 2 \\ 0 \le y \le x^3\end{aligned}$$

The integral is then,

$$\int_0^8 \int_{\sqrt[3]{y}}^2 \sqrt{x^4+1}\,dx\,dy = \int_0^2 \int_0^{x^3} \sqrt{x^4+1}\,dy\,dx$$
$$= \int_0^2 y\sqrt{x^4+1}\Big|_0^{x^3}\,dx$$
$$= \int_0^2 x^3\sqrt{x^4+1}\,dx = \frac{1}{6}\left(17^{\frac{3}{2}}-1\right)$$

The final topic of this section is two geometric interpretations of a double integral. The first interpretation is an extension of the idea that we used to develop the idea of a double integral in the first section of this chapter. We did this by looking at the volume of the solid that was below the surface of the function $z = f(x,y)$ and over the rectangle R in the xy-plane. This idea can be extended to more general regions.

The volume of the solid that lies below the surface given by $z = f(x,y)$ and above the region D in the xy-plane is given by,

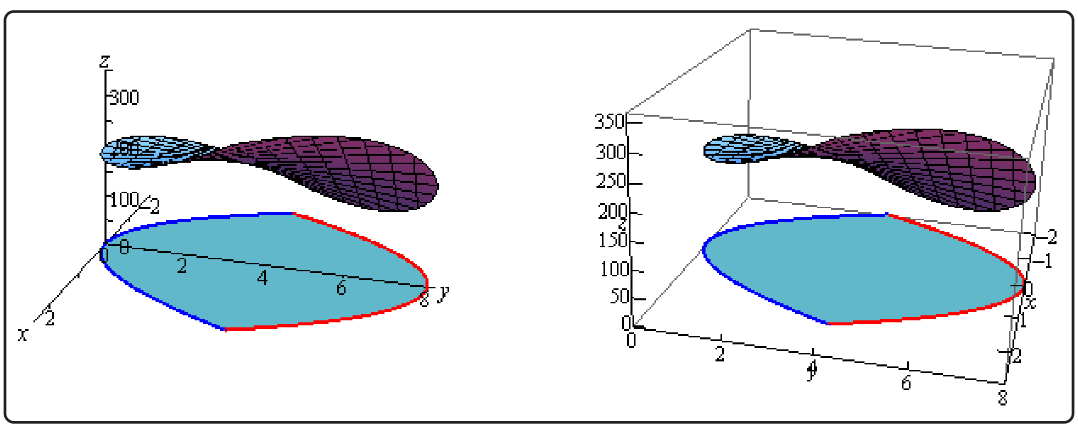
$$V = \iint\limits_D f(x,y)\,dA$$

Example 3

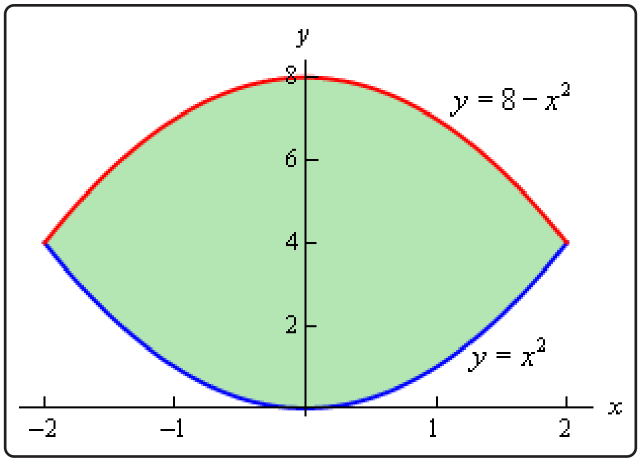
Find the volume of the solid that lies below the surface given by $z = 16xy + 200$ and lies above the region in the xy-plane bounded by $y = x^2$ and $y = 8 - x^2$.

Solution

Here is the graph of the surface and we've tried to show the region in the xy-plane below the surface.



Here is a sketch of the region in the xy-plane by itself.



By setting the two bounding equations equal we can see that they will intersect at $x = 2$ and $x = -2$. So, the inequalities that will define the region D in the xy-plane are,

$$\begin{gathered} -2 \le x \le 2 \\ x^2 \le y \le 8 - x^2 \end{gathered}$$

The volume is then given by,

$$\begin{aligned} V &= \iint\limits_D 16xy + 200\, dA \\ &= \int_{-2}^{2} \int_{x^2}^{8-x^2} 16xy + 200\, dy\, dx \\ &= \int_{-2}^{2} \left(8xy^2 + 200y\right)\Big|_{x^2}^{8-x^2} dx \\ &= \int_{-2}^{2} -128x^3 - 400x^2 + 512x + 1600\, dx \\ &= \left(-32x^4 - \frac{400}{3}x^3 + 256x^2 + 1600x\right)\Big|_{-2}^{2} = \frac{12800}{3} \end{aligned}$$

Example 4

Find the volume of the solid enclosed by the planes $4x + 2y + z = 10$, $y = 3x$, $z = 0$, $x = 0$.

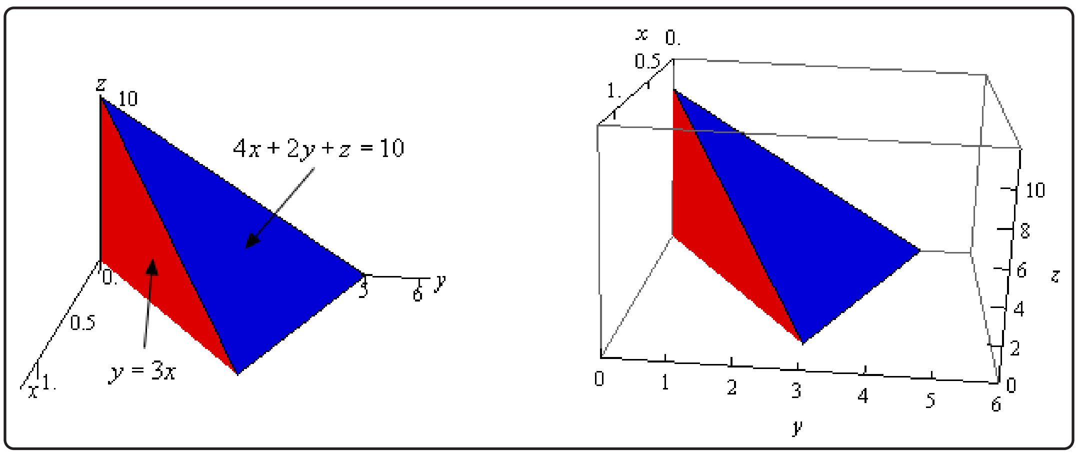
Solution

This example is a little different from the previous one. Here the region D is not explicitly given so we're going to have to find it. First, notice that the last two planes are really telling us that we won't go past the xy-plane and the yz-plane when we reach them.

The first plane, $4x + 2y + z = 10$, is the top of the volume and so we are really looking for the volume under,

$$z = 10 - 4x - 2y$$

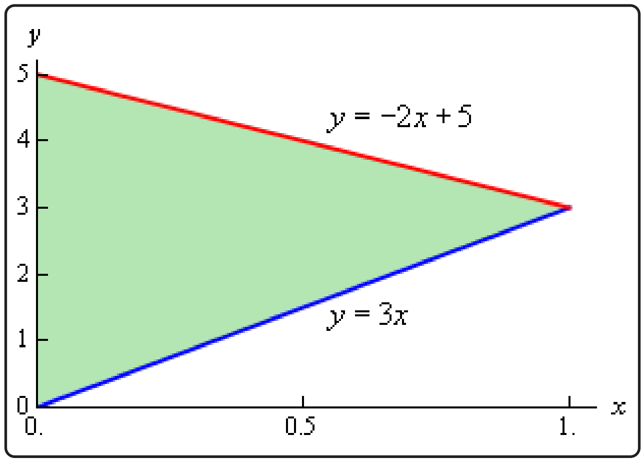
and above the region D in the xy-plane. The second plane, $y = 3x$ (yes that is a plane), gives one of the sides of the volume as shown below.



The region D will be the region in the xy-plane (*i.e.* $z = 0$) that is bounded by $y = 3x$, $x = 0$, and the line where $z + 4x + 2y = 10$ intersects the xy-plane. We can determine where $z + 4x + 2y = 10$ intersects the xy-plane by plugging $z = 0$ into it.

$$0 + 4x + 2y = 10 \quad \Rightarrow \quad 2x + y = 5 \quad \Rightarrow \quad y = -2x + 5$$

So, here is a sketch the region D.



The region D is really where this solid will sit on the xy-plane and here are the inequalities that define the region.

$$\begin{gathered} 0 \leq x \leq 1 \\ 3x \leq y \leq -2x+5 \end{gathered}$$

Here is the volume of this solid.

$$\begin{aligned} V &= \iint\limits_D 10 - 4x - 2y\,dA \\ &= \int_0^1 \int_{3x}^{-2x+5} 10 - 4x - 2y\,dy\,dx \\ &= \int_0^1 \left(10y - 4xy - y^2\right)\Big|_{3x}^{-2x+5} dx \\ &= \int_0^1 25x^2 - 50x + 25\,dx \\ &= \left(\frac{25}{3}x^3 - 25x^2 + 25x\right)\Big|_0^1 = \frac{25}{3} \end{aligned}$$

Note that more generally,

$$V = \iint\limits_D f(x,y)\,dA$$

gives the net volume between the graph of $z = f(x,y)$ and the region D in the xy-plane. Regions that are below the xy-plane have a negative volume and regions that are above the xy-plane have a positive volume.

We saw a similar idea in Calculus I where,

$$A = \int_a^b f(x)\,dx$$

gives the net area between the curve given by $y = f(x)$ and the x-axis on the interval $[a, b]$.

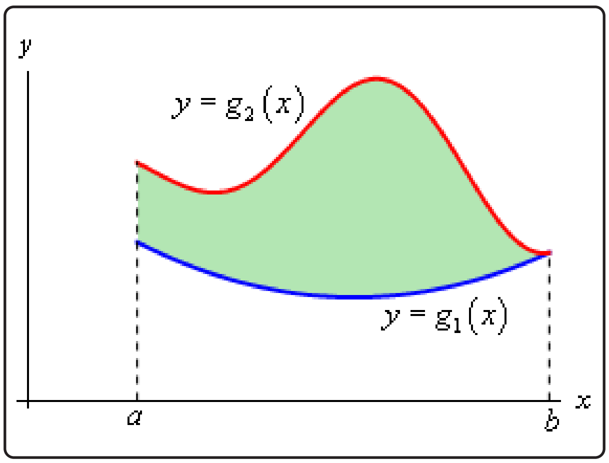
The second geometric interpretation of a double integral is the following.

Fact

$$\text{Area of } D = \iint\limits_D dA$$

This is easy to see why this is true in general. Let's suppose that we want to find the area of the region shown below.

The area between the two functions has been shaded in.



From Calculus I we know that this area can be found by the integral,

$$A = \int_a^b g_2(x) - g_1(x)\,dx$$

Or in terms of a double integral we have,

$$\begin{aligned} \text{Area of } D &= \iint\limits_D dA \\ &= \int_a^b \int_{g_1(x)}^{g_2(x)} dy\,dx \\ &= \int_a^b y\Big|_{g_1(x)}^{g_2(x)}\,dx = \int_a^b g_2(x) - g_1(x)\,dx \end{aligned}$$

This is exactly the same formula we had in Calculus I.

15.4 Double Integrals in Polar Coordinates

To this point we've seen quite a few double integrals. However, in every case we've seen to this point the region D could be easily described in terms of simple functions in Cartesian coordinates. In this section we want to look at some regions that are much easier to describe in terms of polar coordinates. For instance, we might have a region that is a disk, ring, or a portion of a disk or ring. In these cases, using Cartesian coordinates could be somewhat cumbersome. For instance, let's suppose we wanted to do the following integral,

$$\iint\limits_D f(x,y)\,dA, \quad D \text{ is the disk of radius 2}$$

To this we would have to determine a set of inequalities for x and y that describe this region. These would be,

$$\begin{gathered} -2 \le x \le 2 \\ -\sqrt{4-x^2} \le y \le \sqrt{4-x^2} \end{gathered}$$

With these limits the integral would become,

$$\iint\limits_D f(x,y)\,dA = \int_{-2}^{2}\int_{-\sqrt{4-x^2}}^{\sqrt{4-x^2}} f(x,y)\,dy\,dx$$

Due to the limits on the inner integral this is liable to be an unpleasant integral to compute.

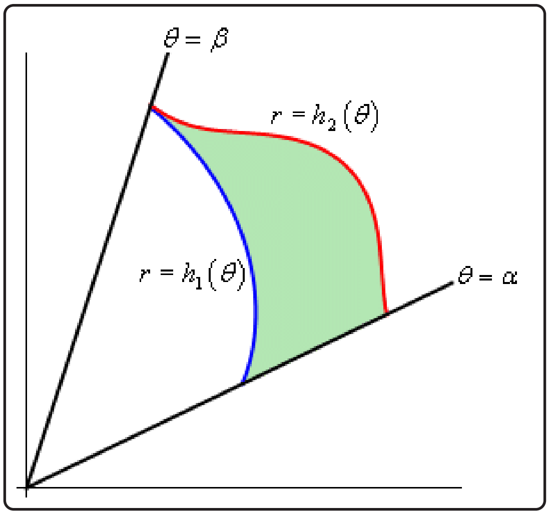
However, a disk of radius 2 can be defined in polar coordinates by the following inequalities,

$$\begin{gathered} 0 \le \theta \le 2\pi \\ 0 \le r \le 2 \end{gathered}$$

These are very simple limits and, in fact, are constant limits of integration which almost always makes integrals somewhat easier.

So, if we could convert our double integral formula into one involving polar coordinates we would be in pretty good shape. The problem is that we can't just convert the dx and the dy into a dr and a $d\theta$. In computing double integrals to this point we have been using the fact that $dA = dx\,dy$ and this really does require Cartesian coordinates to use. Once we've moved into polar coordinates $dA \ne dr\,d\theta$ and so we're going to need to determine just what dA is under polar coordinates.

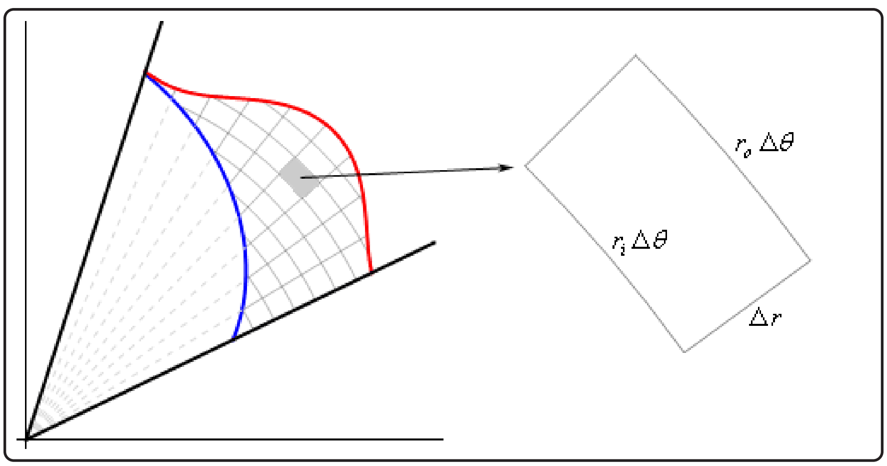
So, let's step back a little bit and start off with a general region in terms of polar coordinates and see what we can do with that. Here is a sketch of some region using polar coordinates.



So, our general region will be defined by inequalities,

$$\alpha \le \theta \le \beta$$
$$h_1(\theta) \le r \le h_2(\theta)$$

Now, to find dA let's redo the figure above as follows,



As shown, we'll break up the region into a mesh of radial lines and arcs. Now, if we pull one of the pieces of the mesh out as shown we have something that is almost, but not quite a rectangle. The area of this piece is ΔA. The two sides of this piece both have length $\Delta r = r_o - r_i$ where r_o is the radius of the outer arc and r_i is the radius of the inner arc. Basic geometry then tells us that the length of the inner edge is $r_i\,\Delta\theta$ while the length of the out edge is $r_o\,\Delta\theta$ where $\Delta\theta$ is the angle between the two radial lines that form the sides of this piece.

Now, let's assume that we've taken the mesh so small that we can assume that $r_i \approx r_o = r$ and with this assumption we can also assume that our piece is close enough to a rectangle that we can also then assume that,

$$\Delta A \approx r\,\Delta\theta\,\Delta r$$

Also, if we assume that the mesh is small enough then we can also assume that,

$$dA \approx \Delta A \qquad d\theta \approx \Delta\theta \qquad dr \approx \Delta r$$

With these assumptions we then get $dA \approx r\,dr\,d\theta$.

In order to arrive at this we had to make the assumption that the mesh was very small. This is not an unreasonable assumption. Recall that the definition of a double integral is in terms of two limits and as limits go to infinity the mesh size of the region will get smaller and smaller. In fact, as the mesh size gets smaller and smaller the formula above becomes more and more accurate and so we can say that,

Fact

$$dA = r\,dr\,d\theta$$

We'll see another way of deriving this once we reach the Change of Variables section later in this chapter. This second way will not involve any assumptions either and so it maybe a little better way of deriving this.

Before moving on it is again important to note that $dA \neq dr\,d\theta$. The actual formula for dA has an r in it. It will be easy to forget this r on occasion, but as you'll see without it some integrals will not be possible to do.

Now, if we're going to be converting an integral in Cartesian coordinates into an integral in polar coordinates we are going to have to make sure that we've also converted all the x's and y's into polar coordinates as well. To do this we'll need to remember the following conversion formulas,

$$x = r\cos(\theta) \qquad y = r\sin(\theta) \qquad r^2 = x^2 + y^2$$

We are now ready to write down a formula for the double integral in terms of polar coordinates.

Fact

$$\iint\limits_D f(x, y)\, dA = \int_\alpha^\beta \int_{h_1(\theta)}^{h_2(\theta)} f(r\cos(\theta), r\sin(\theta))\, r\,dr\,d\theta$$

It is important to not forget the added r and don't forget to convert the Cartesian coordinates in the function over to polar coordinates.

Let's look at a couple of examples of these kinds of integrals.

Example 1

Evaluate the following integrals by converting them into polar coordinates.

(a) $\displaystyle\iint\limits_D 2x\,y\,dA$, D is the portion of the region between the circles of radius 2 and radius 5 centered at the origin that lies in the first quadrant.

(b) $\displaystyle\iint\limits_D \mathbf{e}^{x^2+y^2}\,dA$, D is the unit disk centered at the origin.

Solution

(a) $\displaystyle\iint\limits_D 2x\,y\,dA$, D is the portion of the region between the circles of radius 2 and radius 5 centered at the origin that lies in the first quadrant.

First let's get D in terms of polar coordinates. The circle of radius 2 is given by $r = 2$ and the circle of radius 5 is given by $r = 5$. We want the region between the two circles, so we will have the following inequality for r.

$$2 \le r \le 5$$

Also, since we only want the portion that is in the first quadrant we get the following range of θ's.

$$0 \le \theta \le \frac{\pi}{2}$$

Now that we've got these we can do the integral.

$$\iint\limits_D 2x\,y\,dA = \int_0^{\frac{\pi}{2}}\int_2^5 2\big(r\cos(\theta)\big)\big(r\sin(\theta)\big)r\,dr\,d\theta$$

Don't forget to do the conversions and to add in the extra r. Now, let's simplify and make use of the double angle formula for sine to make the integral a little easier.

$$\begin{aligned}\iint\limits_D 2x\,y\,dA &= \int_0^{\frac{\pi}{2}}\int_2^5 r^3\sin(2\theta)\,dr\,d\theta \\ &= \int_0^{\frac{\pi}{2}} \left.\frac{1}{4}r^4\sin(2\theta)\right|_2^5 d\theta \\ &= \int_0^{\frac{\pi}{2}} \frac{609}{4}\sin(2\theta)\,d\theta \\ &= \left.-\frac{609}{8}\cos(2\theta)\right|_0^{\frac{\pi}{2}} \\ &= \frac{609}{4}\end{aligned}$$

(b) $\displaystyle\iint\limits_D \mathbf{e}^{x^2+y^2}\,dA$, D is the unit disk centered at the origin.

In this case we can't do this integral in terms of Cartesian coordinates. We will however be able to do it in polar coordinates. First, the region D is defined by,

$$\begin{gathered} 0 \le \theta \le 2\pi \\ 0 \le r \le 1 \end{gathered}$$

In terms of polar coordinates the integral is then,

$$\iint\limits_D \mathbf{e}^{x^2+y^2}\,dA = \int_0^{2\pi}\int_0^1 r\,\mathbf{e}^{r^2}\,dr\,d\theta$$

Notice that the addition of the r gives us an integral that we can now do. Here is the work for this integral.

$$\begin{aligned} \iint\limits_D \mathbf{e}^{x^2+y^2}\,dA &= \int_0^{2\pi}\int_0^1 r\,\mathbf{e}^{r^2}\,dr\,d\theta \\ &= \int_0^{2\pi} \left.\frac{1}{2}\mathbf{e}^{r^2}\right|_0^1 d\theta \\ &= \int_0^{2\pi} \frac{1}{2}\left(\mathbf{e}-1\right)\,d\theta \\ &= \pi\left(\mathbf{e}-1\right) \end{aligned}$$

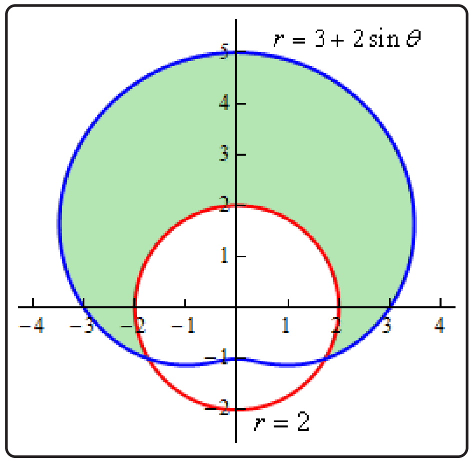
Let's not forget that we still have the two geometric interpretations for these integrals as well.

Example 2

Determine the area of the region that lies inside $r = 3 + 2\sin(\theta)$ and outside $r = 2$.

Solution

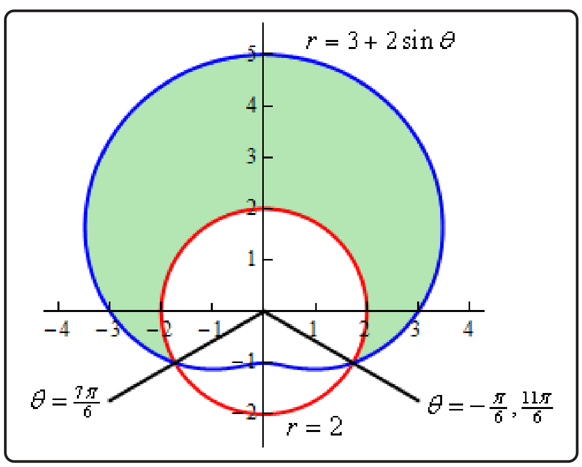
Here is a sketch of the region, D, that we want to determine the area of.



To determine this area we'll need to know that value of θ for which the two curves intersect. We can determine these points by setting the two equations equal and solving.

$$\begin{aligned} 3 + 2\sin(\theta) &= 2 \\ \sin(\theta) &= -\frac{1}{2} \qquad \Rightarrow \qquad \theta = \frac{7\pi}{6}, \frac{11\pi}{6} \end{aligned}$$

Here is a sketch of the figure with these angles added.



Note as well that we've acknowledged that $-\frac{\pi}{6}$ is another representation for the angle $\frac{11\pi}{6}$. This is important since we need the range of θ to actually enclose the regions as we increase from the lower limit to the upper limit. If we'd chosen to use $\frac{11\pi}{6}$ then as we increase from $\frac{7\pi}{6}$ to $\frac{11\pi}{6}$ we would be tracing out the lower portion of the circle and that is not the region that we are after.

So, here are the ranges that will define the region.

$$-\frac{\pi}{6} \le \theta \le \frac{7\pi}{6}$$
$$2 \le r \le 3 + 2\sin(\theta)$$

To get the ranges for r the function that is closest to the origin is the lower bound and the function that is farthest from the origin is the upper bound.

The area of the region D is then,

$$\begin{aligned}
A &= \iint\limits_D dA \\
&= \int_{-\pi/6}^{7\pi/6} \int_2^{3+2\sin(\theta)} r\,dr\,d\theta \\
&= \int_{-\pi/6}^{7\pi/6} \left.\frac{1}{2}r^2\right|_2^{3+2\sin(\theta)} d\theta \\
&= \int_{-\pi/6}^{7\pi/6} \frac{5}{2} + 6\sin(\theta) + 2\sin^2(\theta)\,d\theta \\
&= \int_{-\pi/6}^{7\pi/6} \frac{7}{2} + 6\sin(\theta) - \cos(2\theta)\,d\theta \\
&= \left.\left(\frac{7}{2}\theta - 6\cos(\theta) - \frac{1}{2}\sin(2\theta)\right)\right|_{-\frac{\pi}{6}}^{\frac{7\pi}{6}} \\
&= \frac{11\sqrt{3}}{2} + \frac{14\pi}{3} = 24.187
\end{aligned}$$

Example 3

Determine the volume of the region that lies under the sphere $x^2 + y^2 + z^2 = 9$, above the plane $z = 0$ and inside the cylinder $x^2 + y^2 = 5$.

Solution

We know that the formula for finding the volume of a region is,

$$V = \iint\limits_D f(x,y)\,dA$$

In order to make use of this formula we're going to need to determine the function that we should be integrating and the region D that we're going to be integrating over.

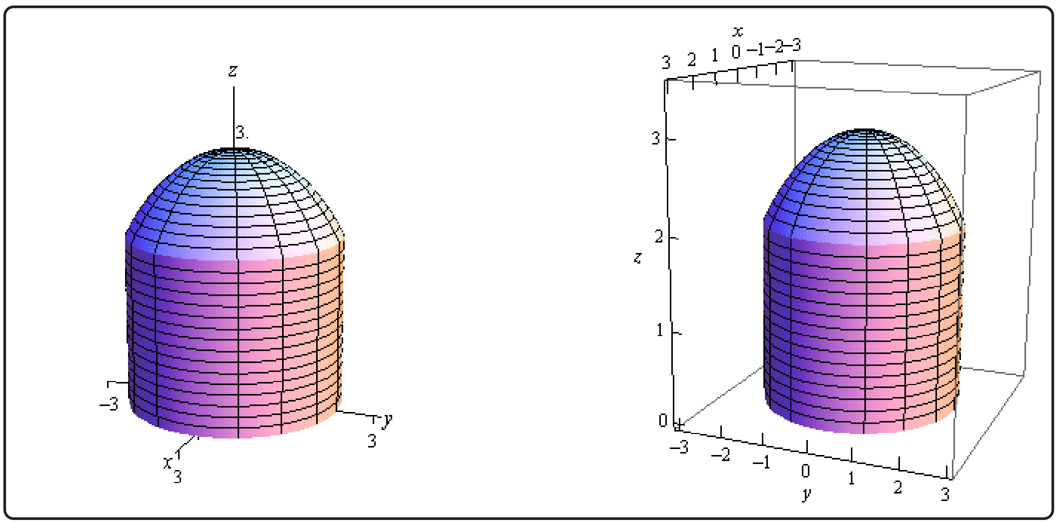
The function isn't too bad. It's just the sphere, however, we do need it to be in the form

$z = f(x, y)$. We are looking at the region that lies under the sphere and above the plane $z = 0$ (just the xy-plane right?) and so all we need to do is solve the equation for z and when taking the square root we'll take the positive one since we are wanting the region above the xy-plane. Here is the function.

$$z = \sqrt{9 - x^2 - y^2}$$

The region D isn't too bad in this case either. As we take points, (x, y), from the region we need to completely graph the portion of the sphere that we are working with. Since we only want the portion of the sphere that actually lies inside the cylinder given by $x^2 + y^2 = 5$ this is also the region D. The region D is the disk $x^2 + y^2 \le 5$ in the xy-plane.

For reference purposes here is a sketch of the region that we are trying to find the volume of.



So, the region that we want the volume for is really a cylinder with a cap that comes from the sphere.

We are definitely going to want to do this integral in terms of polar coordinates so here are the limits (in polar coordinates) for the region,

$$0 \le \theta \le 2\pi$$
$$0 \le r \le \sqrt{5}$$

and we'll need to convert the function to polar coordinates as well.

$$z = \sqrt{9 - (x^2 + y^2)} = \sqrt{9 - r^2}$$

The volume is then,

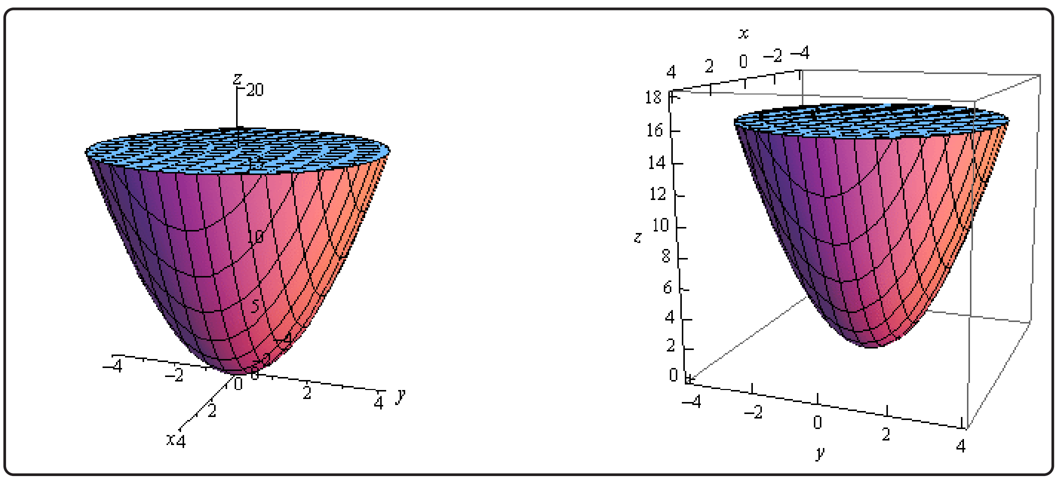
$$\begin{aligned} V &= \iint\limits_{D} \sqrt{9 - x^2 - y^2}\, dA \\ &= \int_0^{2\pi} \int_0^{\sqrt{5}} r\sqrt{9 - r^2}\, dr\, d\theta \\ &= \int_0^{2\pi} -\frac{1}{3}\left(9 - r^2\right)^{\frac{3}{2}}\Big|_0^{\sqrt{5}}\, d\theta \\ &= \int_0^{2\pi} \frac{19}{3}\, d\theta \\ &= \frac{38\pi}{3} \end{aligned}$$

Example 4

Find the volume of the region that lies inside $z = x^2 + y^2$ and below the plane $z = 16$.

Solution

Let's start this example off with a quick sketch of the region.



Now, in this case the standard formula is not going to work. The formula

$$V = \iint\limits_{D} f(x, y)\, dA$$

finds the volume under the function $f(x, y)$ and we're actually after the volume that is above a function. This isn't the problem that it might appear to be however. First, notice that

$$V = \iint\limits_D 16\, dA$$

will be the volume under $z = 16$ (of course we'll need to determine D eventually) while

$$V = \iint\limits_D x^2 + y^2\, dA$$

is the volume under $z = x^2 + y^2$, using the same D.

The volume that we're after is really the difference between these two or,

$$V = \iint\limits_D 16\, dA - \iint\limits_D x^2 + y^2\, dA = \iint\limits_D 16 - \left(x^2 + y^2\right)\, dA$$

Now all that we need to do is to determine the region D and then convert everything over to polar coordinates.

Determining the region D in this case is not too bad. If we were to look straight down the z-axis onto the region we would see a circle of radius 4 centered at the origin. This is because the top of the region, where the elliptic paraboloid intersects the plane, is the widest part of the region. We know the z coordinate at the intersection so, setting $z = 16$ in the equation of the paraboloid gives,

$$16 = x^2 + y^2$$

which is the equation of a circle of radius 4 centered at the origin.

Here are the inequalities for the region and the function we'll be integrating in terms of polar coordinates.

$$0 \le \theta \le 2\pi \qquad 0 \le r \le 4 \qquad z = 16 - r^2$$

The volume is then,

$$\begin{aligned} V &= \iint\limits_D 16 - \left(x^2 + y^2\right)\, dA \\ &= \int_0^{2\pi} \int_0^4 r\left(16 - r^2\right)\, dr\, d\theta \\ &= \int_0^{2\pi} \left(8r^2 - \frac{1}{4}r^4\right)\Big|_0^4\, d\theta \\ &= \int_0^{2\pi} 64\, d\theta \\ &= 128\pi \end{aligned}$$

In both of the previous volume problems we would have not been able to easily compute the volume without first converting to polar coordinates so, as these examples show, it is a good idea to always remember polar coordinates.

There is one more type of example that we need to look at before moving on to the next section. Sometimes we are given an iterated integral that is already in terms of x and y and we need to convert this over to polar so that we can actually do the integral. We need to see an example of how to do this kind of conversion.

Example 5

Evaluate the following integral by first converting to polar coordinates.

$$\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{0} \cos\left(x^2+y^2\right)\,dy\,dx$$

Solution

First, notice that we cannot do this integral in Cartesian coordinates and so converting to polar coordinates may be the only option we have for actually doing the integral. Notice that the function will convert to polar coordinates nicely and so shouldn't be a problem.

Let's first determine the region that we're integrating over and see if it's a region that can be easily converted into polar coordinates. Here are the inequalities that define the region in terms of Cartesian coordinates.

$$\begin{gathered} -1 \le x \le 1 \\ -\sqrt{1-x^2} \le y \le 0 \end{gathered}$$

Now, the lower limit for the y's is,

$$y = -\sqrt{1-x^2}$$

and this looks like the bottom of the circle of radius 1 centered at the origin. Since the upper limit for the y's is $y = 0$ we won't have any portion of the top half of the disk and so it looks like we are going to have a portion (or all) of the bottom of the disk of radius 1 centered at the origin.

The range for the x's in turn, tells us that we are will in fact have the complete bottom part of the disk.

So, we know that the inequalities that will define this region in terms of polar coordinates are then,

$$\begin{gathered} \pi \le \theta \le 2\pi \\ 0 \le r \le 1 \end{gathered}$$

Finally, we just need to remember that,

$$dx\,dy = dA = r\,dr\,d\theta$$

and so the integral becomes,

$$\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{0} \cos\left(x^2+y^2\right)\,dy\,dx = \int_{\pi}^{2\pi}\int_{0}^{1} r\cos\left(r^2\right)\,dr\,d\theta$$

Note that this is an integral that we can do. So, here is the rest of the work for this integral.

$$\begin{aligned}\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{0} \cos\left(x^2+y^2\right)\,dy\,dx &= \int_{\pi}^{2\pi} \frac{1}{2}\sin\left(r^2\right)\Big|_0^1\,d\theta \\ &= \int_{\pi}^{2\pi} \frac{1}{2}\sin\left(1\right)\,d\theta \\ &= \frac{\pi}{2}\sin\left(1\right)\end{aligned}$$

15.5 Triple Integrals

Now that we know how to integrate over a two-dimensional region we need to move on to integrating over a three-dimensional region. We used a double integral to integrate over a two-dimensional region and so it shouldn't be too surprising that we'll use a **triple integral** to integrate over a three dimensional region. The notation for the general triple integrals is,

$$\iiint\limits_E f(x,y,z)\,dV$$

Let's start simple by integrating over the box,

$$B = [a,b] \times [c,d] \times [r,s]$$

Note that when using this notation we list the x's first, the y's second and the z's third.

The triple integral in this case is,

$$\iiint\limits_B f(x,y,z)\,dV = \int_r^s \int_c^d \int_a^b f(x,y,z)\,dx\,dy\,dz$$

Note that we integrated with respect to x first, then y, and finally z here, but in fact there is no reason to the integrals in this order. There are 6 different possible orders to do the integral in and which order you do the integral in will depend upon the function and the order that you feel will be the easiest. We will get the same answer regardless of the order however.

Let's do a quick example of this type of triple integral.

Example 1

Evaluate the following integral.

$$\iiint\limits_B 8xyz\,dV \qquad B = [2,3] \times [1,2] \times [0,1]$$

Solution

Just to make the point that order doesn't matter let's use a different order from that listed above. We'll do the integral in the following order.

$$\iiint\limits_B 8xyz\,dV = \int_1^2 \int_2^3 \int_0^1 8xyz\,dz\,dx\,dy$$
$$= \int_1^2 \int_2^3 4xyz^2\Big|_0^1\,dx\,dy$$
$$= \int_1^2 \int_2^3 4xy\,dx\,dy$$
$$= \int_1^2 2x^2y\Big|_2^3\,dy$$
$$= \int_1^2 10y\,dy = 15$$

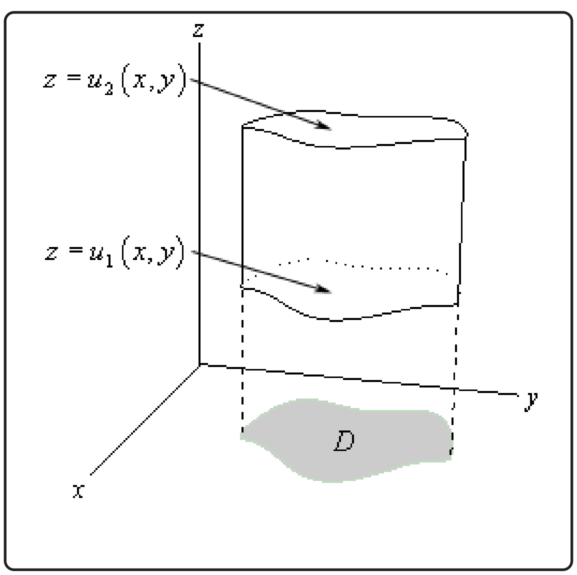
Before moving on to more general regions let's get a nice geometric interpretation about the triple integral out of the way so we can use it in some of the examples to follow.

Fact

The volume of the three-dimensional region E is given by the integral,

$$V = \iiint\limits_E dV$$

Let's now move on the more general three-dimensional regions. We have three different possibilities for a general region. Here is a sketch of the first possibility.



In this case we define the region E as follows,

$$E = \{(x, y, z) \,|\, (x, y) \in D, \;\; u_1(x, y) \le z \le u_2(x, y)\}$$

where $(x, y) \in D$ is the notation that means that the point (x, y) lies in the region D from the xy-plane. In this case we will evaluate the triple integral as follows,

$$\iiint\limits_E f(x, y, z)\, dV = \iint\limits_D \left[\int_{u_1(x,y)}^{u_2(x,y)} f(x, y, z)\, dz\right] dA$$

where the double integral can be evaluated in any of the methods that we saw in the previous couple of sections. In other words, we can integrate first with respect to x, we can integrate first with respect to y, or we can use polar coordinates as needed.

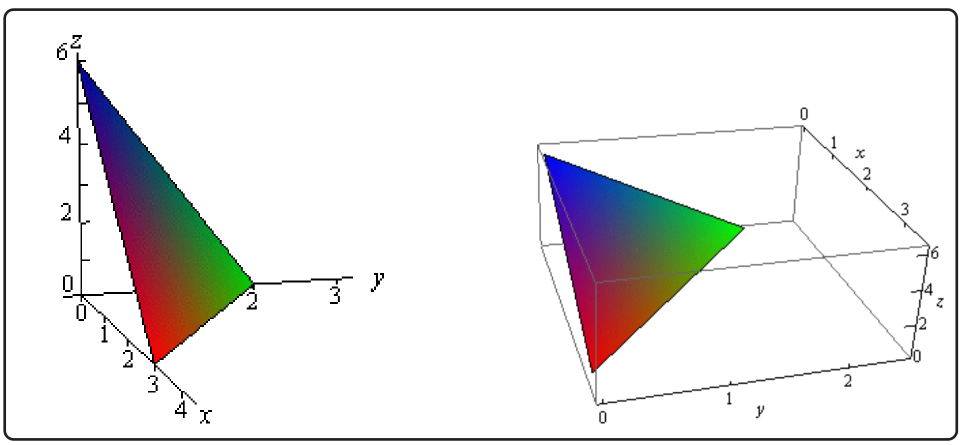
Example 2

Evaluate $\displaystyle\iiint\limits_E 2x\, dV$ where E is the region under the plane $2x + 3y + z = 6$ that lies in the first octant.

Solution

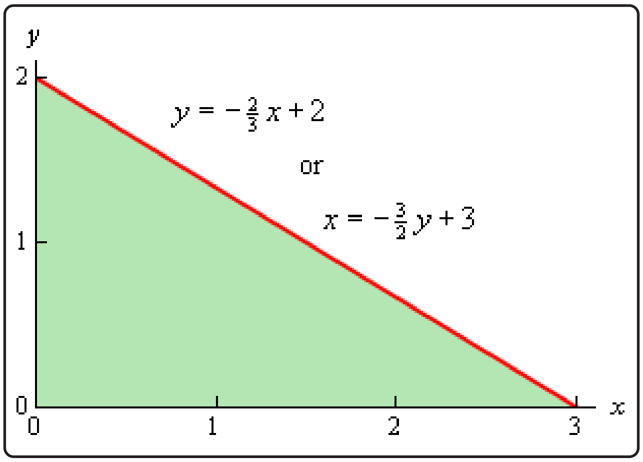
We should first define *octant*. Just as the two-dimensional coordinates system can be divided into four quadrants the three-dimensional coordinate system can be divided into eight octants. The first octant is the octant in which all three of the coordinates are positive.

Here is a sketch of the plane in the first octant.



We now need to determine the region D in the xy-plane. We can get a visualization of the region by pretending to look straight down on the object from above. What we see will be the region D in the xy-plane. So D will be the triangle with vertices at $(0, 0)$, $(3, 0)$, and $(0, 2)$.

Here is a sketch of D.



Now we need the limits of integration. Since we are under the plane and in the first octant (so we're above the plane $z = 0$) we have the following limits for z.

$$0 \le z \le 6 - 2x - 3y$$

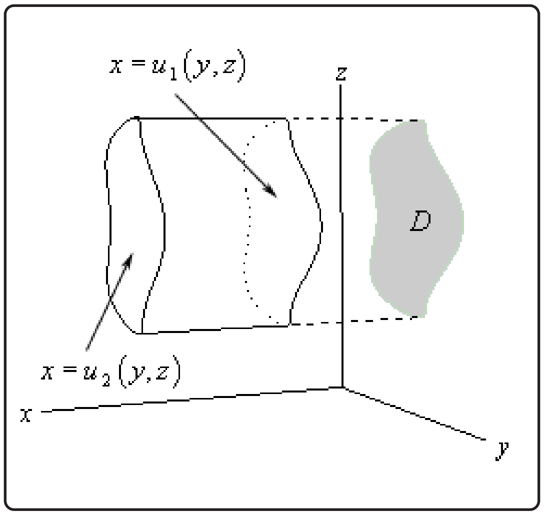
We can integrate the double integral over D using either of the following two sets of inequalities.

$$\begin{array}{cc} 0 \le x \le 3 & 0 \le x \le -\dfrac{3}{2}y + 3 \\ 0 \le y \le -\dfrac{2}{3}x + 2 & 0 \le y \le 2 \end{array}$$

Since neither really holds an advantage over the other we'll use the first one. The integral is then,

$$\begin{aligned} \iiint\limits_E 2x \, dV &= \iint\limits_D \left[\int_0^{6-2x-3y} 2x \, dz \right] dA \\ &= \iint\limits_D 2xz \Big|_0^{6-2x-3y} dA \\ &= \int_0^3 \int_0^{-\frac{2}{3}x+2} 2x\left(6 - 2x - 3y\right) dy\, dx \\ &= \int_0^3 \left(12xy - 4x^2y - 3xy^2\right) \Big|_0^{-\frac{2}{3}x+2} dx \\ &= \int_0^3 \frac{4}{3}x^3 - 8x^2 + 12x \, dx \\ &= \left(\frac{1}{3}x^4 - \frac{8}{3}x^3 + 6x^2\right) \Big|_0^3 \\ &= 9 \end{aligned}$$

Let's now move onto the second possible three-dimensional region we may run into for triple integrals. Here is a sketch of this region.



For this possibility we define the region E as follows,

$$E = \{(x, y, z) \,|\, (y, z) \in D, \;\; u_1(y, z) \le x \le u_2(y, z)\}$$

So, the region D will be a region in the yz-plane. Here is how we will evaluate these integrals.

$$\iiint\limits_E f(x, y, z)\, dV = \iint\limits_D \left[\int_{u_1(y,z)}^{u_2(y,z)} f(x, y, z)\, dx\right] dA$$

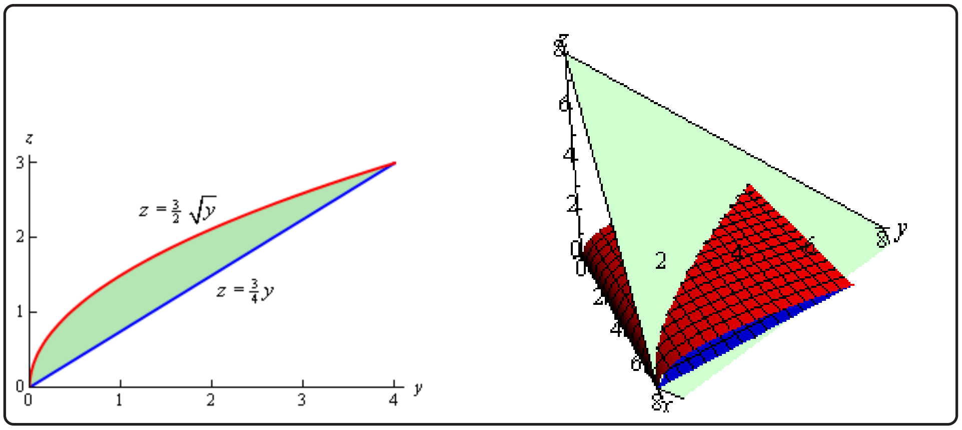
As with the first possibility we will have two options for doing the double integral in the yz-plane as well as the option of using polar coordinates if needed.

Example 3

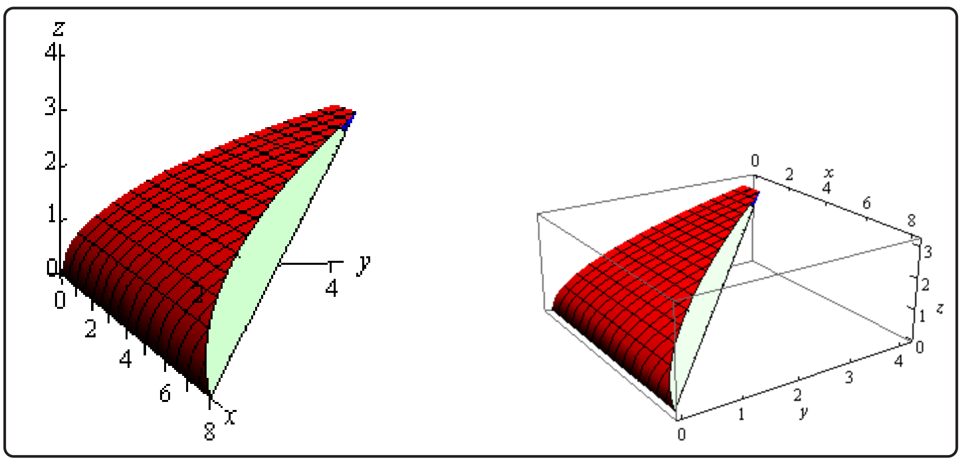
Determine the volume of the region that lies behind the plane $x + y + z = 8$ and in front of the region in the yz-plane that is bounded by $z = \frac{3}{2}\sqrt{y}$ and $z = \frac{3}{4}y$.

Solution

In this case we've been given D and so we won't have to really work to find that. Here is a sketch of the region D as well as a quick sketch of the plane and the curves defining D projected out past the plane so we can get an idea of what the region we're dealing with looks like.



Now, the graph of the region above is all okay, but it doesn't really show us what the region is. So, here is a sketch of the region itself.



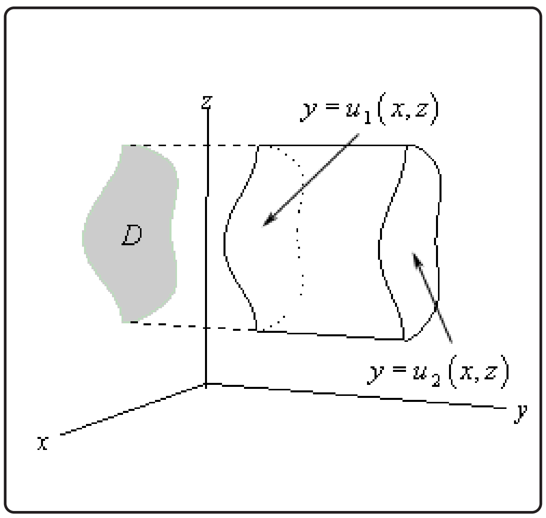
Here are the limits for each of the variables.

$$
\begin{gathered}
0 \le y \le 4 \\
\frac{3}{4}y \le z \le \frac{3}{2}\sqrt{y} \\
0 \le x \le 8 - y - z
\end{gathered}
$$

The volume is then,

$$\begin{aligned} V &= \iiint\limits_E dV = \iint\limits_D \left[\int_0^{8-y-z} dx\right] dA \\ &= \int_0^4 \int_{3y/4}^{3\sqrt{y}/2} 8 - y - z \, dz \, dy \\ &= \int_0^4 \left(8z - yz - \frac{1}{2}z^2\right)\Big|_{\frac{3y}{4}}^{\frac{3\sqrt{y}}{2}} dy \\ &= \int_0^4 12y^{\frac{1}{2}} - \frac{57}{8}y - \frac{3}{2}y^{\frac{3}{2}} + \frac{33}{32}y^2 \, dy \\ &= \left(8y^{\frac{3}{2}} - \frac{57}{16}y^2 - \frac{3}{5}y^{\frac{5}{2}} + \frac{11}{32}y^3\right)\Big|_0^4 = \frac{49}{5} \end{aligned}$$

We now need to look at the third (and final) possible three-dimensional region we may run into for triple integrals. Here is a sketch of this region.



In this final case E is defined as,

$$E = \{(x,y,z) \,|\, (x,z) \in D, \;\; u_1(x,z) \le y \le u_2(x,z)\}$$

and here the region D will be a region in the xz-plane. Here is how we will evaluate these integrals.

$$\iiint\limits_E f(x,y,z) \, dV = \iint\limits_D \left[\int_{u_1(x,z)}^{u_2(x,z)} f(x,y,z) \, dy\right] dA$$

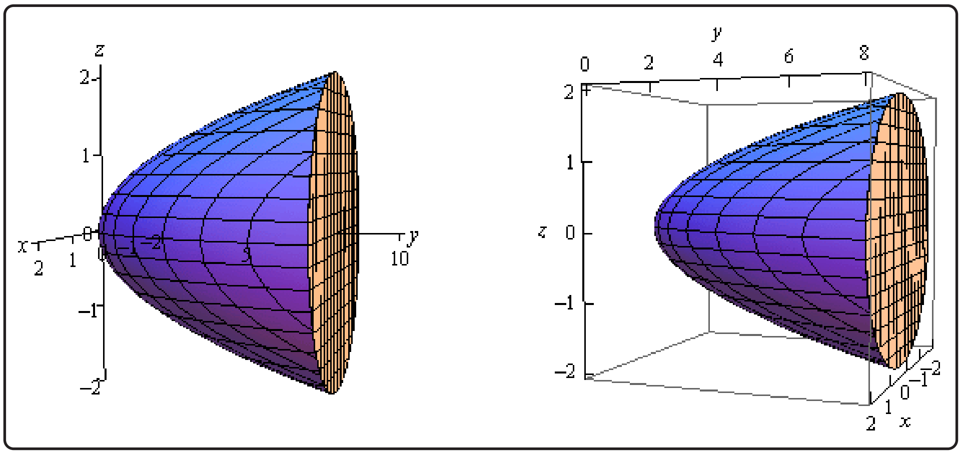
where we will can use either of the two possible orders for integrating D in the xz-plane or we can use polar coordinates if needed.

Example 4

Evaluate $\displaystyle\iiint\limits_E \sqrt{3x^2+3z^2}\,dV$ where E is the solid bounded by $y = 2x^2 + 2z^2$ and the plane $y = 8$.

Solution

Here is a sketch of the solid E.



The region D in the xz-plane can be found by "standing" in front of this solid and we can see that D will be a disk in the xz-plane. This disk will come from the front of the solid and we can determine the equation of the disk by setting the elliptic paraboloid and the plane equal.

$$2x^2 + 2z^2 = 8 \qquad \Rightarrow \qquad x^2 + z^2 = 4$$

This region, as well as the integrand, both seems to suggest that we should use something like polar coordinates. However, we are in the xz-plane and we've only seen polar coordinates in the xy-plane. This is not a problem. We can always "translate" them over to the xz-plane with the following definition.

$$x = r\cos(\theta) \qquad\qquad z = r\sin(\theta)$$

Since the region doesn't have y's we will let z take the place of y in all the formulas. Note that these definitions also lead to the formula,

$$x^2 + z^2 = r^2$$

With this in hand we can arrive at the limits of the variables that we'll need for this integral.

$$\begin{gathered} 2x^2 + 2z^2 \le y \le 8 \\ 0 \le r \le 2 \\ 0 \le \theta \le 2\pi \end{gathered}$$

The integral is then,

$$\begin{aligned} \iiint\limits_E \sqrt{3x^2+3z^2}\,dV &= \iint\limits_D \left[\int_{2x^2+2z^2}^{8} \sqrt{3x^2+3z^2}\,dy\right] dA \\ &= \iint\limits_D \left(y\sqrt{3x^2+3z^2}\right)\Big|_{2x^2+2z^2}^{8} dA \\ &= \iint\limits_D \sqrt{3\left(x^2+z^2\right)}\left(8-\left(2x^2+2z^2\right)\right)\,dA \end{aligned}$$

Now, since we are going to do the double integral in polar coordinates let's get everything converted over to polar coordinates. The integrand is,

$$\begin{aligned} \sqrt{3\left(x^2+z^2\right)}\left(8-\left(2x^2+2z^2\right)\right) &= \sqrt{3r^2}\left(8-2r^2\right) \\ &= \sqrt{3}\,r\left(8-2r^2\right) \\ &= \sqrt{3}\left(8r-2r^3\right) \end{aligned}$$

The integral is then,

$$\begin{aligned} \iiint\limits_E \sqrt{3x^2+3z^2}\,dV &= \iint\limits_D \sqrt{3}\ \left(8r-2r^3\right)dA \\ &= \sqrt{3}\int_0^{2\pi}\int_0^2 \left(8r-2r^3\right) r\,dr\,d\theta \\ &= \sqrt{3}\int_0^{2\pi} \left(\frac{8}{3}r^3-\frac{2}{5}r^5\right)\Big|_0^2 d\theta \\ &= \sqrt{3}\int_0^{2\pi} \frac{128}{15}\,d\theta \\ &= \frac{256\sqrt{3}\,\pi}{15} \end{aligned}$$

15.6 Triple Integrals in Cylindrical Coordinates

In this section we want do take a look at triple integrals done completely in Cylindrical Coordinates. Recall that cylindrical coordinates are really nothing more than an extension of polar coordinates into three dimensions. The following are the conversion formulas for cylindrical coordinates.

$$x = r\cos(\theta) \qquad y = r\sin(\theta) \qquad z = z$$

In order to do the integral in cylindrical coordinates we will need to know what dV will become in terms of cylindrical coordinates. We will be able to show in the Change of Variables section of this chapter that,

Fact

$$dV = r\,dz\,dr\,d\theta$$

The region, E, over which we are integrating becomes,

$$\begin{aligned} E &= \{(x,y,z)\,|\,(x,y)\in D,\ \ u_1(x,y)\le z\le u_2(x,y)\} \\ &= \left\{(r,\theta,z)\,|\alpha\le\theta\le\beta,\ h_1(\theta)\le r\le h_2(\theta),\ \ u_1\big(r\cos(\theta),r\sin(\theta)\big)\le z\le u_2\big(r\cos(\theta),r\sin(\theta)\big)\right\}\end{aligned}$$

Note that we've only given this for E's in which D is in the xy-plane. We can modify this accordingly if D is in the yz-plane or the xz-plane as needed.

In terms of cylindrical coordinates a triple integral is,

Fact

$$\iiint\limits_E f(x,y,z)\,dV = \int_\alpha^\beta\int_{h_1(\theta)}^{h_2(\theta)}\int_{u_1\left(r\cos(\theta),r\sin(\theta)\right)}^{u_2\left(r\cos(\theta),r\sin(\theta)\right)} r\,f\left(r\cos(\theta),r\sin(\theta),z\right)\,dz\,dr\,d\theta$$

Don't forget to add in the r and make sure that all the x's and y's also get converted over into cylindrical coordinates.

Let's see an example.

Example 1

Evaluate $\displaystyle\iiint\limits_E y\,dV$ where E is the region that lies below the plane $z = x + 2$ above the xy-plane and between the cylinders $x^2 + y^2 = 1$ and $x^2 + y^2 = 4$.

Solution

There really isn't too much to do with this one other than do the conversions and then evaluate the integral.

We'll start out by getting the range for z in terms of cylindrical coordinates.

$$0 \le z \le x + 2 \qquad \Rightarrow \qquad 0 \le z \le r\cos(\theta) + 2$$

Remember that we are above the xy-plane and so we are above the plane $z = 0$

Next, the region D is the region between the two circles $x^2 + y^2 = 1$ and $x^2 + y^2 = 4$ in the xy-plane and so the ranges for it are,

$$0 \le \theta \le 2\pi \qquad 1 \le r \le 2$$

Here is the integral.

$$\begin{aligned}
\iiint\limits_E y\,dV &= \int_0^{2\pi}\int_1^2\int_0^{r\cos(\theta)+2} (r\sin(\theta))\,r\,dz\,dr\,d\theta \\
&= \int_0^{2\pi}\int_1^2 r^2\sin(\theta)\big(r\cos(\theta)+2\big)\,dr\,d\theta \\
&= \int_0^{2\pi}\int_1^2 \frac{1}{2}r^3\sin(2\theta) + 2r^2\sin(\theta)\,dr\,d\theta \\
&= \int_0^{2\pi} \left(\frac{1}{8}r^4\sin(2\theta) + \frac{2}{3}r^3\sin(\theta)\right)\Bigg|_1^2\,d\theta \\
&= \int_0^{2\pi} \frac{15}{8}\sin(2\theta) + \frac{14}{3}\sin(\theta)\,d\theta \\
&= \left(-\frac{15}{16}\cos(2\theta) - \frac{14}{3}\cos(\theta)\right)\Bigg|_0^{2\pi} \\
&= 0
\end{aligned}$$

Just as we did with double integral involving polar coordinates we can start with an iterated integral in terms of x, y, and z and convert it to cylindrical coordinates.

Example 2

Convert $\displaystyle\int_{-1}^{1}\int_{0}^{\sqrt{1-y^2}}\int_{x^2+y^2}^{\sqrt{x^2+y^2}} xyz\,dz\,dx\,dy$ into an integral in cylindrical coordinates.

Solution

Here are the ranges of the variables from this iterated integral.

$$\begin{gathered} -1 \le y \le 1 \\ 0 \le x \le \sqrt{1-y^2} \\ x^2+y^2 \le z \le \sqrt{x^2+y^2} \end{gathered}$$

The first two inequalities define the region D and since the upper and lower bounds for the x's are $x = \sqrt{1-y^2}$ and $x = 0$ we know that we've got at least part of the right half a circle of radius 1 centered at the origin. Since the range of y's is $-1 \le y \le 1$ we know that we have the complete right half of the disk of radius 1 centered at the origin. So, the ranges for D in cylindrical coordinates are,

$$\begin{gathered} -\frac{\pi}{2} \le \theta \le \frac{\pi}{2} \\ 0 \le r \le 1 \end{gathered}$$

All that's left to do now is to convert the limits of the z range, but that's not too bad.

$$r^2 \le z \le r$$

On a side note notice that the lower bound here is an elliptic paraboloid and the upper bound is a cone. Therefore, E is a portion of the region between these two surfaces.

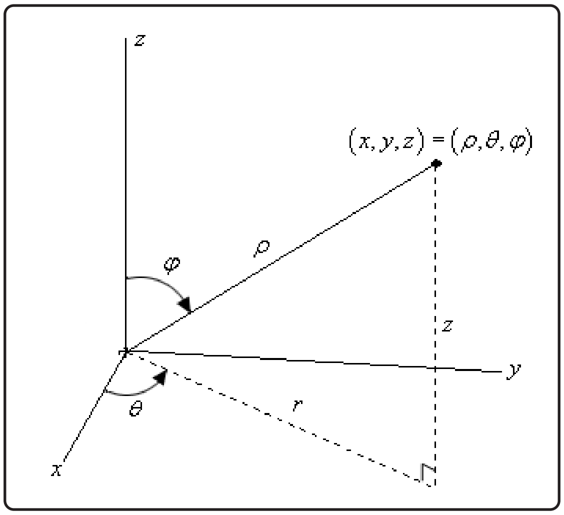
The integral is,

$$\begin{aligned} \int_{-1}^{1}\int_{0}^{\sqrt{1-y^2}}\int_{x^2+y^2}^{\sqrt{x^2+y^2}} xyz\,dz\,dx\,dy &= \int_{-\pi/2}^{\pi/2}\int_{0}^{1}\int_{r^2}^{r} r\left(r\cos(\theta)\right)\left(r\sin(\theta)\right)z\,dz\,dr\,d\theta \\ &= \int_{-\pi/2}^{\pi/2}\int_{0}^{1}\int_{r^2}^{r} zr^3\cos(\theta)\sin(\theta)\,dz\,dr\,d\theta \end{aligned}$$

15.7 Triple Integrals in Spherical Coordinates

In the previous section we looked at doing integrals in terms of cylindrical coordinates and we now need to take a quick look at doing integrals in terms of spherical coordinates.

First, we need to recall just how spherical coordinates are defined. The following sketch shows the relationship between the Cartesian and spherical coordinate systems.



Here are the conversion formulas for spherical coordinates.

$$x = \rho \sin(\varphi)\cos(\theta) \qquad y = \rho \sin(\varphi)\sin(\theta) \qquad z = \rho\cos(\varphi)$$
$$x^2 + y^2 + z^2 = \rho^2$$

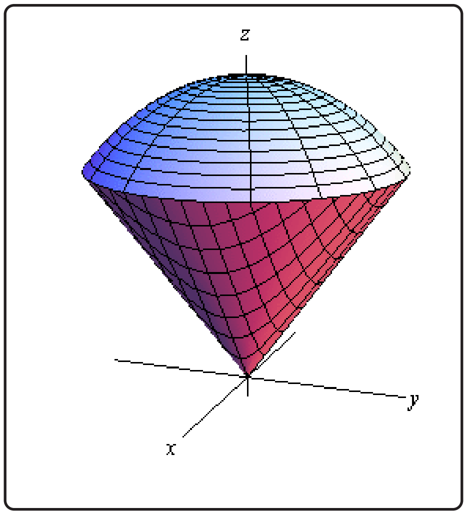
We also have the following restrictions on the coordinates.

$$\rho \geq 0 \qquad\qquad 0 \leq \varphi \leq \pi$$

For our integrals we are going to restrict E down to a spherical wedge. This will mean that we are going to take ranges for the variables as follows,

$$\begin{gathered} a \leq \rho \leq b \\ \alpha \leq \theta \leq \beta \\ \delta \leq \varphi \leq \gamma \end{gathered}$$

Here is a quick sketch of a spherical wedge in which the lower limit for both ρ and φ are zero for reference purposes. Most of the wedges we'll be working with will fit into this pattern.



From this sketch we can see that E is nothing more than the intersection of a sphere and a cone and generally will represent a shape that is reminiscent of an ice cream cone.

In the next section we will show that

Fact

$$dV = \rho^2 \sin(\varphi)\, d\rho\, d\theta\, d\varphi$$

Therefore, the integral will become,

Fact

$$\iiint\limits_E f(x, y, z)\, dV$$
$$= \int_\delta^\gamma \int_\alpha^\beta \int_a^b \rho^2 \sin(\varphi)\; f\left(\rho\sin(\varphi)\cos(\theta), \rho\sin(\varphi)\sin(\theta), \rho\cos(\varphi)\right) d\rho\, d\theta\, d\varphi$$

This looks bad but given that the limits are all constants the integrals here tend to not be too bad.

Example 1

Evaluate $\displaystyle\iiint\limits_E 16z\, dV$ where E is the upper half of the sphere $x^2 + y^2 + z^2 = 1$.

Solution

Since we are taking the upper half of the sphere the limits for the variables are,

$$0 \le \rho \le 1$$
$$0 \le \theta \le 2\pi$$
$$0 \le \varphi \le \frac{\pi}{2}$$

The integral is then,

$$\begin{aligned}
\iiint\limits_E 16z\,dV &= \int_0^{\frac{\pi}{2}} \int_0^{2\pi} \int_0^1 \rho^2 \sin(\varphi)\left(16\rho\cos(\varphi)\right)\,d\rho\,d\theta\,d\varphi \\
&= \int_0^{\frac{\pi}{2}} \int_0^{2\pi} \int_0^1 8\rho^3 \sin(2\varphi)\,\,d\rho\,d\theta\,d\varphi \\
&= \int_0^{\frac{\pi}{2}} \int_0^{2\pi} 2\sin(2\varphi)\,\,d\theta\,d\varphi \\
&= \int_0^{\frac{\pi}{2}} 4\pi\sin(2\varphi)\,\,d\varphi \\
&= -2\pi\cos(2\varphi)\Big|_0^{\frac{\pi}{2}} \\
&= 4\pi
\end{aligned}$$

Example 2

Evaluate $\iiint\limits_E z\,x\,dV$ where E is inside both $x^2+y^2+z^2 = 4$ and the cone (pointing upward) that makes an angle of $\frac{\pi}{3}$ with the negative z-axis and has $x \le 0$.

Solution

First, we need to take care of the limits. The region E is basically an upside down ice cream cone that has been cut in half so that only the portion with $x \le 0$ remains. Therefore, because we are inside a portion of a sphere of radius 2 we must have,

$$0 \le \rho \le 2$$

For φ we need to be careful. The problem statement says that the cone makes an angle of $\frac{\pi}{3}$ with the negative z-axis. However, remember that φ is measured from the positive z-axis. Therefore, the first angle, as measured from the positive z-axis, that will "start" the cone will be $\varphi = \frac{2\pi}{3}$ and it goes to the negative z-axis. Therefore, we get the following limits for φ.

$$\frac{2\pi}{3} \le \varphi \le \pi$$

Finally, for the θ we can use the fact that we are also told that $x \leq 0$. This means we are to the left of the y-axis and so the range of θ must be,

$$\frac{\pi}{2} \leq \theta \leq \frac{3\pi}{2}$$

Now that we have the limits we can evaluate the integral.

$$\begin{aligned}
\iiint\limits_E z\,x\,dV &= \int_{\frac{2\pi}{3}}^{\pi} \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \int_0^2 \left(\rho\cos(\varphi)\right)\left(\rho\sin(\varphi)\cos(\theta)\right)\rho^2 \sin(\varphi)\,d\rho\,d\theta\,d\varphi \\
&= \int_{\frac{2\pi}{3}}^{\pi} \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \int_0^2 \rho^4 \cos(\varphi)\sin^2(\varphi)\cos(\theta)\,d\rho\,d\theta\,d\varphi \\
&= \int_{\frac{2\pi}{3}}^{\pi} \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \frac{32}{5}\cos(\varphi)\sin^2(\varphi)\cos(\theta)\,d\theta\,d\varphi \\
&= \int_{\frac{2\pi}{3}}^{\pi} -\frac{64}{5}\cos(\varphi)\sin^2(\varphi)\,d\varphi \\
&= -\frac{64}{15}\sin^3(\varphi)\Big|_{\frac{2\pi}{3}}^{\pi} \\
&= \frac{8\sqrt{3}}{5}
\end{aligned}$$

Example 3

Convert $\displaystyle\int_0^3 \int_0^{\sqrt{9-y^2}} \int_{\sqrt{x^2+y^2}}^{\sqrt{18-x^2-y^2}} x^2+y^2+z^2\,dz\,dx\,dy$ into spherical coordinates.

Solution

Let's first write down the limits for the variables.

$$\begin{gathered}
0 \leq y \leq 3 \\
0 \leq x \leq \sqrt{9-y^2} \\
\sqrt{x^2+y^2} \leq z \leq \sqrt{18-x^2-y^2}
\end{gathered}$$

The range for x tells us that we have a portion of the right half of a disk of radius 3 centered at the origin. Since we are restricting y's to positive values it looks like we will have the quarter disk in the first quadrant. Therefore, since D is in the first quadrant the region, E, must be in the first octant and this in turn tells us that we have the following range for θ (since this is the angle around the z-axis).

$$0 \leq \theta \leq \frac{\pi}{2}$$

Now, let's see what the range for z tells us. The lower bound, $z = \sqrt{x^2 + y^2}$, is the upper half of a cone. At this point we don't need this quite yet, but we will later. The upper bound, $z = \sqrt{18 - x^2 - y^2}$, is the upper half of the sphere,

$$x^2 + y^2 + z^2 = 18$$

and so from this we now have the following range for ρ

$$0 \le \rho \le \sqrt{18} = 3\sqrt{2}$$

Now all that we need is the range for φ. There are two ways to get this. One is from where the cone and the sphere intersect. Plugging in the equation for the cone into the sphere gives,

$$\begin{aligned} \left(\sqrt{x^2 + y^2}\right)^2 + z^2 &= 18 \\ z^2 + z^2 &= 18 \\ z^2 &= 9 \qquad \Rightarrow \qquad z = \pm 3 \end{aligned}$$

Note that we can assume z is positive here since we know that we have the upper half of the cone and/or sphere. Finally, plug this into the conversion for z and take advantage of the fact that we know that $\rho = 3\sqrt{2}$ since we are intersecting on the sphere. This gives,

$$\begin{aligned} \rho \cos(\varphi) &= 3 \\ 3\sqrt{2} \cos(\varphi) &= 3 \\ \cos(\varphi) &= \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2} \qquad \Rightarrow \qquad \varphi = \frac{\pi}{4} \end{aligned}$$

So, it looks like we have the following range,

$$0 \le \varphi \le \frac{\pi}{4}$$

The other way to get this range is from the cone by itself. By first converting the equation into cylindrical coordinates and then into spherical coordinates we get the following,

$$\begin{aligned} z &= r \\ \rho \cos(\varphi) &= \rho \sin(\varphi) \\ 1 &= \tan(\varphi) \qquad \Rightarrow \qquad \varphi = \frac{\pi}{4} \end{aligned}$$

So, recalling that $\rho^2 = x^2 + y^2 + z^2$, the integral is then,

$$\int_0^3 \int_0^{\sqrt{9-y^2}} \int_{\sqrt{x^2+y^2}}^{\sqrt{18-x^2-y^2}} x^2 + y^2 + z^2 \, dz \, dx \, dy = \int_0^{\pi/4} \int_0^{\pi/2} \int_0^{3\sqrt{2}} \rho^4 \sin(\varphi) \, d\rho \, d\theta \, d\varphi$$

15.8 Change of Variables

Back in Calculus I we had the substitution rule that told us that,

$$\int_a^b f(g(x))\, g'(x)\, dx = \int_c^d f(u)\, du \qquad \text{where } u = g(x)$$

In essence this is taking an integral in terms of x's and changing it into terms of u's. We want to do something similar for double and triple integrals. In fact we've already done this to a certain extent when we converted double integrals to polar coordinates and when we converted triple integrals to cylindrical or spherical coordinates. The main difference is that we didn't actually go through the details of where the formulas came from. If you recall, in each of those cases we commented that we would justify the formulas for dA and dV eventually. Now is the time to do that justification.

While often the reason for changing variables is to get us an integral that we can do with the new variables, another reason for changing variables is to convert the region into a nicer region to work with. When we were converting the polar, cylindrical or spherical coordinates we didn't worry about this change since it was easy enough to determine the new limits based on the given region. That is not always the case however. So, before we move into changing variables with multiple integrals we first need to see how the region may change with a change of variables.

First, we need a little terminology/notation out of the way. We call the equations that define the change of variables a **transformation**. Also, we will typically start out with a region, R, in xy-coordinates and transform it into a region in uv-coordinates.

Example 1

Determine the new region that we get by applying the given transformation to the region R.

(a) R is the ellipse $x^2 + \dfrac{y^2}{36} = 1$ and the transformation is $x = \dfrac{u}{2}$, $y = 3v$.

(b) R is the region bounded by $y = -x+4$, $y = x+1$, and $y = \dfrac{x}{3} - \dfrac{4}{3}$ and the transformation is $x = \dfrac{1}{2}(u+v)$, $y = \dfrac{1}{2}(u-v)$.

Solution

(a) R is the ellipse $x^2 + \dfrac{y^2}{36} = 1$ and the transformation is $x = \dfrac{u}{2}$, $y = 3v$.

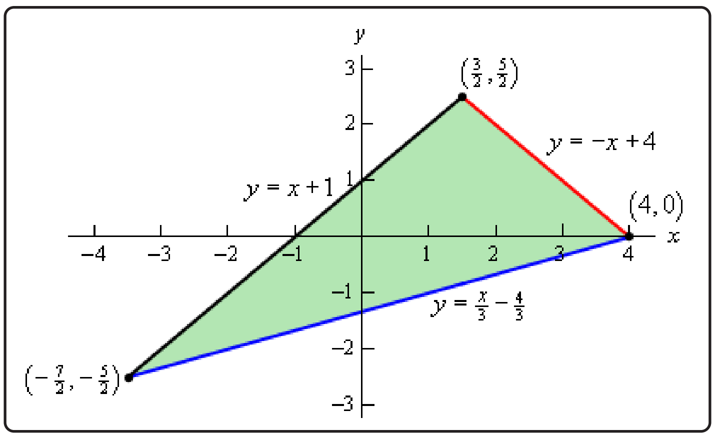
There really isn't too much to do with this one other than to plug the transformation into the equation for the ellipse and see what we get.

$$\begin{aligned}\left(\frac{u}{2}\right)^2 + \frac{(3v)^2}{36} &= 1\\ \frac{u^2}{4} + \frac{9v^2}{36} &= 1\\ u^2 + v^2 &= 4\end{aligned}$$

So, we started out with an ellipse and after the transformation we had a disk of radius 2.

(b) R is the region bounded by $y = -x+4$, $y = x+1$, and $y = \frac{x}{3} - \frac{4}{3}$ and the transformation is $x = \frac{1}{2}(u+v)$, $y = \frac{1}{2}(u-v)$.

As with the first part we'll need to plug the transformation into the equation, however, in this case we will need to do it three times, once for each equation. Before we do that let's sketch the graph of the region and see what we've got.



So, we have a triangle. Now, let's go through the transformation. We will apply the transformation to each edge of the triangle and see where we get.

Let's do $y = -x + 4$ first. Plugging in the transformation gives,

$$\begin{aligned}\frac{1}{2}(u-v) &= -\frac{1}{2}(u+v) + 4\\ u - v &= -u - v + 8\\ 2u &= 8\\ u &= 4\end{aligned}$$

The first boundary transforms very nicely into a much simpler equation.

Now let's take a look at $y = x + 1$,

$$\begin{aligned}\frac{1}{2}(u - v) &= \frac{1}{2}(u + v) + 1 \\ u - v &= u + v + 2 \\ -2v &= 2 \\ v &= -1\end{aligned}$$

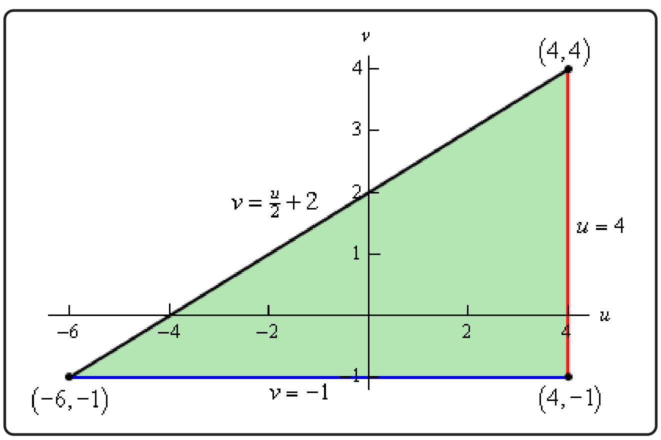
Again, a much nicer equation that what we started with.

Finally, let's transform $y = \frac{x}{3} - \frac{4}{3}$.

$$\begin{aligned}\frac{1}{2}(u - v) &= \frac{1}{3}\left(\frac{1}{2}(u + v)\right) - \frac{4}{3} \\ 3u - 3v &= u + v - 8 \\ 4v &= 2u + 8 \\ v &= \frac{u}{2} + 2\end{aligned}$$

So, again, we got a somewhat simpler equation, although not quite as nice as the first two.

Let's take a look at the new region that we get under the transformation.



We still get a triangle, but a much nicer one.

Note that we can't always expect to transform a specific type of region (a triangle for example) into the same kind of region. It is completely possible to have a triangle transform into a region in which each of the edges are curved and in no way resembles a triangle.

Notice that in each of the above examples we took a two dimensional region that would have been somewhat difficult to integrate over and converted it into a region that would be much nicer in integrate over. As we noted at the start of this set of examples, that is often one of the points behind the transformation. In addition to converting the integrand into something simpler it will often also transform the region into one that is much easier to deal with.

Before proceeding with the next topic let's address another point. On occasion, we will also need to know the range of u and/or v for each of the new equations we get from the transformation. We didn't need that for the two examples above and it is not something that we will often need. However, it can help on occasion in determining the new region.

So, let's work a quick example to see how we do that.

Example 2

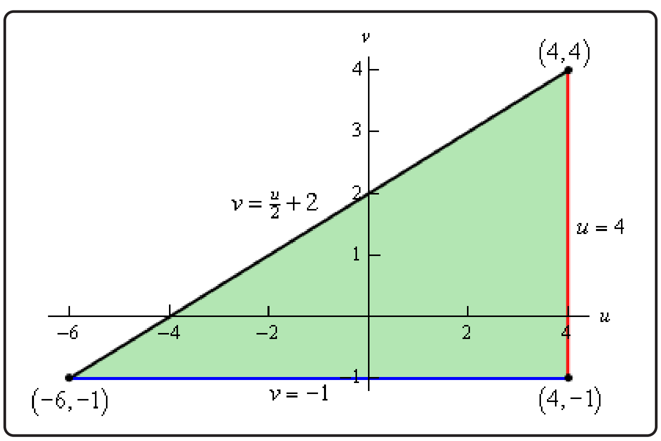
For the region bounded by $y = -x + 4$, $y = x + 1$, and $y = \frac{x}{3} - \frac{4}{3}$ and the transformation $x = \frac{1}{2}(u + v)$, $y = \frac{1}{2}(u - v)$ determine the ranges of u and v for each of the new equations from the transformation.

Solution

Okay, we already know what the new region looks like and what the new equations are from the previous example. So, here is a quick review of the transformation of each of the original equations.

$$\begin{aligned} y &= -x + 4 & \Rightarrow \quad & u = 4 \\ y &= x + 1 & \Rightarrow \quad & v = -1 \\ y &= \frac{x}{3} - \frac{4}{3} & \Rightarrow \quad & v = \frac{u}{2} + 2 \end{aligned}$$

Here is the new region we get under the transformation.



Note, that, in this case, we could determine the range of u and v for each equation from the sketch above. However, in cases where we might actually need the ranges that is usually not an option as we often need the ranges for u and/or v to get an accurate sketch of the new region.

So, let's now actually start working the problem.

Let's start with the equation $u = 4$. First, we don't need a "range" of u's here as equation makes it pretty clear we have a single value of u, namely $u = 4$. So, let's determine the range of v's we should get.

Let's start with the x transformation and plug in the known value of u for this equation. That gives,

$$x = \frac{1}{2}(u+v) = \frac{1}{2}(4+v)$$

Now, we know that the range of x's for the original equation, $y = -x+4$, are $\frac{3}{2} \le x \le 4$. We also know from above what x is in terms of v, so plug that into this range and do a little manipulation as follows,

$$\begin{aligned} \frac{3}{2} &\le x \le 4 \\ \frac{3}{2} \le \frac{1}{2}&(4+v) \le 4 \\ 3 \le\ &4+v \le 8 \\ -1 &\le v \le 4 \end{aligned}$$

So, the range of v's for $u = 4$ must be $-1 \le v \le 4$, which nicely matches with what we would expect from the graph of the new region.

Note that we could just as easily used the y transformation and y range for the original equation and gotten the same result.

Okay, let's now move onto $v = -1$ and we won't put in quite as much explanation for this part.

First, we don't need a range of v for this because we clearly have just a single value of v. So, to get the range of u let's again start with the x transformation, plug $v = -1$ in that and then use the range of x's from the original equation, $y = x+1$.

Here is that work.

$$\begin{aligned} -\frac{7}{2} &\le x \le \frac{3}{2} \\ -\frac{7}{2} \le \frac{1}{2}&(u-1) \le \frac{3}{2} \\ -7 \le\ &u-1 \le 3 \\ -6 &\le u \le 4 \end{aligned}$$

So, the range of u for $v = -1$ is $-6 \le u \le 4$ which, again, matches up with what we see on the graph. Also note that once again, we could have used the y ranges to do this work.

Finally, let's find the range of u and v for $v = \frac{u}{2} + 2$. This time let's use the y transformation so we can say we used that in one these. So, we'll start with the range of y's for the original equation, $y = \frac{x}{3} - \frac{4}{3}$, plug in the y transformation and then plug in for v. Doing this gives,

$$\begin{aligned} -\frac{5}{2} &\le y \le 0 \\ -\frac{5}{2} \le \frac{1}{2}&\left(u - \left(\frac{u}{2} + 2\right)\right) \le 0 \\ -5 &\le \frac{u}{2} - 2 \le 0 \\ -3 &\le \frac{u}{2} \le 2 \\ -6 &\le u \le 4 \end{aligned}$$

So, again we get the range of u's we expect to get from the graph. Once we have those the appropriate range of v can be found from the equation itself as follows,

$$\begin{aligned} -6 &\le u \le 4 \\ -3 &\le \frac{u}{2} \le 2 \\ -1 &\le \frac{u}{2} + 2 \le 4 \\ -1 &\le v \le 4 \end{aligned}$$

Basically, start with the range of u's and "build up" the equation for the side and we get the range of v's for this side.

So, we now know how to get ranges of u and/or v for new equations under a transformation. However, this is not something that is done terribly often but it is a useful skill to have in case it does arise somewhere.

Now that we've seen a couple of examples of transforming regions we need to now talk about how we actually do change of variables in the integral. We will start with double integrals. In order to change variables in a double integral we will need the **Jacobian** of the transformation. Here is the definition of the Jacobian.

Definition

The **Jacobian** of the transformation $x = g(u, v)$, $y = h(u, v)$ is

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}$$

The Jacobian is defined as a determinant of a 2x2 matrix, if you are unfamiliar with this that is okay.

Here is how to compute the determinant.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Therefore, another formula for the determinant is,

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

Now that we have the Jacobian out of the way we can give the formula for change of variables for a double integral.

Change of Variables for a Double Integral

Suppose that we want to integrate $f(x,y)$ over the region R. Under the transformation $x = g(u,v)$, $y = h(u,v)$ the region becomes S and the integral becomes,

$$\iint\limits_R f(x,y)\,dA = \iint\limits_S f(g(u,v), h(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| d\overline{A}$$

Note that we use $d\overline{A}$ in the u & v integral above to denote that it will be in terms of du and dv once we convert to two single integrals rather than the dx and dy we are used to using for dA. This is notational only and we generally just use dA for both and just make sure to remember that the "new" dA is in terms of du and dv.

Also note that we are taking the absolute value of the Jacobian.

If we look just at the differentials in the above formula we can also say that

$$dA = \left|\frac{\partial(x,y)}{\partial(u,v)}\right| d\overline{A}$$

Example 3

Show that when changing to polar coordinates we have $dA = r\,dr\,d\theta$

Solution

So, what we are doing here is justifying the formula that we used back when we were integrating with respect to polar coordinates. All that we need to do is use the formula above for dA.

The transformation here is the standard conversion formulas,

$$x = r\cos(\theta) \qquad\qquad y = r\sin(\theta)$$

The Jacobian for this transformation is,

$$\begin{aligned}\frac{\partial(x,y)}{\partial(r,\theta)} &= \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} \\ &= \begin{vmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{vmatrix} \\ &= r\cos^2(\theta) - \left(-r\sin^2(\theta)\right) \\ &= r\left(\cos^2(\theta) + \sin^2(\theta)\right) \\ &= r\end{aligned}$$

We then get,

$$dA = \left|\frac{\partial(x,y)}{\partial(r,\theta)}\right| dr\,d\theta = |r|\,dr\,d\theta = r\,dr\,d\theta$$

So, the formula we used in the section on polar integrals was correct.

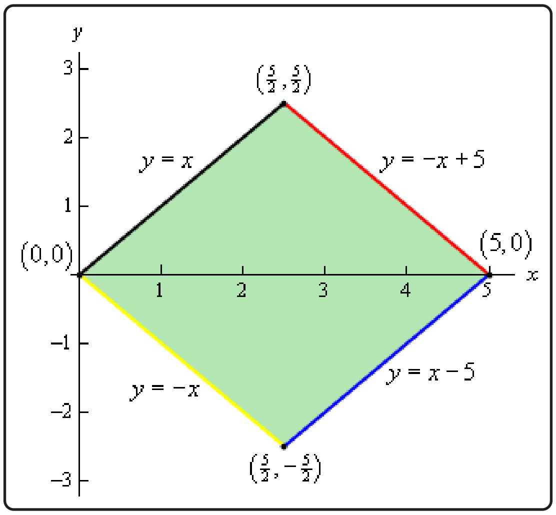
Now, let's do a couple of integrals.

Example 4

Evaluate $\displaystyle\iint\limits_R x + y\,dA$ where R is the trapezoidal region with vertices given by $(0,0)$, $(5,0)$, $\left(\frac{5}{2},\frac{5}{2}\right)$ and $\left(\frac{5}{2},-\frac{5}{2}\right)$ using the transformation $x = 2u + 3v$ and $y = 2u - 3v$.

Solution

First, let's sketch the region R and determine equations for each of the sides.



Each of the equations was found by using the fact that we know two points on each line (*i.e.* the two vertices that form the edge).

While we could do this integral in terms of x and y it would involve two integrals and so would be some work.

Let's use the transformation and see what we get. We'll do this by plugging the transformation into each of the equations above.

Let's start the process off with $y = x$.

$$\begin{aligned} 2u - 3v &= 2u + 3v \\ 6v &= 0 \\ v &= 0 \end{aligned}$$

Transforming $y = -x$ is similar.

$$\begin{aligned} 2u - 3v &= -(2u + 3v) \\ 4u &= 0 \\ u &= 0 \end{aligned}$$

Next, we'll transform $y = -x + 5$.

$$\begin{aligned} 2u - 3v &= -(2u + 3v) + 5 \\ 4u &= 5 \\ u &= \frac{5}{4} \end{aligned}$$

Finally, let's transform $y = x - 5$.

$$\begin{aligned} 2u - 3v &= 2u + 3v - 5 \\ -6v &= -5 \\ v &= \frac{5}{6} \end{aligned}$$

The region S is then a rectangle whose sides are given by $u = 0$, $v = 0$, $u = \frac{5}{4}$ and $v = \frac{5}{6}$ and so the ranges of u and v are,

$$0 \le u \le \frac{5}{4} \qquad\qquad 0 \le v \le \frac{5}{6}$$

Next, we need the Jacobian.

$$\frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} 2 & 3 \\ 2 & -3 \end{vmatrix} = -6 - 6 = -12$$

The integral is then,

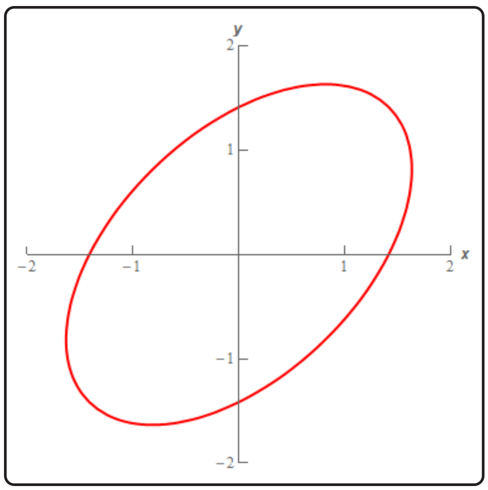
$$\begin{aligned} \iint\limits_R x + y\,dA &= \int_0^{\frac{5}{6}} \int_0^{\frac{5}{4}} \left((2u+3v) + (2u-3v)\right)|-12|\,d\overline{A} \\ &= \int_0^{\frac{5}{6}} \int_0^{\frac{5}{4}} 48u\,d\overline{A} \\ &= \int_0^{\frac{5}{6}} 24u^2\Big|_0^{\frac{5}{4}}\,dv \\ &= \int_0^{\frac{5}{6}} \frac{75}{2}\,dv \\ &= \frac{75}{2}v\Big|_0^{\frac{5}{6}} \\ &= \frac{125}{4} \end{aligned}$$

Example 5

Evaluate $\displaystyle\iint\limits_R x^2 - xy + y^2\, dA$ where R is the ellipse given by $x^2 - xy + y^2 \le 2$ and using the transformation $x = \sqrt{2}\,u - \sqrt{\frac{2}{3}}\,v$, $y = \sqrt{2}\,u + \sqrt{\frac{2}{3}}\,v$.

Solution

Before we proceed with this problem. Let's do a quick graph of the boundary of the region R. We claimed that it is an ellipse, but is clearly not in "standard" form. Here is the boundary of R.



So, it is an ellipse, just one that is at an angle rather than symmetric about the x and y-axis as we are used to dealing with.

Also, note that we used "≤ 2" when "defining" R to make it clear that we are using both the actual ellipse itself as well as the interior of the ellipse for R.

Okay, let's proceed with the problem.

The first thing to do is to plug the transformation into the equation for the ellipse to see what the region transforms into.

$$\begin{aligned}2 &= x^2 - xy + y^2\\ &= \left(\sqrt{2}\,u - \sqrt{\frac{2}{3}}\,v\right)^2 - \left(\sqrt{2}\,u - \sqrt{\frac{2}{3}}\,v\right)\left(\sqrt{2}\,u + \sqrt{\frac{2}{3}}\,v\right) + \left(\sqrt{2}\,u + \sqrt{\frac{2}{3}}\,v\right)^2\\ &= 2u^2 - \frac{4}{\sqrt{3}}uv + \frac{2}{3}v^2 - \left(2u^2 - \frac{2}{3}v^2\right) + 2u^2 + \frac{4}{\sqrt{3}}uv + \frac{2}{3}v^2\\ &= 2u^2 + 2v^2\end{aligned}$$

Or, upon dividing by 2 we see that the equation describing R transforms into

$$u^2 + v^2 = 1$$

or the unit circle. Again, this will be much easier to integrate over than the original region.

Note as well that we've shown that the function that we're integrating is

$$x^2 - xy + y^2 = 2\left(u^2 + v^2\right)$$

in terms of u and v so we won't have to redo that work when the time to do the integral comes around.

Finally, we need to find the Jacobian.

$$\frac{\partial (x,y)}{\partial (u,v)} = \begin{vmatrix} \sqrt{2} & -\sqrt{\frac{2}{3}} \\ \sqrt{2} & \sqrt{\frac{2}{3}} \end{vmatrix} = \frac{2}{\sqrt{3}} + \frac{2}{\sqrt{3}} = \frac{4}{\sqrt{3}}$$

The integral is then,

$$\iint\limits_R x^2 - xy + y^2 \, dA = \iint\limits_S 2\left(u^2 + v^2\right) \left|\frac{4}{\sqrt{3}}\right| du\, dv$$

Before proceeding a word of caution is in order. Do not make the mistake of substituting $x^2 - xy + y^2 = 2$ or $u^2 + v^2 = 1$ in for the integrand. These equations are only valid on the boundary of the region and we are looking at all the points interior to the boundary as well and for those points neither of these equations will be true!

At this point we'll note that this integral will be much easier in terms of polar coordinates and so to finish the integral out will convert to polar coordinates.

$$\begin{aligned} \iint\limits_R x^2 - xy + y^2 \, dA &= \iint\limits_S 2\left(u^2 + v^2\right) \left|\frac{4}{\sqrt{3}}\right| du\, dv \\ &= \frac{8}{\sqrt{3}} \int_0^{2\pi} \int_0^1 \left(r^2\right) r\, dr\, d\theta \\ &= \frac{8}{\sqrt{3}} \int_0^{2\pi} \left.\frac{1}{4} r^4 \right|_0^1 d\theta \\ &= \frac{8}{\sqrt{3}} \int_0^{2\pi} \frac{1}{4}\, d\theta \\ &= \frac{4\pi}{\sqrt{3}} \end{aligned}$$

Let's now briefly look at triple integrals. In this case we will again start with a region R and use the transformation $x = g(u, v, w)$, $y = h(u, v, w)$, and $z = k(u, v, w)$ to transform the region into the new region S. To do the integral we will need a Jacobian, just as we did with double integrals. Here is the definition of the Jacobian for this kind of transformation.

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}$$

In this case the Jacobian is defined in terms of the determinant of a 3x3 matrix. We saw how to evaluate these when we looked at cross products back in Calculus II. If you need a refresher on how to compute them you should go back and review that section.

The integral under this transformation is,

$$\iiint\limits_R f(x, y, z)\, dV = \iiint\limits_S f(g(u, v, w), h(u, v, w), k(u, v, w)) \left|\frac{\partial(x, y, z)}{\partial(u, v, w)}\right| d\overline{V}$$

As with double integrals we used $d\overline{V}$ in the u, v, w integral above to remind ourselves that we will need to use du, dv and dw when converting to single integrals. Again, this is just notation and is usually written as just dV.

We can look at just the differentials and note that we must have

$$dV = \left|\frac{\partial(x, y, z)}{\partial(u, v, w)}\right| d\overline{V}$$

We're not going to do any integrals here, but let's verify the formula for dV for spherical coordinates.

Example 6

Verify that $dV = \rho^2 \sin(\varphi)\, d\rho\, d\theta\, d\varphi$ when using spherical coordinates.

Solution

Here the transformation is just the standard conversion formulas.

$$x = \rho \sin(\varphi) \cos(\theta) \qquad y = \rho \sin(\varphi) \sin(\theta) \qquad z = \rho \cos(\varphi)$$

The Jacobian is,

$$\begin{aligned}\frac{\partial(x,y,z)}{\partial(\rho,\theta,\varphi)} &= \begin{vmatrix} \sin(\varphi)\cos(\theta) & -\rho\sin(\varphi)\sin(\theta) & \rho\cos(\varphi)\cos(\theta) \\ \sin(\varphi)\sin(\theta) & \rho\sin(\varphi)\cos(\theta) & \rho\cos(\varphi)\sin(\theta) \\ \cos(\varphi) & 0 & -\rho\sin(\varphi) \end{vmatrix} \\ &= -\rho^2\sin^3\varphi\cos^2\theta - \rho^2\sin(\varphi)\cos^2\varphi\sin^2\theta + 0 \\ &\qquad - \rho^2\sin^3\varphi\sin^2\theta - 0 - \rho^2\sin(\varphi)\cos^2\varphi\cos^2(\theta) \\ &= -\rho^2\sin^3\varphi\left(\cos^2(\theta)+\sin^2(\theta)\right) - \rho^2\sin(\varphi)\cos^2\varphi\left(\sin^2(\theta)+\cos^2(\theta)\right) \\ &= -\rho^2\sin^3(\varphi) - \rho^2\sin(\varphi)\cos^2\varphi \\ &= -\rho^2\sin(\varphi)\left(\sin^2(\varphi)+\cos^2(\varphi)\right) \\ &= -\rho^2\sin(\varphi)\end{aligned}$$

Finally, dV becomes,

$$dV = \left|-\rho^2\sin(\varphi)\right| d\rho\, d\theta\, d\varphi = \rho^2\sin(\varphi)\, d\rho\, d\theta\, d\varphi$$

Recall that we restricted φ to the range $0 \le \varphi \le \pi$ for spherical coordinates and so we know that $\sin(\varphi) \ge 0$ and so we don't need the absolute value bars on the sine.

We will leave it to you to check the formula for dV for cylindrical coordinates if you'd like to. It is a much easier formula to check.

15.9 Surface Area

In this section we will look at the lone application (aside from the area and volume interpretations) of multiple integrals in this material. This is not the first time that we've looked at surface area We first saw surface area in Calculus II, however, in that setting we were looking at the surface area of a solid of revolution. In other words, we were looking at the surface area of a solid obtained by rotating a function about the x or y axis. In this section we want to look at a much more general setting although you will note that the formula here is very similar to the formula we saw back in Calculus II.

Here we want to find the surface area of the surface given by $z = f(x, y)$ where (x, y) is a point from the region D in the xy-plane. In this case the surface area is given by,

Fact

$$S = \iint\limits_D \sqrt{[f_x]^2 + [f_y]^2 + 1}\, dA$$

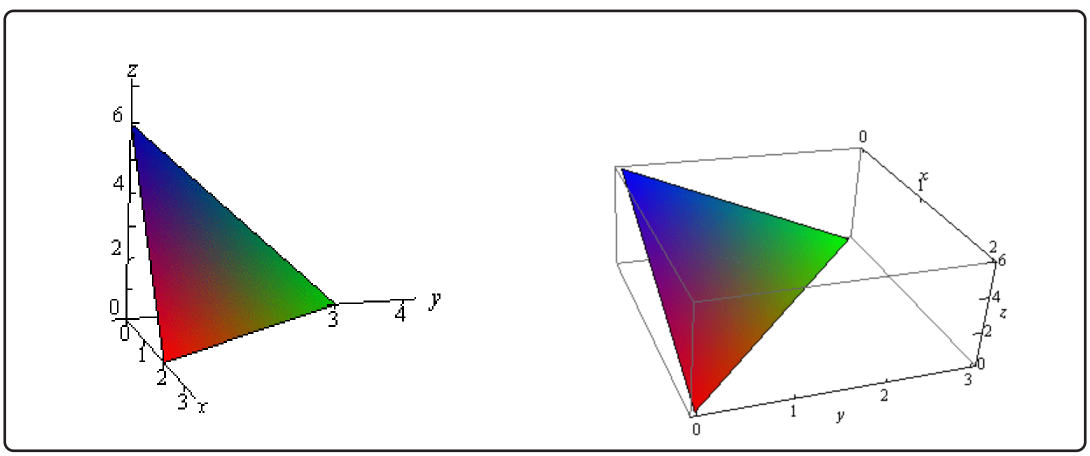
Let's take a look at a couple of examples.

Example 1

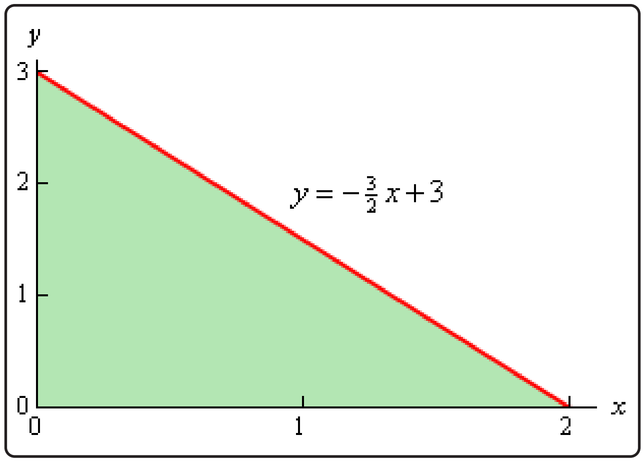
Find the surface area of the part of the plane $3x + 2y + z = 6$ that lies in the first octant.

Solution

Remember that the first octant is the portion of the *xyz*-axis system in which all three variables are positive. Let's first get a sketch of the part of the plane that we are interested in.



We'll also need a sketch of the region D.



Remember that to get the region D we can pretend that we are standing directly over the plane and what we see is the region D. We can get the equation for the hypotenuse of the triangle by realizing that this is nothing more than the line where the plane intersects the xy-plane and we also know that $z = 0$ on the xy-plane. Plugging $z = 0$ into the equation of the plane will give us the equation for the hypotenuse.

Notice that in order to use the surface area formula we need to have the function in the form $z = f(x, y)$ and so solving for z and taking the partial derivatives gives,

$$z = 6 - 3x - 2y \qquad f_x = -3 \qquad f_y = -2$$

The limits defining D are,

$$0 \le x \le 2 \qquad 0 \le y \le -\frac{3}{2}x + 3$$

The surface area is then,

$$\begin{aligned} S &= \iint\limits_D \sqrt{[-3]^2 + [-2]^2 + 1}\, dA \\ &= \int_0^2 \int_0^{-\frac{3}{2}x+3} \sqrt{14}\, dy\, dx \\ &= \sqrt{14} \int_0^2 -\frac{3}{2}x + 3\, dx \\ &= \sqrt{14} \left(-\frac{3}{4}x^2 + 3x \right) \Big|_0^2 \\ &= 3\sqrt{14} \end{aligned}$$

Example 2

Determine the surface area of the part of $z = xy$ that lies in the cylinder given by $x^2 + y^2 = 1$.

Solution

In this case we are looking for the surface area of the part of $z = xy$ where (x, y) comes from the disk of radius 1 centered at the origin since that is the region that will lie inside the given cylinder.

Here are the partial derivatives,

$$f_x = y \qquad\qquad f_y = x$$

The integral for the surface area is,

$$S = \iint_D \sqrt{x^2 + y^2 + 1}\, dA$$

Given that D is a disk it makes sense to do this integral in polar coordinates.

$$\begin{aligned} S &= \iint_D \sqrt{x^2 + y^2 + 1}\, dA \\ &= \int_0^{2\pi} \int_0^1 r\sqrt{1 + r^2}\, dr\, d\theta \\ &= \int_0^{2\pi} \frac{1}{2}\left(\frac{2}{3}\right)\left(1 + r^2\right)^{\frac{3}{2}}\Big|_0^1\, d\theta \\ &= \int_0^{2\pi} \frac{1}{3}\left(2^{\frac{3}{2}} - 1\right)\, d\theta \\ &= \frac{2\pi}{3}\left(2^{\frac{3}{2}} - 1\right) \end{aligned}$$

15.10 Area and Volume Revisited

This section is here only so we can summarize the geometric interpretations of the double and triple integrals that we saw in this chapter. Since the purpose of this section is to summarize these formulas we aren't going to be doing any examples in this section.

We'll first look at the area of a region. The area of the region D is given by,

Area of region

$$\text{Area of } D = \iint\limits_D dA$$

Now let's give the two volume formulas. First the volume of the region E is given by,

Volume of solid

$$\text{Volume of } E = \iiint\limits_E dV$$

Finally, if the region E can be defined as the region under the function $z = f(x, y)$ and above the region D in xy-plane then,

Volume of solid

$$\text{Volume of } E = \iint\limits_D f(x, y)\ dA$$

Note as well that there are similar formulas for the other planes. For instance, the volume of the region behind the function $y = f(x, z)$ and in front of the region D in the xz-plane is given by,

Volume of solid

$$\text{Volume of } E = \iint\limits_D f(x, z)\ dA$$

Likewise, the volume of the region behind the function $x = f(y, z)$ and in front of the region D in the yz-plane is given by,

Volume of solid

$$\text{Volume of } E = \iint\limits_D f(y, z)\ dA$$

16 Line Integrals

We now need to move on to a new kind of integral. When doing single variable definite integrals we integrated a function of one variable over an interval. In the last chapter we integrated a function of two variables over a two dimensional region and we integrated a function of three variables over a three dimensional solid. In this chapter we are going to look at Line Integrals. The difference in this chapter versus the last chapter is where the values of the variables will come from. For a line integral of a function of two variables the variables will all be on the graph of a two dimensional curve C. Similarly, for a line integral of a function of three variables, the variables will all be on the graph of a three dimensional curve C.

The other main difference in this chapter versus previous chapters in which we evaluated integrals is that in addition to evaluating line integrals over functions we will also, for the first time, be integrating a vector field (which we'll also define).

Once we have a grasp on line integrals and how to compute them we'll take a look at the Fundamental Theorem of Calculus for Line Integrals and it's relationship with conservative vector fields. We will, in addition, discuss a method for determining if a two dimensional vector field is conservative or not and if it is conservative how to find the potential function for the vector field.

Finally, we'll discuss a very important theorem, Green's Theorem. Green's theorem gives a very important relationship between certain line integrals and double integrals.

16.1 Vector Fields

We need to start this chapter off with the definition of a vector field as they will be a major component of both this chapter and the next. Let's start off with the formal definition of a vector field.

Definition

A vector field on two (or three) dimensional space is a function $\vec{F}$ that assigns to each point (x, y) (or (x, y, z)) a two (or three dimensional) vector given by $\vec{F}(x, y)$ (or $\vec{F}(x, y, z)$).

That may not make a lot of sense, but most people do know what a vector field is, or at least they've seen a sketch of a vector field. If you've seen a current sketch giving the direction and magnitude of a flow of a fluid or the direction and magnitude of the winds then you've seen a sketch of a vector field.

The standard notation for the function $\vec{F}$ is,

$$\vec{F}(x, y) = P(x, y)\,\vec{i} + Q(x, y)\,\vec{j}$$
$$\vec{F}(x, y, z) = P(x, y, z)\,\vec{i} + Q(x, y, z)\,\vec{j} + R(x, y, z)\,\vec{k}$$

depending on whether or not we're in two or three dimensions. The function P, Q, R (if it is present) are sometimes called **scalar functions**.

Let's take a quick look at a couple of examples.

Example 1

Sketch each of the following vector fields.

(a) $\vec{F}(x, y) = -y\,\vec{i} + x\,\vec{j}$

(b) $\vec{F}(x, y, z) = 2x\,\vec{i} - 2y\,\vec{j} - 2x\,\vec{k}$

Solution

(a) $\vec{F}(x, y) = -y\,\vec{i} + x\,\vec{j}$

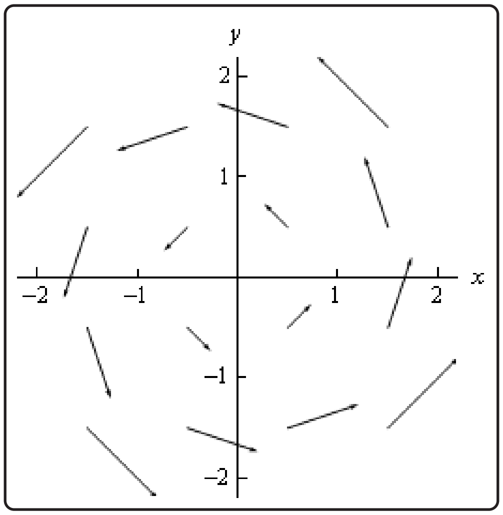
Okay, to graph the vector field we need to get some "values" of the function. This means plugging in some points into the function. Here are a couple of evaluations.

$$\vec{F}\left(\frac{1}{2}, \frac{1}{2}\right) = -\frac{1}{2}\vec{i} + \frac{1}{2}\vec{j} \qquad \vec{F}\left(\frac{3}{2}, \frac{1}{4}\right) = -\frac{1}{4}\vec{i} + \frac{3}{2}\vec{j}$$

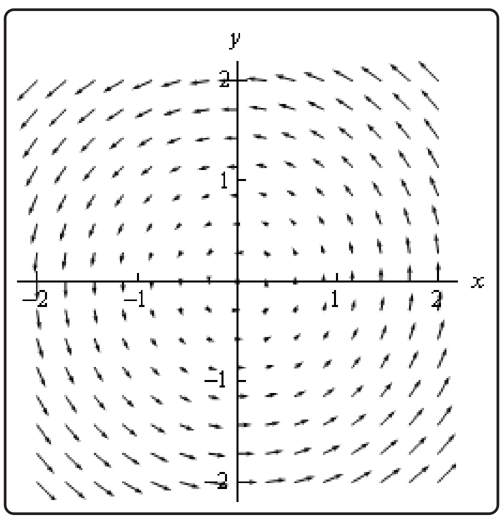
$$\vec{F}\left(\frac{1}{2}, -\frac{1}{2}\right) = -\left(-\frac{1}{2}\right)\vec{i} + \frac{1}{2}\vec{j} = \frac{1}{2}\vec{i} + \frac{1}{2}\vec{j}$$

So, just what do these evaluations tell us? Well the first one tells us that at the point $\left(\frac{1}{2}, \frac{1}{2}\right)$ we will plot the vector $-\frac{1}{2}\vec{i} + \frac{1}{2}\vec{j}$. Likewise, the third evaluation tells us that at the point $\left(\frac{3}{2}, \frac{1}{4}\right)$ we will plot the vector $-\frac{1}{4}\vec{i} + \frac{3}{2}\vec{j}$.

We can continue in this fashion plotting vectors for several points and we'll get the following sketch of the vector field.



If we want significantly more points plotted, then it is usually best to use a computer aided graphing system such as Maple or Mathematica. Here is a sketch with many more vectors included that was generated with Mathematica.



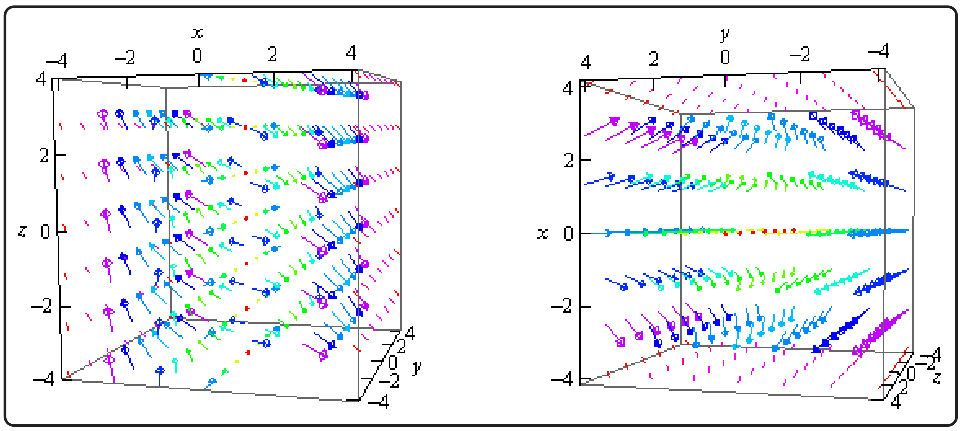
(b) $\vec{F}(x, y, z) = 2x\,\vec{i} - 2y\,\vec{j} - 2x\,\vec{k}$

In the case of three dimensional vector fields it is almost always better to use Maple, Mathematica, or some other such tool. Despite that let's go ahead and do a couple of evaluations anyway.

$$\vec{F}(1,-3,2) = 2\,\vec{i} + 6\,\vec{j} - 2\,\vec{k}$$
$$\vec{F}(0,5,3) = -10\,\vec{j}$$

Notice that z only affects the placement of the vector in this case and does not affect the direction or the magnitude of the vector. Sometimes this will happen so don't get excited about it when it does.

Here is a couple of sketches generated by Mathematica. The sketch on the left is from the "front" and the sketch on the right is from "above".



Now that we've seen a couple of vector fields let's notice that we've already seen a vector field function. In the second chapter we looked at the gradient vector. Recall that given a function $f(x,y,z)$ the gradient vector is defined by,

Fact

$$\nabla f = \langle f_x, f_y, f_z \rangle$$

This is a vector field and is often called a **gradient vector field**.

In these cases, the function $f(x,y,z)$ is often called a scalar function to differentiate it from the vector field.

Example 2

Find the gradient vector field of the following functions.

(a) $f(x,y) = x^2 \sin(5y)$

(b) $f(x,y,z) = z\mathbf{e}^{-xy}$

Solution

(a) $f(x,y) = x^2 \sin(5y)$

Note that we only gave the gradient vector definition for a three dimensional function, but don't forget that there is also a two dimension definition. All that we need to drop off the third component of the vector.

Here is the gradient vector field for this function.

$$\nabla f = \langle 2x\sin(5y), 5x^2\cos(5y)\rangle$$

(b) $f(x,y,z) = z\mathbf{e}^{-xy}$

There isn't much to do here other than take the gradient.

$$\nabla f = \left\langle -yz\mathbf{e}^{-xy}, -xz\mathbf{e}^{-xy}, \mathbf{e}^{-xy}\right\rangle$$

Let's do another example that will illustrate the relationship between the gradient vector field of a function and its contours.

Example 3

Sketch the gradient vector field for $f(x,y) = x^2 + y^2$ as well as several contours for this function.

Solution

Recall that the contours for a function are nothing more than curves defined by,

$$f(x,y) = k$$

for various values of k. So, for our function the contours are defined by the equation,

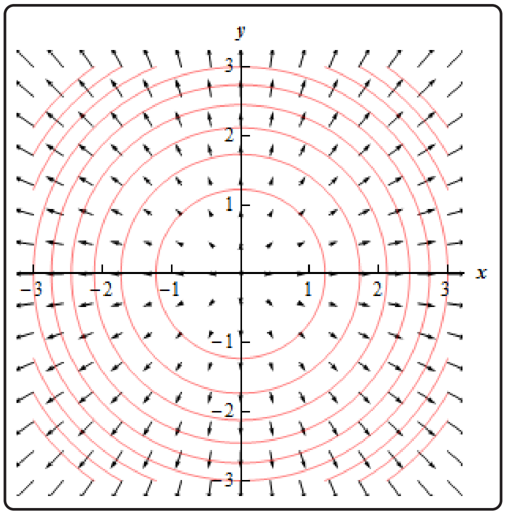
$$x^2 + y^2 = k$$

and so they are circles centered at the origin with radius $\sqrt{k}$.

Here is the gradient vector field for this function.

$$\nabla f(x, y) = 2x\,\vec{i} + 2y\,\vec{j}$$

Here is a sketch of several of the contours as well as the gradient vector field.



Notice that the vectors of the vector field are all orthogonal (or perpendicular) to the contours. This will always be the case when we are dealing with the contours of a function as well as its gradient vector field.

The k's we used for the graph above were 1.5, 3, 4.5, 6, 7.5, 9, 10.5, 12, and 13.5. Now notice that as we increased k by 1.5 the contour curves get closer together and that as the contour curves get closer together the larger the vectors become. In other words, the closer the contour curves are (as k is increased by a fixed amount) the faster the function is changing at that point. Also recall that the direction of fastest change for a function is given by the gradient vector at that point. Therefore, it should make sense that the two ideas should match up as they do here.

The final topic of this section is that of conservative vector fields. A vector field $\vec{F}$ is called a **conservative vector field** if there exists a function f such that $\vec{F} = \nabla f$. If $\vec{F}$ is a conservative vector field then the function, f, is called a **potential function** for $\vec{F}$.

All this definition is saying is that a vector field is conservative if it is also a gradient vector field for some function.

For instance the vector field $\vec{F} = y\,\vec{i} + x\,\vec{j}$ is a conservative vector field with a potential function of $f(x, y) = xy$ because $\nabla f = \langle y, x \rangle$.

On the other hand, $\vec{F} = -y\,\vec{i} + x\,\vec{j}$ is not a conservative vector field since there is no function f such that $\vec{F} = \nabla f$. If you're not sure that you believe this at this point be patient, we will be able to prove this in a couple of sections. In that section we will also show how to find the potential function for a conservative vector field.

16.2 Line Integrals - Part I

In this section we are now going to introduce a new kind of integral. However, before we do that it is important to note that you will need to remember how to parameterize equations, or put another way, you will need to be able to write down a set of parametric equations for a given curve. You should have seen some of this in your Calculus II course. If you need some review you should go back and review some of the basics of parametric equations and curves.

Here are some of the more basic curves that we'll need to know how to do as well as limits on the parameter if they are required.

Curve	Parametric Equations
$\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=1$ (Ellipse)	Counter-Clockwise: $x=a\cos(t)$, $y=b\sin(t)$, $0\le t\le 2\pi$ Clockwise: $x=a\cos(t)$, $y=-b\sin(t)$, $0\le t\le 2\pi$
$x^2+y^2=r^2$ (Circle)	Counter-Clockwise: $x=r\cos(t)$, $y=r\sin(t)$, $0\le t\le 2\pi$ Clockwise: $x=r\cos(t)$, $y=-r\sin(t)$, $0\le t\le 2\pi$
$y=f(x)$	$x=t$ $y=f(t)$
$x=g(y)$	$x=g(t)$ $y=t$
Line Segment From (x_0,y_0,z_0) to (x_1,y_1,z_1)	$\vec r(t)=(1-t)\langle x_0,y_0,z_0\rangle+t\langle x_1,y_1,z_1\rangle\ ,\ 0\le t\le 1$ or $x=(1-t)x_0+tx_1$ $y=(1-t)y_0+ty_1\ ,\ 0\le t\le 1$ $z=(1-t)z_0+tz_1$

With the final one we gave both the vector form of the equation as well as the parametric form and if we need the two-dimensional version then we just drop the z components. In fact, we will be using the two-dimensional version of this in this section.

For the ellipse and the circle we've given two parameterizations, one tracing out the curve clockwise and the other counter-clockwise. As we'll eventually see the direction that the curve is traced out can, on occasion, change the answer. Also, both of these "start" on the positive x-axis at $t = 0$.

Now let's move on to line integrals. In Calculus I we integrated $f(x)$, a function of a single variable, over an interval $[a, b]$. In this case we were thinking of x as taking all the values in this interval starting at a and ending at b. With line integrals we will start with integrating the function $f(x, y)$, a function of two variables, and the values of x and y that we're going to use will be the points, (x, y), that lie on a curve C. Note that this is different from the double integrals that we were working with in the previous chapter where the points came out of some two-dimensional region.

Let's start with the curve C that the points come from. We will assume that the curve is *smooth* (defined shortly) and is given by the parametric equations,

$$x = h(t) \qquad y = g(t) \qquad a \le t \le b$$

We will often want to write the parameterization of the curve as a vector function. In this case the curve is given by,

$$\vec{r}(t) = h(t)\,\vec{i} + g(t)\,\vec{j} \qquad a \le t \le b$$

The curve is called **smooth** if $\vec{r}'(t)$ is continuous and $\vec{r}'(t) \ne 0$ for all t.

Defintion

The **line integral** of $f(x, y)$ along C is denoted by,

$$\int_C f(x, y)\, ds$$

We use a ds here to acknowledge the fact that we are moving along the curve, C, instead of the x-axis (denoted by dx) or the y-axis (denoted by dy). Because of the ds this is sometimes called the **line integral of f with respect to arc length**.

We've seen the notation ds before. If you recall from Calculus II when we looked at the arc length of a curve given by parametric equations we found it to be,

$$L = \int_a^b ds\,, \qquad \text{where } ds = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt$$

It is no coincidence that we use ds for both of these problems. The ds is the same for both the arc length integral and the notation for the line integral.

So, to compute a line integral we will convert everything over to the parametric equations. The line integral is then,

Fact

$$\int\limits_C f(x,y)\,ds = \int_a^b f(h(t),g(t))\sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2}\,dt$$

Don't forget to plug the parametric equations into the function as well.

If we use the vector form of the parameterization we can simplify the notation up somewhat by noticing that,

$$\sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2} = \left\|\vec{r}\,'(t)\right\|$$

where $\left\|\vec{r}\,'(t)\right\|$ is the magnitude or norm of $\vec{r}\,'(t)$. Using this notation, the line integral becomes,

Fact

$$\int\limits_C f(x,y)\,ds = \int_a^b f(h(t),g(t))\,\left\|\vec{r}\,'(t)\right\|\,dt$$

Note that as long as the parameterization of the curve C is traced out exactly once as t increases from a to b the value of the line integral will be independent of the parameterization of the curve.

Let's take a look at an example of a line integral.

Example 1

Evaluate $\int\limits_C xy^4\,ds$ where C is the right half of the circle, $x^2+y^2=16$ traced out in a counter clockwise direction.

Solution

We first need a parameterization of the circle. This is given by,

$$x = 4\cos(t) \qquad y = 4\sin(t)$$

We now need a range of t's that will give the right half of the circle. The following range of t's will do this.

$$-\frac{\pi}{2} \le t \le \frac{\pi}{2}$$

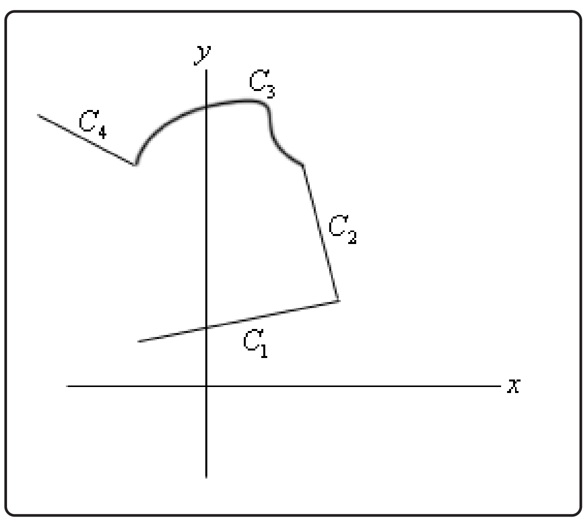
Now, we need the derivatives of the parametric equations and let's compute ds.

$$\frac{dx}{dt} = -4\sin(t) \qquad \frac{dy}{dt} = 4\cos(t)$$
$$ds = \sqrt{16\sin^2(t) + 16\cos^2(t)}\,dt = 4\,dt$$

The line integral is then,

$$\begin{aligned}\int_C xy^4\,ds &= \int_{-\pi/2}^{\pi/2} 4\cos(t)(4\sin(t))^4\,(4)\,dt \\ &= 4096\int_{-\pi/2}^{\pi/2} \cos(t)\ \sin^4(t)\,dt \\ &= \frac{4096}{5}\sin^5(t)\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \\ &= \frac{8192}{5}\end{aligned}$$

Next we need to talk about line integrals over **piecewise smooth curves**. A piecewise smooth curve is any curve that can be written as the union of a finite number of smooth curves, C_1,..., C_n where the end point of C_i is the starting point of C_{i+1}. Below is an illustration of a piecewise smooth curve.



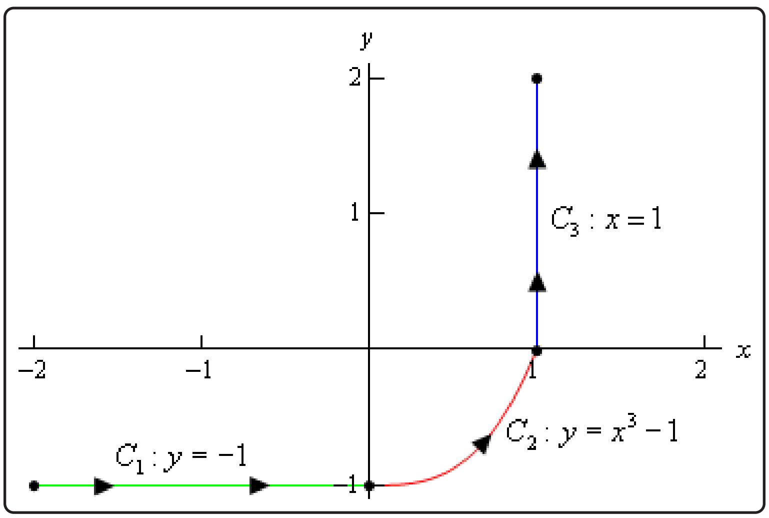
Evaluation of line integrals over piecewise smooth curves is a relatively simple thing to do. All we do is evaluate the line integral over each of the pieces and then add them up. The line integral for some function over the above piecewise curve would be,

$$\int_C f(x,y)\,ds = \int_{C_1} f(x,y)\,ds + \int_{C_2} f(x,y)\,ds + \int_{C_3} f(x,y)\,ds + \int_{C_4} f(x,y)\,ds$$

Let's see an example of this.

Example 2

Evaluate $\int_C 4x^3\,ds$ where C is the curve shown below.



Solution

So, first we need to parameterize each of the curves.

$$\begin{aligned} C_1 &: \quad x = t,\ y = -1, & -2 \le t \le 0 \\ C_2 &: \quad x = t,\ y = t^3 - 1, & 0 \le t \le 1 \\ C_3 &: \quad x = 1,\ y = t, & 0 \le t \le 2 \end{aligned}$$

Now let's do the line integral over each of these curves.

$$\int_{C_1} 4x^3\,ds = \int_{-2}^{0} 4t^3\sqrt{(1)^2 + (0)^2}\,dt = \int_{-2}^{0} 4t^3\,dt = t^4\Big|_{-2}^{0} = -16$$

$$\int_{C_2} 4x^3\,ds = \int_0^1 4t^3\sqrt{(1)^2 + (3t^2)^2}\,dt$$
$$= \int_0^1 4t^3\sqrt{1+9t^4}\,dt$$
$$= \frac{1}{9}\left(\frac{2}{3}\right)\left(1+9t^4\right)^{\frac{3}{2}}\Big|_0^1 = \frac{2}{27}\left(10^{\frac{3}{2}} - 1\right) = 2.268$$

$$\int_{C_3} 4x^3\,ds = \int_0^2 4(1)^3\sqrt{(0)^2 + (1)^2}\,dt = \int_0^2 4\,dt = 8$$

Finally, the line integral that we were asked to compute is,

$$\int_C 4x^3\,ds = \int_{C_1} 4x^3\,ds + \int_{C_2} 4x^3\,ds + \int_{C_3} 4x^3\,ds$$
$$= -16 + 2.268 + 8$$
$$= -5.732$$

Notice that we put direction arrows on the curve in the above example. The direction of motion along a curve *may* change the value of the line integral as we will see in the next section. Also note that the curve can be thought of a curve that takes us from the point $(-2,-1)$ to the point $(1,2)$. Let's first see what happens to the line integral if we change the path between these two points.

Example 3

Evaluate $\int_C 4x^3\,ds$ where C is the line segment from $(-2,-1)$ to $(1,2)$.

Solution

From the parameterization formulas at the start of this section we know that the line segment starting at $(-2,-1)$ and ending at $(1,2)$ is given by,

$$\vec{r}(t) = (1-t)\langle -2,-1\rangle + t\langle 1,2\rangle$$
$$= \langle -2+3t, -1+3t\rangle$$

for $0 \le t \le 1$. This means that the individual parametric equations are,

$$x = -2+3t \qquad y = -1+3t$$

Using this path the line integral is,

$$\begin{aligned}\int_C 4x^3\,ds &= \int_0^1 4(-2+3t)^3\sqrt{9+9}\,dt\\ &= 12\sqrt{2}\left(\frac{1}{12}\right)(-2+3t)^4\Big|_0^1\\ &= 12\sqrt{2}\left(-\frac{5}{4}\right)\\ &= -15\sqrt{2} = -21.213\end{aligned}$$

When doing these integrals don't forget simple Calc I substitutions to avoid having to do things like cubing out a term. Cubing it out is not that difficult, but it is more work than a simple substitution.

So, the previous two examples seem to suggest that if we change the path between two points then the value of the line integral (with respect to arc length) will change. While this will happen fairly regularly we can't assume that it will always happen. In a later section we will investigate this idea in more detail.

Next, let's see what happens if we change the direction of a path.

Example 4

Evaluate $\displaystyle\int_C 4x^3\,ds$ where C is the line segment from $(1,2)$ to $(-2,-1)$.

Solution

This one isn't much different, work wise, from the previous example. Here is the parameterization of the curve.

$$\begin{aligned}\vec{r}(t) &= (1-t)\langle 1,2\rangle + t\langle -2,-1\rangle\\ &= \langle 1-3t, 2-3t\rangle\end{aligned}$$

for $0 \le t \le 1$. Remember that we are switching the direction of the curve and this will also change the parameterization so we can make sure that we start/end at the proper point.

Here is the line integral.

$$\begin{aligned}\int_C 4x^3\,ds &= \int_0^1 4(1-3t)^3\sqrt{9+9}\,dt\\ &= 12\sqrt{2}\left(-\frac{1}{12}\right)(1-3t)^4\Big|_0^1\\ &= 12\sqrt{2}\left(-\frac{5}{4}\right)\\ &= -15\sqrt{2} = -21.213\end{aligned}$$

So, it looks like when we switch the direction of the curve the line integral (with respect to arc length) will not change. This will always be true for these kinds of line integrals. However, there are other kinds of line integrals in which this won't be the case. We will see more examples of this in the next couple of sections so don't get it into your head that changing the direction will never change the value of the line integral.

Before working another example let's formalize this idea up somewhat. Let's suppose that the curve C has the parameterization $x = h(t)$, $y = g(t)$. Let's also suppose that the initial point on the curve is A and the final point on the curve is B. The parameterization $x = h(t)$, $y = g(t)$ will then determine an **orientation** for the curve where the positive direction is the direction that is traced out as t increases. Finally, let $-C$ be the curve with the same points as C, however in this case the curve has B as the initial point and A as the final point, again t is increasing as we traverse this curve. In other words, given a curve C, the curve $-C$ is the same curve as C except the direction has been reversed.

We then have the following fact about line integrals with respect to arc length.

Fact

$$\int_C f(x,y)\,ds = \int_{-C} f(x,y)\,ds$$

So, for a line integral with respect to arc length we can change the direction of the curve and not change the value of the integral. This is a useful fact to remember as some line integrals will be easier in one direction than the other.

Now, let's work another example

Example 5

Evaluate $\int\limits_C x\,ds$ for each of the following curves.

(a) $C_1 : y = x^2, \ \ -1 \le x \le 1$

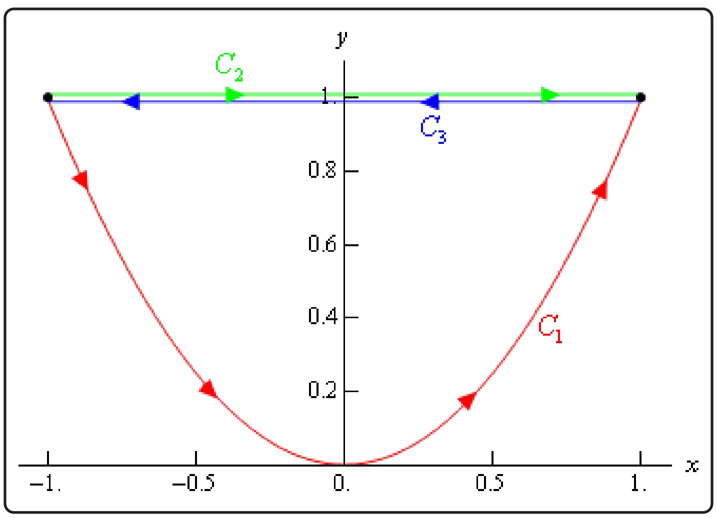
(b) C_2: The line segment from $(-1,1)$ to $(1,1)$.

(c) C_3: The line segment from $(1,1)$ to $(-1,1)$.

Solution

Before working any of these line integrals let's notice that all of these curves are paths that connect the points $(-1,1)$ and $(1,1)$. Also notice that $C_3 = -C_2$ and so by the fact above these two should give the same answer.

Here is a sketch of the three curves and note that the curves illustrating C_2 and C_3 have been separated a little to show that they are separate curves in some way even though they are the same line.



(a) $C_1 : y = x^2, \ \ -1 \le x \le 1$

Here is a parameterization for this curve.

$$C_1 : x = t, \ y = t^2, \ \ -1 \le t \le 1$$

Here is the line integral.

$$\int\limits_{C_1} x\,ds = \int_{-1}^{1} t\sqrt{1+4t^2}\,dt = \frac{1}{12}\left(1+4t^2\right)^{\frac{3}{2}}\Big|_{-1}^{1} = 0$$

(b) C_2: The line segment from $(-1, 1)$ to $(1, 1)$.

There are two parameterizations that we could use here for this curve. The first is to use the formula we used in the previous couple of examples. That parameterization is,

$$\begin{aligned} C_2 : \vec{r}(t) &= (1-t)\langle -1, 1\rangle + t\langle 1, 1\rangle \\ &= \langle 2t-1, 1\rangle \end{aligned}$$

for $0 \le t \le 1$.

Sometimes we have no choice but to use this parameterization. However, in this case there is a second (probably) easier parameterization. The second one uses the fact that we are really just graphing a portion of the line $y = 1$. Using this the parameterization is,

$$C_2 : x = t, \ y = 1, \ -1 \le t \le 1$$

This will be a much easier parameterization to use so we will use this. Here is the line integral for this curve.

$$\int_{C_2} x\,ds = \int_{-1}^{1} t\sqrt{1+0}\,dt = \frac{1}{2}t^2\Big|_{-1}^{1} = 0$$

Note that this time, unlike the line integral we worked with in Examples 2, 3, and 4 we got the same value for the integral despite the fact that the path is different. This will happen on occasion. We should also not expect this integral to be the same for all paths between these two points. At this point all we know is that for these two paths the line integral will have the same value. It is completely possible that there is another path between these two points that will give a different value for the line integral.

(c) C_3: The line segment from $(1, 1)$ to $(-1, 1)$.

Now, according to our fact above we really don't need to do anything here since we know that $C_3 = -C_2$. The fact tells us that this line integral should be the same as the second part (*i.e.* zero). However, let's verify that, plus there is a point we need to make here about the parameterization.

Here is the parameterization for this curve.

$$\begin{aligned} C_3 : \vec{r}(t) &= (1-t)\langle 1, 1\rangle + t\langle -1, 1\rangle \\ &= \langle 1-2t, 1\rangle \end{aligned}$$

for $0 \le t \le 1$.

Note that this time we can't use the second parameterization that we used in part (b) since we need to move from right to left as the parameter increases and the second parameterization used in the previous part will move in the opposite direction.

Here is the line integral for this curve.

$$\int\limits_{C_3} x\,ds = \int_0^1 (1-2t)\sqrt{4+0}\,dt = 2\left(t-t^2\right)\Big|_0^1 = 0$$

Sure enough we got the same answer as the second part.

To this point in this section we've only looked at line integrals over a two-dimensional curve. However, there is no reason to restrict ourselves like that. We can do line integrals over three-dimensional curves as well.

Let's suppose that the three-dimensional curve C is given by the parameterization,

$$x = x(t), \qquad y = y(t) \qquad z = z(t) \qquad a \le t \le b$$

then the line integral is given by,

Fact

$$\int\limits_C f(x,y,z)\,ds = \int_a^b f(x(t),y(t),z(t))\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\,dt$$

Note that often when dealing with three-dimensional space the parameterization will be given as a vector function.

$$\vec{r}(t) = \langle x(t), y(t), z(t)\rangle$$

Notice that we changed up the notation for the parameterization a little. Since we rarely use the function names we simply kept the x, y, and z and added on the (t) part to denote that they may be functions of the parameter.

Also notice that, as with two-dimensional curves, we have,

$$\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} = \left\|\vec{r}\,'(t)\right\|$$

and the line integral can again be written as,

Fact

$$\int\limits_C f(x,y,z)\,ds = \int_a^b f(x(t),y(t),z(t))\,\left\|\vec{r}\,'(t)\right\|\,dt$$

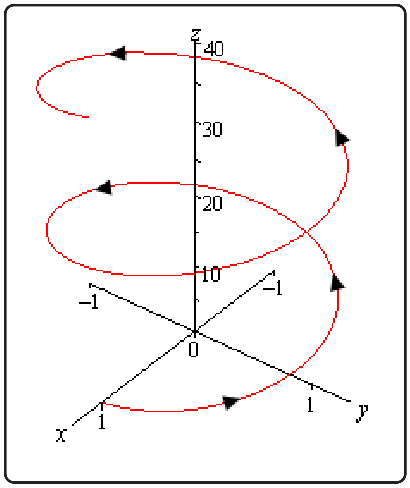
So, outside of the addition of a third parametric equation line integrals in three-dimensional space work the same as those in two-dimensional space. Let's work a quick example.

Example 6

Evaluate $\int\limits_C xyz\,ds$ where C is the helix given by, $\vec{r}(t) = \langle \cos(t), \sin(t), 3t\rangle$, $0 \le t \le 4\pi$.

Solution

Note that we first saw the vector equation for a helix back in the Vector Functions section. Here is a quick sketch of the helix.



Here is the line integral.

$$\begin{aligned}\int\limits_C xyz\,ds &= \int_0^{4\pi} 3t\cos(t)\sin(t)\sqrt{\sin^2(t)+\cos^2(t)+9}\,dt \\ &= \int_0^{4\pi} 3t\left(\frac{1}{2}\sin(2t)\right)\sqrt{1+9}\,dt \\ &= \frac{3\sqrt{10}}{2}\int_0^{4\pi} t\sin(2t)\ dt \\ &= \frac{3\sqrt{10}}{2}\left(\frac{1}{4}\sin(2t)-\frac{t}{2}\cos(2t)\right)\Bigg|_0^{4\pi} \\ &= -3\sqrt{10}\,\pi\end{aligned}$$

You were able to do that integral right? It required integration by parts.

So, as we can see there really isn't too much difference between two- and three-dimensional line integrals.

16.3 Line Integrals - Part II

In the previous section we looked at line integrals with respect to arc length. In this section we want to look at line integrals with respect to x and/or y.

As with the last section we will start with a two-dimensional curve C with parameterization,

$$x = x(t) \qquad y = y(t) \qquad a \le t \le b$$

Fact

The **line integral of f with respect to x** is,

$$\int\limits_C f(x, y)\, dx = \int_a^b f\big(x(t), y(t)\big)x'(t)\, dt$$

The **line integral of f with respect to y** is,

$$\int\limits_C f(x, y)\, dy = \int_a^b f\big(x(t), y(t)\big)y'(t)\, dt$$

Note that the only notational difference between these two and the line integral with respect to arc length (from the previous section) is the differential. These have a dx or dy while the line integral with respect to arc length has a ds. So, when evaluating line integrals be careful to first note which differential you've got so you don't work the wrong kind of line integral.

These two integral often appear together and so we have the following shorthand notation for these cases.

Fact

$$\int\limits_C P dx + Q\, dy = \int\limits_C P(x, y)\, dx + \int\limits_C Q(x, y)\, dy$$

Let's take a quick look at an example of this kind of line integral.

Example 1

Evaluate $\displaystyle\int_C \sin(\pi y)\,dy + yx^2\,dx$ where C is the line segment from $(0,2)$ to $(1,4)$.

Solution

Here is the parameterization of the curve.

$$\vec{r}(t) = (1-t)\langle 0,2\rangle + t\langle 1,4\rangle = \langle t, 2+2t\rangle \qquad 0 \le t \le 1$$

The line integral is,

$$\begin{aligned}
\int_C \sin(\pi y)\,dy + yx^2\,dx &= \int_C \sin(\pi y)\,dy + \int_C yx^2\,dx \\
&= \int_0^1 \sin(\pi(2+2t))(2)\,dt + \int_0^1 (2+2t)(t)^2(1)\,dt \\
&= -\frac{1}{\pi}\cos(2\pi + 2\pi t)\Big|_0^1 + \left(\frac{2}{3}t^3 + \frac{1}{2}t^4\right)\Big|_0^1 \\
&= \frac{7}{6}
\end{aligned}$$

In the previous section we saw that changing the direction of the curve for a line integral with respect to arc length doesn't change the value of the integral. Let's see what happens with line integrals with respect to x and/or y.

Example 2

Evaluate $\displaystyle\int_C \sin(\pi y)\,dy + yx^2\,dx$ where C is the line segment from $(1,4)$ to $(0,2)$.

Solution

So, we simply changed the direction of the curve. Here is the new parameterization.

$$\vec{r}(t) = (1-t)\langle 1,4\rangle + t\langle 0,2\rangle = \langle 1-t, 4-2t\rangle \qquad 0 \le t \le 1$$

The line integral in this case is,

$$\begin{aligned}\int_C \sin(\pi y)\,dy + yx^2\,dx &= \int_C \sin(\pi y)\,dy + \int_C yx^2\,dx\\ &= \int_0^1 \sin(\pi(4-2t))(-2)\,dt + \int_0^1 (4-2t)(1-t)^2(-1)\,dt\\ &= -\frac{1}{\pi}\cos(4\pi - 2\pi t)\Big|_0^1 - \left(-\frac{1}{2}t^4 + \frac{8}{3}t^3 - 5t^2 + 4t\right)\Big|_0^1\\ &= -\frac{7}{6}\end{aligned}$$

So, switching the direction of the curve got us a different value or at least the opposite sign of the value from the first example. In fact this will always happen with these kinds of line integrals.

Fact

If C is any curve then,

$$\int_{-C} f(x,y)\,dx = -\int_C f(x,y)\,dx \qquad \text{and} \qquad \int_{-C} f(x,y)\,dy = -\int_C f(x,y)\,dy$$

With the combined form of these two integrals we get,

$$\int_{-C} P\,dx + Q\,dy = -\int_C P\,dx + Q\,dy$$

We can also do these integrals over three-dimensional curves as well. In this case we will pick up a third integral (with respect to z) and the three integrals will be.

Fact

$$\begin{aligned}\int_C f(x,y,z)\,dx &= \int_a^b f\big(x(t),y(t),z(t)\big)x'(t)\,dt\\ \int_C f(x,y,z)\,dy &= \int_a^b f\big(x(t),y(t),z(t)\big)y'(t)\,dt\\ \int_C f(x,y,z)\,dz &= \int_a^b f\big(x(t),y(t),z(t)\big)z'(t)\,dt\end{aligned}$$

where the curve C is parameterized by

$$x = x(t) \qquad y = y(t) \qquad z = z(t) \qquad a \le t \le b$$

As with the two-dimensional version these three will often occur together so the shorthand we'll be using here is,

Fact

$$\int_C P\,dx + Q\,dy + R\,dz = \int_C P\left(x,y,z\right)dx + \int_C Q\left(x,y,z\right)\,dy + \int_C R\left(x,y,z\right)dz$$

Let's work an example.

Example 3

Evaluate $\displaystyle\int_C y\,dx + x\,dy + z\,dz$ where C is given by $x = \cos(t)$, $y = \sin(t)$, $z = t^2$, $0 \le t \le 2\pi$.

Solution

So, we already have the curve parameterized so there really isn't much to do other than evaluate the integral.

$$\begin{aligned}\int_C y\,dx + x\,dy + z\,dz &= \int_C y\,dx + \int_C x\,dy + \int_C z\,dz \\ &= \int_0^{2\pi} \sin(t)\big(-\sin(t)\big)\,dt + \int_0^{2\pi} \cos(t)\big(\cos(t)\big)\,dt + \int_0^{2\pi} t^2\left(2t\right)\,dt \\ &= -\int_0^{2\pi} \sin^2(t)\,dt + \int_0^{2\pi} \cos^2(t)\,dt + \int_0^{2\pi} 2t^3\,dt \\ &= -\frac{1}{2}\int_0^{2\pi} 1 - \cos\left(2t\right)dt + \frac{1}{2}\int_0^{2\pi} 1 + \cos\left(2t\right)dt + \int_0^{2\pi} 2t^3\,dt \\ &= \left(-\frac{1}{2}\left(t - \frac{1}{2}\sin\left(2t\right)\right) + \frac{1}{2}\left(t + \frac{1}{2}\sin\left(2t\right)\right) + \frac{1}{2}t^4\right)\Bigg|_0^{2\pi} \\ &= 8\pi^4\end{aligned}$$

16.4 Line Integrals of Vector Fields

In the previous two sections we looked at line integrals of functions. In this section we are going to evaluate line integrals of vector fields. We'll start with the vector field,

$$\vec{F}(x,y,z) = P(x,y,z)\,\vec{i} + Q(x,y,z)\,\vec{j} + R(x,y,z)\,\vec{k}$$

and the three-dimensional, smooth curve given by

$$\vec{r}(t) = x(t)\,\vec{i} + y(t)\,\vec{j} + z(t)\,\vec{k} \qquad a \le t \le b$$

Fact

The **line integral of $\vec{F}$ along** C is

$$\int\limits_C \vec{F}\cdot d\,\vec{r} = \int_a^b \vec{F}\big(\vec{r}(t)\big)\cdot\vec{r}\,'(t)\,dt$$

Note the notation in the integral on the left side. That really is a dot product of the vector field and the differential really is a vector. Also, $\vec{F}(\vec{r}(t))$ is a shorthand for,

$$\vec{F}(\vec{r}(t)) = \vec{F}\big(x(t), y(t), z(t)\big)$$

We can also write line integrals of vector fields as a line integral with respect to arc length as follows,

$$\int\limits_C \vec{F}\cdot d\,\vec{r} = \int\limits_C \vec{F}\cdot\vec{T}\,ds$$

where $\vec{T}(t)$ is the unit tangent vector and is given by,

$$\vec{T}(t) = \frac{\vec{r}\,'(t)}{\|\vec{r}\,'(t)\|}$$

If we use our knowledge on how to compute line integrals with respect to arc length we can see that this second form is equivalent to the first form given above.

$$\begin{aligned}\int\limits_C \vec{F}\cdot d\,\vec{r} &= \int\limits_C \vec{F}\cdot\vec{T}\,ds\\ &= \int_a^b \vec{F}\big(\vec{r}(t)\big)\cdot\frac{\vec{r}\,'(t)}{\|\vec{r}\,'(t)\|}\,\|\vec{r}\,'(t)\|\,dt\\ &= \int_a^b \vec{F}\big(\vec{r}(t)\big)\cdot\vec{r}\,'(t)\,dt\end{aligned}$$

In general, we use the first form to compute these line integral as it is usually much easier to use. Let's take a look at a couple of examples.

Example 1

Evaluate $\int\limits_C \vec{F} \cdot d\,\vec{r}$ where $\vec{F}(x,y,z) = 8x^2 y\,z\,\vec{i} + 5z\,\vec{j} - 4x\,y\,\vec{k}$ and C is the curve given by $\vec{r}(t) = t\,\vec{i} + t^2\,\vec{j} + t^3\,\vec{k}$, $0 \le t \le 1$.

Solution

Okay, we first need the vector field evaluated along the curve.

$$\vec{F}(\vec{r}(t)) = 8t^2\left(t^2\right)\left(t^3\right)\vec{i} + 5t^3\,\vec{j} - 4t\left(t^2\right)\vec{k} = 8t^7\,\vec{i} + 5t^3\,\vec{j} - 4t^3\,\vec{k}$$

Next, we need the derivative of the parameterization.

$$\vec{r}\,'(t) = \vec{i} + 2t\,\vec{j} + 3t^2\,\vec{k}$$

Finally, let's get the dot product taken care of.

$$\vec{F}(\vec{r}(t)) \cdot \vec{r}\,'(t) = 8t^7 + 10t^4 - 12t^5$$

The line integral is then,

$$\begin{aligned}\int\limits_C \vec{F} \cdot d\,\vec{r} &= \int_0^1 8t^7 + 10t^4 - 12t^5\,dt \\ &= \left(t^8 + 2t^5 - 2t^6\right)\Big|_0^1 \\ &= 1\end{aligned}$$

Example 2

Evaluate $\int\limits_C \vec{F} \cdot d\,\vec{r}$ where $\vec{F}(x,y,z) = x\,z\,\vec{i} - y\,z\,\vec{k}$ and C is the line segment from $(-1,2,0)$ to $(3,0,1)$.

Solution

We'll first need the parameterization of the line segment. We saw how to get the parameterization of line segments in the first section on line integrals. We've been using the two dimensional version of this over the last couple of sections. Here is the parameterization for the line.

$$\vec{r}(t) = (1-t)\langle -1, 2, 0\rangle + t\langle 3, 0, 1\rangle$$
$$= \langle 4t-1, 2-2t, t\rangle, \qquad 0 \le t \le 1$$

So, let's get the vector field evaluated along the curve.

$$\vec{F}\left(\vec{r}(t)\right) = (4t-1)(t)\,\vec{i} - (2-2t)(t)\,\vec{k}$$
$$= \left(4t^2 - t\right)\vec{i} - \left(2t - 2t^2\right)\vec{k}$$

Now we need the derivative of the parameterization.

$$\vec{r}\,'(t) = \langle 4, -2, 1\rangle$$

The dot product is then,

$$\vec{F}(\vec{r}(t)) \cdot \vec{r}\,'(t) = 4\left(4t^2 - t\right) - \left(2t - 2t^2\right) = 18t^2 - 6t$$

The line integral becomes,

$$\int\limits_C \vec{F} \cdot d\,\vec{r} = \int_0^1 18t^2 - 6t\,dt = \left(6t^3 - 3t^2\right)\Big|_0^1 = 3$$

Let's close this section out by doing one of these in general to get a nice relationship between line integrals of vector fields and line integrals with respect to x, y, and z.

Given the vector field $\vec{F}(x,y,z) = P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}$ and the curve C parameterized by $\vec{r}(t) = x(t)\,\vec{i} + y(t)\,\vec{j} + z(t)\,\vec{k}$, $a \le t \le b$ the line integral is,

$$\begin{aligned}
\int\limits_C \vec{F} \cdot d\,\vec{r} &= \int_a^b \left(P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}\right) \cdot \left(x'\,\vec{i} + y'\,\vec{j} + z'\,\vec{k}\right)\,dt \\
&= \int_a^b Px' + Qy' + Rz'\,dt \\
&= \int_a^b Px'\,dt + \int_a^b Qy'\,dt + \int_a^b Rz'\,dt \\
&= \int\limits_C P\,dx + \int\limits_C Q\,dy + \int\limits_C R\,dz \\
&= \int\limits_C P\,dx + Q\,dy + R\,dz
\end{aligned}$$

So, we see that,

Fact

$$\int_C \vec{F} \cdot d\,\vec{r} = \int_C P\,dx + Q\,dy + R\,dz$$

Note that this gives us another method for evaluating line integrals of vector fields.

This also allows us to say the following about reversing the direction of the path with line integrals of vector fields.

Fact

$$\int_{-C} \vec{F} \cdot d\,\vec{r} = -\int_C \vec{F} \cdot d\,\vec{r}$$

This should make some sense given that we know that this is true for line integrals with respect to x, y, and/or z and that line integrals of vector fields can be defined in terms of line integrals with respect to x, y, and z.

16.5 Fundamental Theorem for Line Integrals

In Calculus I we had the Fundamental Theorem of Calculus that told us how to evaluate definite integrals. This told us,

$$\int_a^b F'(x)\,dx = F(b) - F(a)$$

It turns out that there is a version of this for line integrals over certain kinds of vector fields. Here it is.

Theorem

Suppose that C is a smooth curve given by $\vec{r}(t)$, $a \le t \le b$. Also suppose that f is a function whose gradient vector, ∇f, is continuous on C. Then,

$$\int\limits_C \nabla f \cdot d\vec{r} = f\big(\vec{r}(b)\big) - f\big(\vec{r}(a)\big)$$

Note that $\vec{r}(a)$ represents the initial point on C while $\vec{r}(b)$ represents the final point on C. Also, we did not specify the number of variables for the function since it is really immaterial to the theorem. The theorem will hold regardless of the number of variables in the function.

Proof

This is a fairly straight forward proof.

For the purposes of the proof we'll assume that we're working in three dimensions, but it can be done in any dimension.

Let's start by just computing the line integral.

$$\begin{aligned}\int\limits_C \nabla f \cdot d\vec{r} &= \int_a^b \nabla f\big(\vec{r}(t)\big) \cdot \vec{r}\,'(t)\,dt \\ &= \int_a^b \left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}\right) dt\end{aligned}$$

Now, at this point we can use the Chain Rule to simplify the integrand as follows,

$$\begin{aligned}\int\limits_C \nabla f \cdot d\vec{r} &= \int_a^b \left(\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}\right) dt \\ &= \int_a^b \frac{d}{dt}\Big[f\left(\vec{r}(t)\right)\Big]\,dt\end{aligned}$$

To finish this off we just need to use the Fundamental Theorem of Calculus for single integrals.

$$\int_C \nabla f \cdot d\vec{r} = f\left(\vec{r}\left(b\right)\right) - f\left(\vec{r}\left(a\right)\right)$$

Let's take a quick look at an example of using this theorem.

Example 1

Evaluate $\displaystyle\int_C \nabla f \cdot \, d\vec{r}$ where $f\left(x,y,z\right) = \cos\left(\pi x\right) + \sin\left(\pi y\right) - xyz$ and C is *any* path that starts at $\left(1, \frac{1}{2}, 2\right)$ and ends at $\left(2, 1, -1\right)$.

Solution

First let's notice that we didn't specify the path for getting from the first point to the second point. The reason for this is simple. The theorem above tells us that all we need are the initial and final points on the curve in order to evaluate this kind of line integral.

So, let $\vec{r}\left(t\right)$, $a \le t \le b$ be any path that starts at $\left(1, \frac{1}{2}, 2\right)$ and ends at $\left(2, 1, -1\right)$. Then,

$$\vec{r}\left(a\right) = \left\langle 1, \frac{1}{2}, 2\right\rangle \qquad\qquad \vec{r}\left(b\right) = \left\langle 2, 1, -1\right\rangle$$

The integral is then,

$$\begin{aligned}\int_C \nabla f \cdot \, d\vec{r} &= f\left(2,1,-1\right) - f\left(1, \frac{1}{2}, 2\right) \\ &= \cos\left(2\pi\right) + \sin\pi - 2\left(1\right)\left(-1\right) - \left(\cos\pi + \sin\left(\frac{\pi}{2}\right) - 1\left(\frac{1}{2}\right)\left(2\right)\right) \\ &= 4\end{aligned}$$

Notice that we also didn't need the gradient vector to actually do this line integral. However, for the practice of finding gradient vectors here it is,

$$\nabla f = \left\langle -\pi\sin\left(\pi x\right) - yz, \pi\cos\left(\pi y\right) - xz, -xy\right\rangle$$

The most important idea to get from this example is not how to do the integral as that's pretty simple, all we do is plug the final point and initial point into the function and subtract the two results. The important idea from this example (and hence about the Fundamental Theorem of Calculus) is that, for these kinds of line integrals, we didn't really need to know the path to get the answer. In other words, we could use any path we want and we'll always get the same results.

In the first section on line integrals (even though we weren't looking at vector fields) we saw that often when we change the path we will change the value of the line integral. We now have a type of line integral for which we know that changing the path will NOT change the value of the line integral.

Let's formalize this idea up a little. Here are some definitions. The first one we've already seen before, but it's been a while and it's important in this section so we'll give it again. The remaining definitions are new.

Definitions

First suppose that $\vec{F}$ is a continuous vector field in some domain D.

1. $\vec{F}$ is a **conservative** vector field if there is a function f such that $\vec{F} = \nabla f$. The function f is called a **potential function** for the vector field. We first saw this definition in the first section of this chapter.

2. $\int_C \vec{F} \cdot d\vec{r}$ is **independent of path** if $\int_{C_1} \vec{F} \cdot d\vec{r} = \int_{C_2} \vec{F} \cdot d\vec{r}$ for any two paths C_1 and C_2 in D with the same initial and final points.

3. A path C is called **closed** if its initial and final points are the same point. For example, a circle is a closed path.

4. A path C is **simple** if it doesn't cross itself. A circle is a simple curve while a figure 8 type curve is not simple.

5. A region D is **open** if it doesn't contain any of its boundary points.

6. A region D is **connected** if we can connect any two points in the region with a path that lies completely in D.

7. A region D is **simply-connected** if it is connected and it contains no holes. We won't need this one until the next section, but it fits in with all the other definitions given here so this was a natural place to put the definition.

With these definitions we can now give some nice facts.

Facts

1. $\int\limits_C \nabla f \cdot d\vec{r}$ is independent of path. This is easy enough to prove since all we need to do is look at the theorem above. The theorem tells us that in order to evaluate this integral all we need are the initial and final points of the curve. This in turn tells us that the line integral must be independent of path.

2. If $\vec{F}$ is a conservative vector field then $\int\limits_C \vec{F} \cdot d\vec{r}$ is independent of path. This fact is also easy enough to prove. If $\vec{F}$ is conservative then it has a potential function, f, and so the line integral becomes $\int\limits_C \vec{F} \cdot d\vec{r} = \int\limits_C \nabla f \cdot d\vec{r}$. Then using the first fact we know that this line integral must be independent of path.

3. If $\vec{F}$ is a continuous vector field on an open connected region D and if $\int\limits_C \vec{F} \cdot d\vec{r}$ is independent of path (for any path in D) then $\vec{F}$ is a conservative vector field on D.

4. If $\int\limits_C \vec{F} \cdot d\vec{r}$ is independent of path then $\int\limits_C \vec{F} \cdot d\vec{r} = 0$ for every closed path C.

5. If $\int\limits_C \vec{F} \cdot d\vec{r} = 0$ for every closed path C then $\int\limits_C \vec{F} \cdot d\vec{r}$ is independent of path.

These are some nice facts to remember as we work with line integrals over vector fields. Also notice that 2 & 3 and 4 & 5 are converses of each other.

16.6 Conservative Vector Fields

In the previous section we saw that if we knew that the vector field $\vec{F}$ was conservative then $\int\limits_C \vec{F} \cdot d\,\vec{r}$ was independent of path. This in turn means that we can easily evaluate this line integral provided we can find a potential function for $\vec{F}$.

In this section we want to look at two questions. First, given a vector field $\vec{F}$ is there any way of determining if it is a conservative vector field? Secondly, if we know that $\vec{F}$ is a conservative vector field how do we go about finding a potential function for the vector field?

The first question is easy to answer at this point if we have a two-dimensional vector field. For higher dimensional vector fields we'll need to wait until the final section in this chapter to answer this question. With that being said let's see how we do it for two-dimensional vector fields.

Theorem

Let $\vec{F} = P\,\vec{i} + Q\,\vec{j}$ be a vector field on an open and simply-connected region D. Then if P and Q have continuous first order partial derivatives in D and

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$$

the vector field $\vec{F}$ is conservative.

Let's take a look at a couple of examples.

Example 1

Determine if the following vector fields are conservative or not.

(a) $\vec{F}(x,y) = \left(x^2 - yx\right)\vec{i} + \left(y^2 - xy\right)\vec{j}$

(b) $\vec{F}(x,y) = \left(2x\mathbf{e}^{xy} + x^2y\mathbf{e}^{xy}\right)\vec{i} + \left(x^3\mathbf{e}^{xy} + 2y\right)\vec{j}$

Solution

Okay, there really isn't too much to these. All we do is identify P and Q then take a couple of derivatives and compare the results.

(a) $\vec{F}(x,y) = (x^2 - yx)\,\vec{i} + (y^2 - xy)\,\vec{j}$

In this case here is P and Q and the appropriate partial derivatives.

$$P = x^2 - yx \qquad \frac{\partial P}{\partial y} = -x$$
$$Q = y^2 - xy \qquad \frac{\partial Q}{\partial x} = -y$$

So, since the two partial derivatives are not the same this vector field is NOT conservative.

(b) $\vec{F}(x,y) = (2x\mathbf{e}^{xy} + x^2y\mathbf{e}^{xy})\,\vec{i} + (x^3\mathbf{e}^{xy} + 2y)\,\vec{j}$

Here is P and Q as well as the appropriate derivatives.

$$P = 2x\mathbf{e}^{xy} + x^2y\mathbf{e}^{xy} \qquad \frac{\partial P}{\partial y} = 2x^2\mathbf{e}^{xy} + x^2\mathbf{e}^{xy} + x^3y\mathbf{e}^{xy} = 3x^2\mathbf{e}^{xy} + x^3y\mathbf{e}^{xy}$$
$$Q = x^3\mathbf{e}^{xy} + 2y \qquad \frac{\partial Q}{\partial x} = 3x^2\mathbf{e}^{xy} + x^3y\mathbf{e}^{xy}$$

The two partial derivatives are equal and so this is a conservative vector field.

Now that we know how to identify if a two-dimensional vector field is conservative we need to address how to find a potential function for the vector field. This is actually a fairly simple process. First, let's assume that the vector field is conservative and so we know that a potential function, $f(x,y)$ exists. We can then say that,

$$\nabla f = \frac{\partial f}{\partial x}\vec{i} + \frac{\partial f}{\partial y}\vec{j} = P\,\vec{i} + Q\,\vec{j} = \vec{F}$$

Or by setting components equal we have,

Fact

$$\frac{\partial f}{\partial x} = P \qquad \text{and} \qquad \frac{\partial f}{\partial y} = Q$$

By integrating each of these with respect to the appropriate variable we can arrive at the following two equations.

Fact

$$f(x,y) = \int P(x,y)\,dx \qquad \text{or} \qquad f(x,y) = \int Q(x,y)\,dy$$

We saw this kind of integral briefly at the end of the section on iterated integrals in the previous chapter.

It is usually best to see how we use these two facts to find a potential function in an example or two.

Example 2

Determine if the following vector fields are conservative and find a potential function for the vector field if it is conservative.

(a) $\vec{F} = \left(2x^3y^4 + x\right)\vec{i} + \left(2x^4y^3 + y\right)\vec{j}$

(b) $\vec{F}(x,y) = \left(2x\mathbf{e}^{xy} + x^2y\mathbf{e}^{xy}\right)\vec{i} + \left(x^3\mathbf{e}^{xy} + 2y\right)\vec{j}$

Solution

(a) $\vec{F} = \left(2x^3y^4 + x\right)\vec{i} + \left(2x^4y^3 + y\right)\vec{j}$

Let's first identify P and Q and then check that the vector field is conservative.

$$P = 2x^3y^4 + x \qquad \frac{\partial P}{\partial y} = 8x^3y^3$$
$$Q = 2x^4y^3 + y \qquad \frac{\partial Q}{\partial x} = 8x^3y^3$$

So, the vector field is conservative. Now let's find the potential function. From the first fact above we know that,

$$\frac{\partial f}{\partial x} = 2x^3y^4 + x \qquad \frac{\partial f}{\partial y} = 2x^4y^3 + y$$

From these we can see that

$$f(x,y) = \int 2x^3y^4 + x\,dx \qquad \text{or} \qquad f(x,y) = \int 2x^4y^3 + y\,dy$$

We can use either of these to get the process started. Recall that we are going to have to be careful with the "constant of integration" which ever integral we choose to use. For this example let's work with the first integral and so that means that we are asking what function did we differentiate with respect to x to get the integrand. This means that the "constant of integration" is going to have to be a function of y since any function consisting only of y and/or constants will differentiate to zero when taking the partial derivative with respect to x.

Here is the first integral.

$$f(x,y) = \int 2x^3y^4 + x\,dx$$
$$= \frac{1}{2}x^4y^4 + \frac{1}{2}x^2 + h(y)$$

where $h(y)$ is the "constant of integration".

We now need to determine $h(y)$. This is easier than it might at first appear to be. To get to this point we've used the fact that we knew P, but we will also need to use the fact that we know Q to complete the problem. Recall that Q is really the derivative of f with respect to y. So, if we differentiate our function with respect to y we know what it should be.

So, let's differentiate f (including the $h(y)$) with respect to y and set it equal to Q since that is what the derivative is supposed to be.

$$\frac{\partial f}{\partial y} = 2x^4y^3 + h'(y) = 2x^4y^3 + y = Q$$

From this we can see that,

$$h'(y) = y$$

Notice that since $h'(y)$ is a function only of y so if there are any x's in the equation at this point we will know that we've made a mistake. At this point finding $h(y)$ is simple.

$$h(y) = \int h'(y)\,dy = \int y\,dy = \frac{1}{2}y^2 + c$$

So, putting this all together we can see that a potential function for the vector field is,

$$f(x,y) = \frac{1}{2}x^4y^4 + \frac{1}{2}x^2 + \frac{1}{2}y^2 + c$$

Note that we can always check our work by verifying that $\nabla f = \vec{F}$. Also note that because the c can be anything there are an infinite number of possible potential functions, although they will only vary by an additive constant.

(b) $\vec{F}(x,y) = \left(2x\mathbf{e}^{xy} + x^2y\mathbf{e}^{xy}\right)\vec{i} + \left(x^3\mathbf{e}^{xy} + 2y\right)\vec{j}$

Okay, this one will go a lot faster since we don't need to go through as much explanation. We've already verified that this vector field is conservative in the first set of examples so we won't bother redoing that.

Let's start with the following,

$$\frac{\partial f}{\partial x} = 2x\mathbf{e}^{xy} + x^2y\mathbf{e}^{xy} \qquad \frac{\partial f}{\partial y} = x^3\mathbf{e}^{xy} + 2y$$

This means that we can do either of the following integrals,

$$f(x,y) = \int 2x\mathbf{e}^{xy} + x^2 y\mathbf{e}^{xy}\,dx \qquad \text{or} \qquad f(x,y) = \int x^3\mathbf{e}^{xy} + 2y\,dy$$

While we can do either of these the first integral would be somewhat unpleasant as we would need to do integration by parts on each portion. On the other hand, the second integral is fairly simple since the second term only involves y's and the first term can be done with the substitution $u = xy$. So, from the second integral we get,

$$f(x,y) = x^2\mathbf{e}^{xy} + y^2 + h(x)$$

Notice that this time the "constant of integration" will be a function of x. If we differentiate this with respect to x and set equal to P we get,

$$\frac{\partial f}{\partial x} = 2x\mathbf{e}^{xy} + x^2 y\mathbf{e}^{xy} + h'(x) = 2x\mathbf{e}^{xy} + x^2 y\mathbf{e}^{xy} = P$$

So, in this case it looks like,

$$h'(x) = 0 \qquad \Rightarrow \qquad h(x) = c$$

So, in this case the "constant of integration" really was a constant. Sometimes this will happen and sometimes it won't.

Here is the potential function for this vector field.

$$f(x,y) = x^2\mathbf{e}^{xy} + y^2 + c$$

Now, as noted above we don't have a way (yet) of determining if a three-dimensional vector field is conservative or not. However, if we are given that a three-dimensional vector field is conservative finding a potential function is similar to the above process, although the work will be a little more involved.

In this case we will use the fact that,

$$\nabla f = \frac{\partial f}{\partial x}\,\vec{i} + \frac{\partial f}{\partial y}\,\vec{j} + \frac{\partial f}{\partial z}\,\vec{k} = P\,\vec{i} + Q\,\vec{j} + R\,\vec{k} = \vec{F}$$

Let's take a quick look at an example.

Example 3

Find a potential function for the vector field,

$$\vec{F} = 2xy^3z^4\,\vec{i} + 3x^2y^2z^4\,\vec{j} + 4x^2y^3z^3\,\vec{k}$$

Solution

Okay, we'll start off with the following equalities.

$$\frac{\partial f}{\partial x} = 2xy^3z^4 \qquad \frac{\partial f}{\partial y} = 3x^2y^2z^4 \qquad \frac{\partial f}{\partial z} = 4x^2y^3z^3$$

To get started we can integrate the first one with respect to x, the second one with respect to y, or the third one with respect to z. Let's integrate the first one with respect to x.

$$f\left(x,y,z\right) = \int 2xy^3z^4\,dx = x^2y^3z^4 + g\left(y,z\right)$$

Note that this time the "constant of integration" will be a function of both y and z since differentiating anything of that form with respect to x will differentiate to zero.

Now, we can differentiate this with respect to y and set it equal to Q. Doing this gives,

$$\frac{\partial f}{\partial y} = 3x^2y^2z^4 + g_y\left(y,z\right) = 3x^2y^2z^4 = Q$$

Of course we'll need to take the partial derivative of the constant of integration since it is a function of two variables. It looks like we've now got the following,

$$g_y\left(y,z\right) = 0 \qquad \Rightarrow \qquad g\left(y,z\right) = h\left(z\right)$$

Since differentiating $g\left(y,z\right)$ with respect to y gives zero then $g\left(y,z\right)$ could at most be a function of z. This means that we now know the potential function must be in the following form.

$$f\left(x,y,z\right) = x^2y^3z^4 + h\left(z\right)$$

To finish this out all we need to do is differentiate with respect to z and set the result equal to R.

$$\frac{\partial f}{\partial z} = 4x^2y^3z^3 + h'\left(z\right) = 4x^2y^3z^3 = R$$

So,

$$h'\left(z\right) = 0 \qquad \Rightarrow \qquad h\left(z\right) = c$$

The potential function for this vector field is then,

$$f\left(x,y,z\right) = x^2y^3z^4 + c$$

Note that to keep the work to a minimum we used a fairly simple potential function for this example. It might have been possible to guess what the potential function was based simply on the vector field. However, we should be careful to remember that this usually won't be the case and often this process is required.

Also, there were several other paths that we could have taken to find the potential function. Each would have gotten us the same result.

Let's work one more slightly (and only slightly) more complicated example.

Example 4

Find a potential function for the vector field,

$$\vec{F} = \left(2x\cos\left(y\right) - 2z^3\right)\,\vec{i} + \left(3 + 2y\mathbf{e}^{\,z} - x^2\sin\left(y\right)\right)\,\vec{j} + \left(y^2\mathbf{e}^{\,z} - 6xz^2\right)\,\vec{k}$$

Solution

Here are the equalities for this vector field.

$$\frac{\partial f}{\partial x} = 2x\cos\left(y\right) - 2z^3 \qquad \frac{\partial f}{\partial y} = 3 + 2y\mathbf{e}^{\,z} - x^2\sin\left(y\right) \qquad \frac{\partial f}{\partial z} = y^2\mathbf{e}^{\,z} - 6xz^2$$

For this example let's integrate the third one with respect to z.

$$f\left(x,y,z\right) = \int y^2\mathbf{e}^{\,z} - 6xz^2\,dz = y^2\mathbf{e}^{\,z} - 2xz^3 + g\left(x,y\right)$$

The "constant of integration" for this integration will be a function of both x and y.

Now, we can differentiate this with respect to x and set it equal to P. Doing this gives,

$$\frac{\partial f}{\partial x} = -2z^3 + g_x\left(x,y\right) = 2x\cos\left(y\right) - 2z^3 = P$$

So, it looks like we've now got the following,

$$g_x\left(x,y\right) = 2x\cos\left(y\right) \qquad \Rightarrow \qquad g\left(x,y\right) = x^2\cos\left(y\right) + h\left(y\right)$$

The potential function for this problem is then,

$$f\left(x,y,z\right) = y^2\mathbf{e}^{\,z} - 2xz^3 + x^2\cos\left(y\right) + h\left(y\right)$$

To finish this out all we need to do is differentiate with respect to y and set the result equal to Q.

$$\frac{\partial f}{\partial y} = 2y\mathbf{e}^{\,z} - x^2\sin\left(y\right) + h'\left(y\right) = 3 + 2y\mathbf{e}^{\,z} - x^2\sin\left(y\right) = Q$$

So,

$$h'(y) = 3 \qquad \Rightarrow \qquad h(y) = 3y + c$$

The potential function for this vector field is then,

$$f(x,y,z) = y^2 \mathbf{e}^{z} - 2xz^3 + x^2 \cos(y) + 3y + c$$

So, a little more complicated than the others and there are again many different paths that we could have taken to get the answer.

We need to work one final example in this section.

Example 5

Evaluate $\int\limits_C \vec{F} \cdot d\,\vec{r}$ where $\vec{F} = \left(2x^3y^4 + x\right)\vec{i} + \left(2x^4y^3 + y\right)\vec{j}$ and C is given by $\vec{r}(t) = (t\cos(\pi t) - 1)\,\vec{i} + \sin\left(\frac{\pi t}{2}\right)\vec{j}$, $0 \le t \le 1$.

Solution

Now, we could use the techniques we discussed when we first looked at line integrals of vector fields however that would be particularly unpleasant solution.

Instead, let's take advantage of the fact that we know from Example 2a above this vector field is conservative and that a potential function for the vector field is,

$$f(x,y) = \frac{1}{2}x^4y^4 + \frac{1}{2}x^2 + \frac{1}{2}y^2 + c$$

Using this we know that integral must be independent of path and so all we need to do is use the theorem from the previous section to do the evaluation.

$$\int\limits_C \vec{F} \cdot d\,\vec{r} = \int\limits_C \nabla f \cdot d\,\vec{r} = f\left(\vec{r}(1)\right) - f\left(\vec{r}(0)\right)$$

where,

$$\vec{r}(1) = \langle -2, 1 \rangle \qquad \vec{r}(0) = \langle -1, 0 \rangle$$

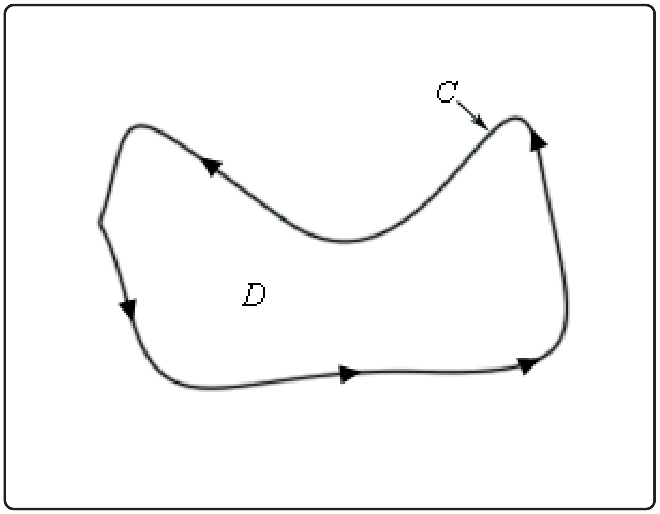
So, the integral is,

$$\begin{aligned} \int\limits_C \vec{F} \cdot d\,\vec{r} &= f(-2,1) - f(-1,0) \\ &= \left(\frac{21}{2} + c\right) - \left(\frac{1}{2} + c\right) \\ &= 10 \end{aligned}$$

16.7 Green's Theorem

In this section we are going to investigate the relationship between certain kinds of line integrals (on closed paths) and double integrals.

Let's start off with a simple (recall that this means that it doesn't cross itself) closed curve C and let D be the region enclosed by the curve. Here is a sketch of such a curve and region.



First, notice that because the curve is simple and closed there are no holes in the region D. Also notice that a direction has been put on the curve. We will use the convention here that the curve C has a **positive orientation** if it is traced out in a counter-clockwise direction. Another way to think of a positive orientation (that will cover much more general curves as well see later) is that as we traverse the path following the positive orientation the region D must always be on the left.

Given curves/regions such as this we have the following theorem.

Green's Theorem

Let C be a positively oriented, piecewise smooth, simple, closed curve and let D be the region enclosed by the curve. If P and Q have continuous first order partial derivatives on D then,

$$\int\limits_C P\,dx + Q\,dy = \iint\limits_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dA$$

Before working some examples there are some alternate notations that we need to acknowledge. When working with a line integral in which the path satisfies the condition of Green's Theorem we will often denote the line integral as,

$$\oint_C P\,dx + Q\,dy \qquad \text{or} \qquad \oint\limits_C P\,dx + Q\,dy$$

Both of these notations do assume that C satisfies the conditions of Green's Theorem so be careful in using them.

Also, sometimes the curve C is not thought of as a separate curve but instead as the boundary of some region D and in these cases you may see C denoted as ∂D.

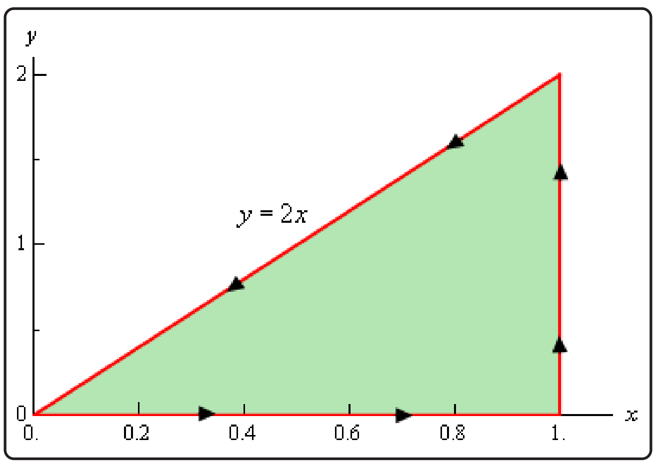
Let's work a couple of examples.

Example 1

Use Green's Theorem to evaluate $\oint_C xy\,dx + x^2y^3\,dy$ where C is the triangle with vertices $(0,0)$, $(1,0)$, $(1,2)$ with positive orientation.

Solution

Let's first sketch C and D for this case to make sure that the conditions of Green's Theorem are met for C and will need the sketch of D to evaluate the double integral.



So, the curve does satisfy the conditions of Green's Theorem and we can see that the following inequalities will define the region enclosed.

$$0 \le x \le 1 \qquad 0 \le y \le 2x$$

We can identify P and Q from the line integral. Here they are.

$$P = xy \qquad Q = x^2y^3$$

So, using Green's Theorem the line integral becomes,

$$\begin{aligned}\oint_C xy\,dx + x^2y^3\,dy &= \iint_D 2xy^3 - x\,dA \\ &= \int_0^1 \int_0^{2x} 2xy^3 - x\,dy\,dx \\ &= \int_0^1 \left(\frac{1}{2}xy^4 - xy\right)\Big|_0^{2x}\,dx \\ &= \int_0^1 8x^5 - 2x^2\,dx \\ &= \left(\frac{4}{3}x^6 - \frac{2}{3}x^3\right)\Big|_0^1 \\ &= \frac{2}{3}\end{aligned}$$

Example 2

Evaluate $\oint_C y^3\,dx - x^3\,dy$ where C is the positively oriented circle of radius 2 centered at the origin.

Solution

Okay, a circle will satisfy the conditions of Green's Theorem since it is closed and simple and so there really isn't a reason to sketch it.

Let's first identify P and Q from the line integral.

$$P = y^3 \qquad\qquad Q = -x^3$$

Be careful with the minus sign on Q!

Now, using Green's theorem on the line integral gives,

$$\oint_C y^3\,dx - x^3\,dy = \iint_D -3x^2 - 3y^2\,dA$$

where D is a disk of radius 2 centered at the origin.

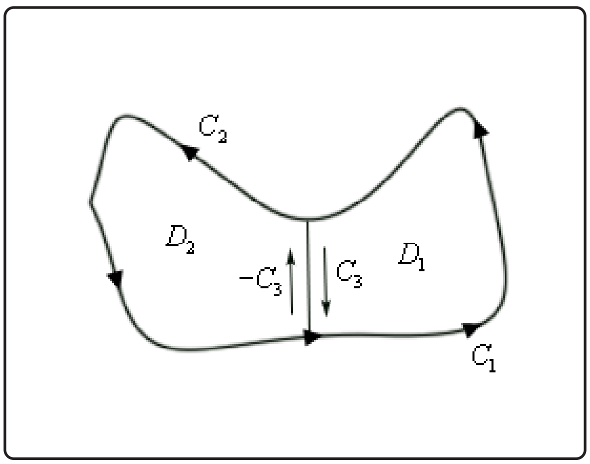
Since D is a disk it seems like the best way to do this integral is to use polar coordinates.

Here is the evaluation of the integral.

$$\begin{aligned}\oint_C y^3\,dx - x^3\,dy &= -3\iint\limits_D \left(x^2+y^2\right)\,dA \\ &= -3\int_0^{2\pi}\int_0^2 r^3\,dr\,d\theta \\ &= -3\int_0^{2\pi} \left.\frac{1}{4}r^4\right|_0^2\,d\theta \\ &= -3\int_0^{2\pi} 4\,d\theta \\ &= -24\pi\end{aligned}$$

So, Green's theorem, as stated, will not work on regions that have holes in them. However, many regions do have holes in them. So, let's see how we can deal with those kinds of regions.

Let's start with the following region. Even though this region doesn't have any holes in it the arguments that we're going to go through will be similar to those that we'd need for regions with holes in them, except it will be a little easier to deal with and write down.



The region D will be $D_1 \cup D_2$ and recall that the symbol $\cup$ is called the union and means that D consists of both D_1 and D_2. The boundary of D_1 is $C_1 \cup C_3$ while the boundary of D_2 is $C_2 \cup (-C_3)$ and notice that both of these boundaries are positively oriented. As we traverse each boundary the corresponding region is always on the left. Finally, also note that we can think of the whole boundary, C, as,

$$C = (C_1 \cup C_3) \cup (C_2 \cup (-C_3)) = C_1 \cup C_2$$

since both C_3 and $-C_3$ will "cancel" each other out.

Now, let's start with the following double integral and use a basic property of double integrals to break it up.

$$\iint\limits_{D} (Q_x - P_y)\, dA = \iint\limits_{D_1 \cup D_2} (Q_x - P_y)\, dA = \iint\limits_{D_1} (Q_x - P_y)\, dA + \iint\limits_{D_2} (Q_x - P_y)\, dA$$

Next, use Green's theorem on each of these and again use the fact that we can break up line integrals into separate line integrals for each portion of the boundary.

$$\begin{aligned}
\iint\limits_{D} (Q_x - P_y)\, dA &= \iint\limits_{D_1} (Q_x - P_y)\, dA + \iint\limits_{D_2} (Q_x - P_y)\, dA \\
&= \oint\limits_{C_1 \cup C_3} P dx + Q dy + \oint\limits_{C_2 \cup (-C_3)} P dx + Q dy \\
&= \oint\limits_{C_1} P dx + Q dy + \oint\limits_{C_3} P dx + Q dy + \oint\limits_{C_2} P dx + Q dy + \oint\limits_{-C_3} P dx + Q dy
\end{aligned}$$

Next, we'll use the fact that,

$$\oint\limits_{-C_3} P dx + Q dy = -\oint\limits_{C_3} P dx + Q dy$$

Recall that changing the orientation of a curve with line integrals with respect to x and/or y will simply change the sign on the integral. Using this fact we get,

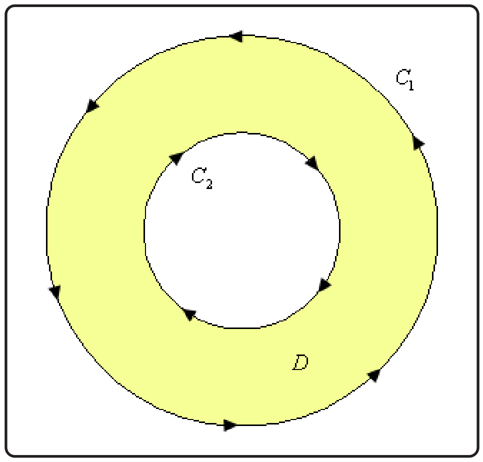
$$\begin{aligned}
\iint\limits_{D} (Q_x - P_y)\, dA &= \oint\limits_{C_1} P dx + Q dy + \oint\limits_{C_3} P dx + Q dy + \oint\limits_{C_2} P dx + Q dy - \oint\limits_{C_3} P dx + Q dy \\
&= \oint\limits_{C_1} P dx + Q dy + \oint\limits_{C_2} P dx + Q dy
\end{aligned}$$

Finally, put the line integrals back together and we get,

$$\begin{aligned}
\iint\limits_{D} (Q_x - P_y)\, dA &= \oint\limits_{C_1} P dx + Q dy + \oint\limits_{C_2} P dx + Q dy \\
&= \oint\limits_{C_1 \cup C_2} P dx + Q dy \\
&= \oint\limits_{C} P dx + Q dy
\end{aligned}$$

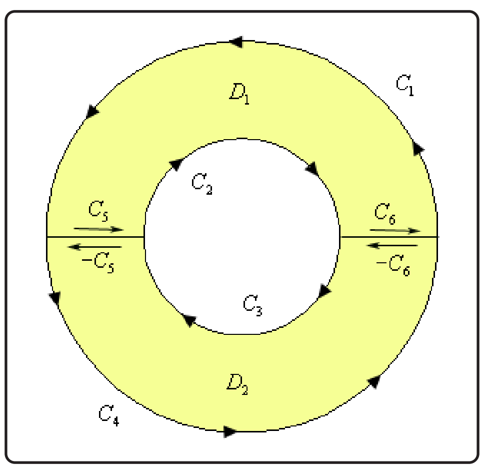
So, what did we learn from this? If you think about it this was just a lot of work and all we got out of it was the result from Green's Theorem which we already knew to be true. What this exercise has shown us is that if we break a region up as we did above then the portion of the line integral on the pieces of the curve that are in the middle of the region (each of which are in the opposite direction) will cancel out. This idea will help us in dealing with regions that have holes in them.

To see this let's look at a ring.



Notice that both of the curves are oriented positively since the region D is on the left side as we traverse the curve in the indicated direction. Note as well that the curve C_2 seems to violate the original definition of positive orientation. We originally said that a curve had a positive orientation if it was traversed in a counter-clockwise direction. However, this was only for regions that do not have holes. For the boundary of the hole this definition won't work and we need to resort to the second definition that we gave above.

Now, since this region has a hole in it we will apparently not be able to use Green's Theorem on any line integral with the curve $C = C_1 \cup C_2$. However, if we cut the disk in half and rename all the various portions of the curves we get the following sketch.



The boundary of the upper portion (D_1)of the disk is $C_1 \cup C_2 \cup C_5 \cup C_6$ and the boundary on the lower portion (D_2)of the disk is $C_3 \cup C_4 \cup (-C_5) \cup (-C_6)$. Also notice that we can use Green's Theorem on each of these new regions since they don't have any holes in them. This means that we can do the following,

$$\begin{aligned}\iint\limits_{D}\left(Q_x - P_y\right)\,dA &= \iint\limits_{D_1}\left(Q_x - P_y\right)\,dA + \iint\limits_{D_2}\left(Q_x - P_y\right)\,dA \\ &= \oint_{C_1\cup C_2\cup C_5\cup C_6} P dx + Q dy + \oint_{C_3\cup C_4\cup(-C_5)\cup(-C_6)} P dx + Q dy\end{aligned}$$

Now, we can break up the line integrals into line integrals on each piece of the boundary. Also recall from the work above that boundaries that have the same curve, but opposite direction will cancel. Doing this gives,

$$\begin{aligned}\iint\limits_{D}\left(Q_x - P_y\right)\,dA &= \iint\limits_{D_1}\left(Q_x - P_y\right)\,dA + \iint\limits_{D_2}\left(Q_x - P_y\right)\,dA \\ &= \oint_{C_1} P dx + Q dy + \oint_{C_2} P dx + Q dy + \oint_{C_3} P dx + Q dy + \oint_{C_4} P dx + Q dy\end{aligned}$$

But at this point we can add the line integrals back up as follows,

$$\begin{aligned}\iint\limits_{D}\left(Q_x - P_y\right)\,dA &= \oint_{C_1\cup C_2\cup C_3\cup C_4} P dx + Q dy \\ &= \oint_{C} P dx + Q dy\end{aligned}$$

The end result of all of this is that we could have just used Green's Theorem on the disk from the start even though there is a hole in it. This will be true in general for regions that have holes in them.

Let's take a look at an example.

Example 3

Evaluate $\displaystyle\oint_C y^3\,dx - x^3\,dy$ where C are the two circles of radius 2 and radius 1 centered at the origin with positive orientation.

Solution

Notice that this is the same line integral as we looked at in the second example and only the curve has changed. In this case the region D will now be the region between these two circles and that will only change the limits in the double integral so we'll not put in some of the details here.

Here is the work for this integral.

$$\begin{aligned}\oint_C y^3\,dx - x^3\,dy &= -3\iint\limits_D \left(x^2+y^2\right)\,dA\\ &= -3\int_0^{2\pi}\int_1^2 r^3\,dr\,d\theta\\ &= -3\int_0^{2\pi} \left.\frac{1}{4}r^4\right|_1^2\,d\theta\\ &= -3\int_0^{2\pi}\frac{15}{4}\,d\theta\\ &= -\frac{45\pi}{2}\end{aligned}$$

We will close out this section with an interesting application of Green's Theorem. Recall that we can determine the area of a region D with the following double integral.

$$A = \iint\limits_D dA$$

Let's think of this double integral as the result of using Green's Theorem. In other words, let's assume that

$$Q_x - P_y = 1$$

and see if we can get some functions P and Q that will satisfy this.

There are many functions that will satisfy this. Here are some of the more common functions.

$$\begin{array}{lll} P = 0 & P = -y & P = -\frac{y}{2}\\ Q = x & Q = 0 & Q = \frac{x}{2}\end{array}$$

Then, if we use Green's Theorem in reverse we see that the area of the region D can also be computed by evaluating any of the following line integrals.

Fact

$$A = \oint\limits_C x\,dy = -\oint\limits_C y\,dx = \frac{1}{2}\oint\limits_C x\,dy - y\,dx$$

where C is the boundary of the region D.

Let's take a quick look at an example of this.

Example 4

Use Green's Theorem to find the area of a disk of radius a.

Solution

We can use either of the integrals above, but the third one is probably the easiest. So,

$$A = \frac{1}{2}\oint_C x\,dy - y\,dx$$

where C is the circle of radius a. So, to do this we'll need a parameterization of C. This is,

$$x = a\cos(t) \qquad y = a\sin(t) \qquad 0 \le t \le 2\pi$$

The area is then,

$$\begin{aligned} A &= \frac{1}{2}\oint_C x\,dy - y\,dx \\ &= \frac{1}{2}\left(\int_0^{2\pi} a\cos(t)\big(a\cos(t)\big)\,dt - \int_0^{2\pi} a\sin(t)\big(-a\sin(t)\big)\,dt\right) \\ &= \frac{1}{2}\int_0^{2\pi} a^2\cos^2(t) + a^2\sin^2(t)\,dt \\ &= \frac{1}{2}\int_0^{2\pi} a^2 dt \\ &= \pi a^2 \end{aligned}$$

17 Surface Integrals

In this chapter we are going to take a look at surface integrals. In the previous chapter we integrated a line integral of a function of three variables where the variables came from a three dimensional curve. In this chapter we want to integrate a function of three variables but now the variables will come from a three dimensional solid. As with line integrals we will integrate both functions and vector fields.

We will also introduce the concept of the curl and divergence of a vector field. In addition, we will discuss how to write down a set of parametric equations for a surface.

We will close out the chapter by discussing Stokes' Theorem and the Divergence Theorem. Stokes' Theorem will give a nice relationship between line integrals and surface integrals. The Divergence Theorem will give a relationship between surface integrals and triple integrals.

17.1 Curl and Divergence

Before we can get into surface integrals we need to get some introductory material out of the way. That is the purpose of the first two sections of this chapter.

In this section we are going to introduce the concepts of the curl and the divergence of a vector.

Let's start with the curl. Given the vector field $\vec{F} = P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}$ the curl is defined to be,

Curl

$$\text{curl}\vec{F} = (R_y - Q_z)\,\vec{i} + (P_z - R_x)\,\vec{j} + (Q_x - P_y)\,\vec{k}$$

There is another (potentially) easier definition of the curl of a vector field. To use it we will first need to define the ∇ **operator**. This is defined to be,

Definition

$$\nabla = \frac{\partial}{\partial x}\,\vec{i} + \frac{\partial}{\partial y}\,\vec{j} + \frac{\partial}{\partial z}\,\vec{k}$$

We use this as if it's a function in the following manner.

$$\nabla f = \frac{\partial f}{\partial x}\,\vec{i} + \frac{\partial f}{\partial y}\,\vec{j} + \frac{\partial f}{\partial z}\,\vec{k}$$

So, whatever function is listed after the ∇ is substituted into the partial derivatives. Note as well that when we look at it in this light we simply get the gradient vector.

Using the ∇ we can define the curl as the following cross product,

Curl, Alternate Formula

$$\text{curl}\vec{F} = \nabla \times \vec{F} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}$$

We have a couple of nice facts that use the curl of a vector field.

Facts

1. If $f(x, y, z)$ has continuous second order partial derivatives then curl$(\nabla f) = \vec{0}$. This is easy enough to check by plugging into the definition of the derivative so we'll leave it to you to check.
2. If $\vec{F}$ is a conservative vector field then curl$\vec{F} = \vec{0}$. This is a direct result of what it means to be a conservative vector field and the previous fact.
3. If $\vec{F}$ is defined on all of $\mathbb{R}^3$ whose components have continuous first order partial derivative and curl$\vec{F} = \vec{0}$ then $\vec{F}$ is a conservative vector field. This is not so easy to verify and so we won't try.

Example 1

Determine if $\vec{F} = x^2y\,\vec{i} + xyz\,\vec{j} - x^2y^2\,\vec{k}$ is a conservative vector field.

Solution

So, all that we need to do is compute the curl and see if we get the zero vector or not.

$$\begin{aligned}\text{curl}\vec{F} &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ x^2y & xyz & -x^2y^2 \end{vmatrix} \\ &= -2x^2y\,\vec{i} + yz\,\vec{k} - \left(-2xy^2\,\vec{j}\right) - xy\,\vec{i} - x^2\vec{k} \\ &= -\left(2x^2y + xy\right)\vec{i} + 2xy^2\,\vec{j} + \left(yz - x^2\right)\vec{k} \\ &\neq \vec{0}\end{aligned}$$

So, the curl isn't the zero vector and so this vector field is not conservative.

Next, we should talk about a physical interpretation of the curl. Suppose that $\vec{F}$ is the velocity field of a flowing fluid. Then curl$\vec{F}$ represents the tendency of particles at the point (x, y, z) to rotate about the axis that points in the direction of curl$\vec{F}$. If curl$\vec{F} = \vec{0}$ then the fluid is called irrotational.

Let's now talk about the second new concept in this section. Given the vector field $\vec{F} = P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}$ the divergence is defined to be,

Divergence

$$\text{div}\vec{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}$$

There is also a definition of the divergence in terms of the ∇ operator. The divergence can be defined in terms of the following dot product.

Divergence, Alternative Formula

$$\text{div}\vec{F} = \nabla \cdot \vec{F}$$

Example 2

Compute div$\vec{F}$ for $\vec{F} = x^2y\,\vec{i} + xyz\,\vec{j} - x^2y^2\,\vec{k}$

Solution

There really isn't much to do here other than compute the divergence.

$$\text{div}\vec{F} = \frac{\partial}{\partial x}\left(x^2y\right) + \frac{\partial}{\partial y}\left(xyz\right) + \frac{\partial}{\partial z}\left(-x^2y^2\right) = 2xy + xz$$

We also have the following fact about the relationship between the curl and the divergence.

$$\text{div}\left(\text{curl}\vec{F}\right) = 0$$

Example 3

Verify the above fact for the vector field $\vec{F} = yz^2\,\vec{i} + xy\,\vec{j} + yz\,\vec{k}$.

Solution

Let's first compute the curl.

$$\begin{aligned}\text{curl}\vec{F} &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ yz^2 & xy & yz \end{vmatrix} \\ &= z\,\vec{i} + 2yz\,\vec{j} + y\,\vec{k} - z^2\vec{k} \\ &= z\vec{i} + 2yz\,\vec{j} + \left(y - z^2\right)\vec{k}\end{aligned}$$

Now compute the divergence of this.

$$\text{div}\left(\text{curl}\vec{F}\right) = \frac{\partial}{\partial x}\left(z\right) + \frac{\partial}{\partial y}\left(2yz\right) + \frac{\partial}{\partial z}\left(y - z^2\right) = 2z - 2z = 0$$

We also have a physical interpretation of the divergence. If we again think of $\vec{F}$ as the velocity field of a flowing fluid then div$\vec{F}$ represents the net rate of change of the mass of the fluid flowing from the point (x, y, z) per unit volume. This can also be thought of as the tendency of a fluid to diverge from a point. If div$\vec{F} = 0$ then the $\vec{F}$ is called incompressible.

The next topic that we want to briefly mention is the **Laplace** operator. Let's first take a look at,

$$\text{div}\,(\nabla f) = \nabla \cdot \nabla f = f_{xx} + f_{yy} + f_{zz}$$

The Laplace operator is then defined as,

$$\nabla^2 = \nabla \cdot \nabla$$

The Laplace operator arises naturally in many fields including heat transfer and fluid flow.

The final topic in this section is to give two vector forms of Green's Theorem. The first form uses the curl of the vector field and is,

Fact

$$\oint_C \vec{F} \cdot d\,\vec{r} = \iint\limits_D \left(\text{curl}\vec{F}\right) \cdot \vec{k}\,dA$$

where $\vec{k}$ is the standard unit vector in the positive z direction.

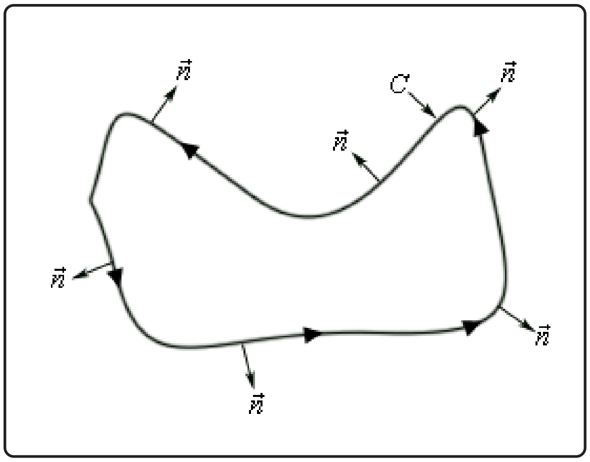
The second form uses the divergence. In this case we also need the outward unit normal to the curve C. If the curve is parameterized by

$$\vec{r}\,(t) = x\,(t)\,\vec{i} + y\,(t)\,\vec{j}$$

then the outward unit normal is given by,

$$\vec{n} = \frac{y'\,(t)}{\|\vec{r}'\,(t)\|}\vec{i} - \frac{x'\,(t)}{\|\vec{r}'\,(t)\|}\vec{j}$$

Here is a sketch illustrating the outward unit normal for some curve C at various points.



The vector form of Green's Theorem that uses the divergence is given by,

Fact

$$\oint_C \vec{F} \cdot \vec{n}\, ds = \iint\limits_D \text{div}\vec{F}\, dA$$

17.2 Parametric Surfaces

The final topic that we need to discuss before getting into surface integrals is how to parameterize a surface. When we parameterized a curve we took values of t from some interval $[a, b]$ and plugged them into

$$\vec{r}(t) = x(t)\,\vec{i} + y(t)\,\vec{j} + z(t)\,\vec{k}$$

and the resulting set of vectors will be the position vectors for the points on the curve.

With surfaces we'll do something similar. We will take points, (u, v), out of some two-dimensional space D and plug them into

$$\vec{r}(u,v) = x(u,v)\,\vec{i} + y(u,v)\,\vec{j} + z(u,v)\,\vec{k}$$

and the resulting set of vectors will be the position vectors for the points on the surface S that we are trying to parameterize. This is often called the **parametric representation** of the **parametric surface** S.

We will sometimes need to write the **parametric equations** for a surface. There are really nothing more than the components of the parametric representation explicitly written down.

$$x = x(u,v) \qquad y = y(u,v) \qquad z = z(u,v)$$

Example 1

Determine the surface given by the parametric representation.

$$\vec{r}(u,v) = u\,\vec{i} + u\cos(v)\,\vec{j} + u\sin(v)\,\vec{k}$$

Solution

Let's first write down the parametric equations.

$$x = u \qquad y = u\cos(v) \qquad z = u\sin(v)$$

Now if we square y and z and then add them together we get,

$$y^2 + z^2 = u^2\cos^2(v) + u^2\sin^2(v) = u^2\left(\cos^2(v) + \sin^2(v)\right) = u^2 = x^2$$

So, we were able to eliminate the parameters and the equation in x, y, and z is given by,

$$x^2 = y^2 + z^2$$

From the Quadric Surfaces section notes we can see that this is a cone that opens along the x-axis.

We are much more likely to need to be able to write down the parametric equations of a surface than identify the surface from the parametric representation so let's take a look at some examples of this.

Example 2

Give parametric representations for each of the following surfaces.

(a) The elliptic paraboloid $x = 5y^2 + 2z^2 - 10$.

(b) The elliptic paraboloid $x = 5y^2 + 2z^2 - 10$ that is in front of the yz-plane.

(c) The sphere $x^2 + y^2 + z^2 = 30$.

(d) The cylinder $y^2 + z^2 = 25$.

Solution

(a) The elliptic paraboloid $x = 5y^2 + 2z^2 - 10$.

This one is probably the easiest one of the four to see how to do. Since the surface is in the form $x = f(y, z)$ we can quickly write down a set of parametric equations as follows,

$$x = 5y^2 + 2z^2 - 10 \qquad y = y \qquad z = z$$

The last two equations are just there to acknowledge that we can choose y and z to be anything we want them to be. The parametric representation is then,

$$\vec{r}(y, z) = \left(5y^2 + 2z^2 - 10\right)\vec{i} + y\,\vec{j} + z\,\vec{k}$$

(b) The elliptic paraboloid $x = 5y^2 + 2z^2 - 10$ that is in front of the yz-plane.

This is really a restriction on the previous parametric representation. The parametric representation stays the same.

$$\vec{r}(y, z) = \left(5y^2 + 2z^2 - 10\right)\vec{i} + y\,\vec{j} + z\,\vec{k}$$

However, since we only want the surface that lies in front of the yz-plane we also need to require that $x \geq 0$. This is equivalent to requiring,

$$5y^2 + 2z^2 - 10 \geq 0 \qquad \text{or} \qquad 5y^2 + 2z^2 \geq 10$$

(c) The sphere $x^2 + y^2 + z^2 = 30$.

This one can be a little tricky until you see how to do it. In spherical coordinates we know that the equation of a sphere of radius a is given by,

$$\rho = a$$

and so the equation of this sphere (in spherical coordinates) is $\rho = \sqrt{30}$. Now, we also have the following conversion formulas for converting Cartesian coordinates into spherical coordinates.

$$x = \rho \sin(\varphi) \cos(\theta) \qquad y = \rho \sin(\varphi) \sin(\theta) \qquad z = \rho \cos(\varphi)$$

However, we know what ρ is for our sphere and so if we plug this into these conversion formulas we will arrive at a parametric representation for the sphere. Therefore, the parametric representation is,

$$\vec{r}(\theta, \varphi) = \sqrt{30} \sin(\varphi) \cos(\theta) \vec{i} + \sqrt{30} \sin(\varphi) \sin(\theta) \vec{j} + \sqrt{30} \cos(\varphi) \vec{k}$$

All we need to do now is come up with some restriction on the variables. First, we know that we have the following restriction.

$$0 \le \varphi \le \pi$$

This is enforced upon us by choosing to use spherical coordinates. Also, to make sure that we only trace out the sphere once we will also have the following restriction.

$$0 \le \theta \le 2\pi$$

(d) The cylinder $y^2 + z^2 = 25$.

As with the last one this can be tricky until you see how to do it. In this case it makes some sense to use cylindrical coordinates since they can be easily used to write down the equation of a cylinder.

In cylindrical coordinates the equation of a cylinder of radius a is given by

$$r = a$$

and so the equation of the cylinder in this problem is $r = 5$.

Next, we have the following conversion formulas.

$$x = x \qquad y = r \sin(\theta) \qquad z = r \cos(\theta)$$

Notice that they are slightly different from those that we are used to seeing. We needed to change them up here since the cylinder was centered upon the x-axis.

Finally, we know what r is so we can easily write down a parametric representation for this cylinder.

$$\vec{r}(x,\theta) = x\,\vec{i} + 5\sin(\theta)\,\vec{j} + 5\cos(\theta)\,\vec{k}$$

We will also need the restriction $0 \le \theta \le 2\pi$ to make sure that we don't retrace any portion of the cylinder. Since we haven't put any restrictions on the "height" of the cylinder there won't be any restriction on x.

In the first part of this example we used the fact that the function was in the form $x = f(y,z)$ to quickly write down a parametric representation. This can always be done for functions that are in this basic form.

Fact

$$\begin{aligned} z &= f(x,y) & \Rightarrow \quad & \vec{r}(x,y) = x\,\vec{i} + y\,\vec{j} + f(x,y)\,\vec{k} \\ x &= f(y,z) & \Rightarrow \quad & \vec{r}(y,z) = f(y,z)\,\vec{i} + y\,\vec{j} + z\,\vec{k} \\ y &= f(x,z) & \Rightarrow \quad & \vec{r}(x,z) = x\,\vec{i} + f(x,z)\,\vec{j} + z\,\vec{k} \end{aligned}$$

Okay, now that we have practice writing down some parametric representations for some surfaces let's take a quick look at a couple of applications.

Let's take a look at finding the tangent plane to the parametric surface S given by,

$$\vec{r}(u,v) = x(u,v)\,\vec{i} + y(u,v)\,\vec{j} + z(u,v)\,\vec{k}$$

First, define

$$\begin{aligned} \vec{r}_u(u,v) &= \frac{\partial x}{\partial u}(u,v)\,\vec{i} + \frac{\partial y}{\partial u}(u,v)\,\vec{j} + \frac{\partial z}{\partial u}(u,v)\,\vec{k} \\ \vec{r}_v(u,v) &= \frac{\partial x}{\partial v}(u,v)\,\vec{i} + \frac{\partial y}{\partial v}(u,v)\,\vec{j} + \frac{\partial z}{\partial v}(u,v)\,\vec{k} \end{aligned}$$

If we hold $v = v_0$ fixed then $\vec{r}_u(u,v_0)$ will be tangent to the curve given by $\vec{r}(u,v_0)$ (and yes this is a curve given that only one of the variables, u, is changing....) provided $\vec{r}_u(u,v_0) \ne \vec{0}$. Similarly, if we hold $u = u_0$ fixed then $\vec{r}_v(u_0,v)$ will be tangent to the curve given by $\vec{r}(u_0,v)$ (again, because only v is changing this is a curve) provided $\vec{r}_v(u_0,v) \ne \vec{0}$.

Therefore, both $\vec{r}_u(u_0,v_0)$ and $\vec{r}_v(u_0,v_0)$, provided neither one is the zero vector) will be tangent to the surface, S, given by $\vec{r}(u,v)$ at (u_0,v_0) and the tangent plane to the surface at (u_0,v_0) will be the plane containing both $\vec{r}_u(u_0,v_0)$ and $\vec{r}_v(u_0,v_0)$.

To help make things a little clearer we did the work at a particular point, but this fact is true at any point for which neither $\vec{r}_u$ or $\vec{r}_v$ are the zero vector.

This, in turn, means that provided $\vec{r}_u \times \vec{r}_v \neq \vec{0}$ the vector $\vec{r}_u \times \vec{r}_v$ will be orthogonal to the surface S and so it can be used for the normal vector that we need in order to write down the equation of a tangent plane. This is an important idea that will be used many times throughout the next couple of sections.

Let's take a look at an example.

Example 3

Find the equation of the tangent plane to the surface given by

$$\vec{r}(u,v) = u\,\vec{i} + 2v^2\,\vec{j} + \left(u^2 + v\right)\vec{k}$$

at the point $(2, 2, 3)$.

Solution

Let's first compute $\vec{r}_u \times \vec{r}_v$. Here are the two individual vectors.

$$\vec{r}_u(u,v) = \vec{i} + 2u\,\vec{k} \qquad\qquad \vec{r}_v(u,v) = 4v\,\vec{j} + \vec{k}$$

Now the cross product (which will give us the normal vector $\vec{n}$) is,

$$\vec{n} = \vec{r}_u \times \vec{r}_v = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & 0 & 2u \\ 0 & 4v & 1 \end{vmatrix} = -8uv\,\vec{i} - \vec{j} + 4v\,\vec{k}$$

Now, this is all fine, but in order to use it we will need to determine the value of u and v that will give us the point in question. We can easily do this by setting the individual components of the parametric representation equal to the coordinates of the point in question. Doing this gives,

$$\begin{aligned} 2 &= u & &\Rightarrow & u &= 2 \\ 2 &= 2v^2 & &\Rightarrow & v &= \pm 1 \\ 3 &= u^2 + v & & & & \end{aligned}$$

Now, as shown, we have the value of u, but there are two possible values of v. To determine the correct value of v let's plug u into the third equation and solve for v. This should tell us what the correct value is.

$$3 = 4 + v \qquad \Rightarrow \qquad v = -1$$

Okay so we now know that we'll be at the point in question when $u = 2$ and $v = -1$. At this point the normal vector is,

$$\vec{n} = 16\,\vec{i} - \vec{j} - 4\,\vec{k}$$

The tangent plane is then,

$$\begin{aligned} 16(x-2)-(y-2)-4(z-3)&=0 \\ 16x-y-4z&=18 \end{aligned}$$

You do remember how to write down the equation of a plane, right?

The second application that we want to take a quick look at is the surface area of the parametric surface S given by,

$$\vec{r}(u,v)=x(u,v)\,\vec{i}+y(u,v)\,\vec{j}+z(u,v)\,\vec{k}$$

and as we will see it again comes down to needing the vector $\vec{r}_u \times \vec{r}_v$.

So, provided S is traced out exactly once as (u,v) ranges over the points in D the surface area of S is given by,

Fact

$$A=\iint\limits_D \left\|\vec{r}_u \times \vec{r}_v\right\| \, dA$$

Let's take a look at an example.

Example 4

Find the surface area of the portion of the sphere of radius 4 that lies inside the cylinder $x^2+y^2=12$ and above the xy-plane.

Solution

Okay we've got a couple of things to do here. First, we need the parameterization of the sphere. We parameterized a sphere earlier in this section so there isn't too much to do at this point. Here is the parameterization.

$$\vec{r}(\theta,\varphi)=4\sin(\varphi)\cos(\theta)\,\vec{i}+4\sin(\varphi)\sin(\theta)\,\vec{j}+4\cos(\varphi)\,\vec{k}$$

Next, we need to determine D. Since we are not restricting how far around the z-axis we are rotating with the sphere we can take the following range for θ.

$$0 \le \theta \le 2\pi$$

Now, we need to determine a range for φ. This will take a little work, although it's not too bad. First, let's start with the equation of the sphere.

$$x^2 + y^2 + z^2 = 16$$

Now, if we substitute the equation for the cylinder into this equation we can find the value of z where the sphere and the cylinder intersect.

$$\begin{aligned} x^2 + y^2 + z^2 &= 16 \\ 12 + z^2 &= 16 \\ z^2 &= 4 \qquad \Rightarrow \qquad z = \pm 2 \end{aligned}$$

Now, since we also specified that we only want the portion of the sphere that lies above the xy-plane we know that we need $z = 2$. We also know that $\rho = 4$. Plugging this into the following conversion formula we get,

$$\begin{aligned} z &= \rho \cos(\varphi) \\ 2 &= 4\cos(\varphi) \\ \cos(\varphi) &= \frac{1}{2} \qquad \Rightarrow \qquad \varphi = \frac{\pi}{3} \end{aligned}$$

So, it looks like the range of φ will be,

$$0 \le \varphi \le \frac{\pi}{3}$$

Finally, we need to determine $\vec{r}_\theta \times \vec{r}_\varphi$. Here are the two individual vectors.

$$\begin{aligned} \vec{r}_\theta(\theta, \varphi) &= -4\sin(\varphi)\sin(\theta)\,\vec{i} + 4\sin(\varphi)\cos(\theta)\,\vec{j} \\ \vec{r}_\varphi(\theta, \varphi) &= 4\cos(\varphi)\cos(\theta)\,\vec{i} + 4\cos(\varphi)\sin(\theta)\,\vec{j} - 4\sin(\varphi)\,\vec{k} \end{aligned}$$

Now let's take the cross product.

$$\begin{aligned} \vec{r}_\theta \times \vec{r}_\varphi &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ -4\sin(\varphi)\sin(\theta) & 4\sin(\varphi)\cos(\theta) & 0 \\ 4\cos(\varphi)\cos(\theta) & 4\cos(\varphi)\sin(\theta) & -4\sin(\varphi) \end{vmatrix} \\ &= -16\sin^2(\varphi)\cos(\theta)\,\vec{i} - 16\sin(\varphi)\cos(\varphi)\sin^2(\theta)\,\vec{k} - 16\sin^2(\varphi)\sin(\theta)\,\vec{j} \\ &\qquad - 16\sin(\varphi)\cos(\varphi)\cos^2(\theta)\,\vec{k} \\ &= -16\sin^2(\varphi)\cos(\theta)\,\vec{i} - 16\sin^2(\varphi)\sin(\theta)\,\vec{j} - 16\sin(\varphi)\cos(\varphi)\left(\sin^2(\theta) + \cos^2(\theta)\right)\vec{k} \\ &= -16\sin^2(\varphi)\cos\theta\,\vec{i} - 16\sin^2(\varphi)\sin(\theta)\,\vec{j} - 16\sin(\varphi)\cos(\varphi)\,\vec{k} \end{aligned}$$

We now need the magnitude of this,

$$\begin{aligned}\|\vec{r}_\theta \times \vec{r}_\varphi\| &= \sqrt{256\sin^4(\varphi)\cos^2(\theta) + 256\sin^4(\varphi)\sin^2(\theta) + 256\sin^2(\varphi)\cos^2(\varphi)} \\ &= \sqrt{256\sin^4(\varphi)\left(\cos^2(\theta) + \sin^2(\theta)\right) + 256\sin^2(\varphi)\cos^2(\varphi)} \\ &= \sqrt{256\sin^2(\varphi)\left(\sin^2(\varphi) + \cos^2(\varphi)\right)} \\ &= 16\sqrt{\sin^2(\varphi)} \\ &= 16\left|\sin(\varphi)\right| \\ &= 16\sin(\varphi)\end{aligned}$$

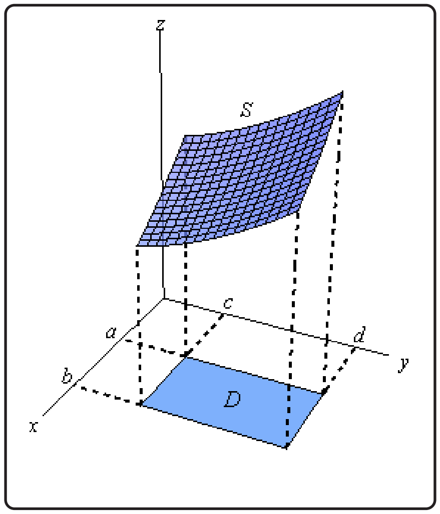
We can drop the absolute value bars in the sine because sine is positive in the range of φ that we are working with.

We can finally get the surface area.

$$\begin{aligned}A &= \iint\limits_D 16\sin(\varphi)\,dA \\ &= \int_0^{2\pi}\int_0^{\frac{\pi}{3}} 16\sin(\varphi)\,d\varphi\,d\theta \\ &= \int_0^{2\pi} -16\cos(\varphi)\Big|_0^{\pi/3}\,d\theta \\ &= \int_0^{2\pi} 8\,d\theta \\ &= 16\pi\end{aligned}$$

17.3 Surface Integrals

It is now time to think about integrating functions over some surface, S, in three-dimensional space. Let's start off with a sketch of the surface S since the notation can get a little confusing once we get into it. Here is a sketch of some surface S.



The region S will lie above (in this case) some region D that lies in the xy-plane. We used a rectangle here, but it doesn't have to be of course. Also note that we could just as easily looked at a surface S that was in front of some region D in the yz-plane or the xz-plane. Do not get so locked into the xy-plane that you can't do problems that have regions in the other two planes.

Now, how we evaluate the surface integral will depend upon how the surface is given to us. There are essentially two separate methods here, although as we will see they are really the same.

First, let's look at the surface integral in which the surface S is given by $z = g(x, y)$. In this case the surface integral is,

Surface Integral

$$\iint\limits_S f(x,y,z)\, dS = \iint\limits_D f(x,y,g(x,y))\sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + 1}\, dA$$

Now, we need to be careful here as both of these look like standard double integrals. In fact the integral on the right is a standard double integral. The integral on the left however is a surface integral. The way to tell them apart is by looking at the differentials. The surface integral will have a dS while the standard double integral will have a dA.

In order to evaluate a surface integral we will substitute the equation of the surface in for z in the integrand and then add on the often messy square root. After that the integral is a standard double integral and by this point we should be able to deal with that.

Note as well that there are similar formulas for surfaces given by $y = g(x, z)$ (with D in the xz-plane) and $x = g(y, z)$ (with D in the yz-plane). We will see one of these formulas in the examples and we'll leave the other to you to write down.

The second method for evaluating a surface integral is for those surfaces that are given by the parameterization,

$$\vec{r}(u, v) = x(u, v)\,\vec{i} + y(u, v)\,\vec{j} + z(u, v)\,\vec{k}$$

In these cases the surface integral is,

Surface Integral

$$\iint\limits_S f(x, y, z)\, dS = \iint\limits_D f(\vec{r}(u, v))\, \|\vec{r}_u \times \vec{r}_v\|\, dA$$

where D is the range of the parameters that trace out the surface S.

Before we work some examples let's notice that since we can parameterize a surface given by $z = g(x, y)$ as,

$$\vec{r}(x, y) = x\vec{i} + y\vec{j} + g(x, y)\,\vec{k}$$

we can always use this form for these kinds of surfaces as well. In fact, it can be shown that,

$$\|\vec{r}_x \times \vec{r}_y\| = \sqrt{\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 + 1}$$

for these kinds of surfaces. You might want to verify this for the practice of computing these cross products.

Let's work some examples.

Example 1

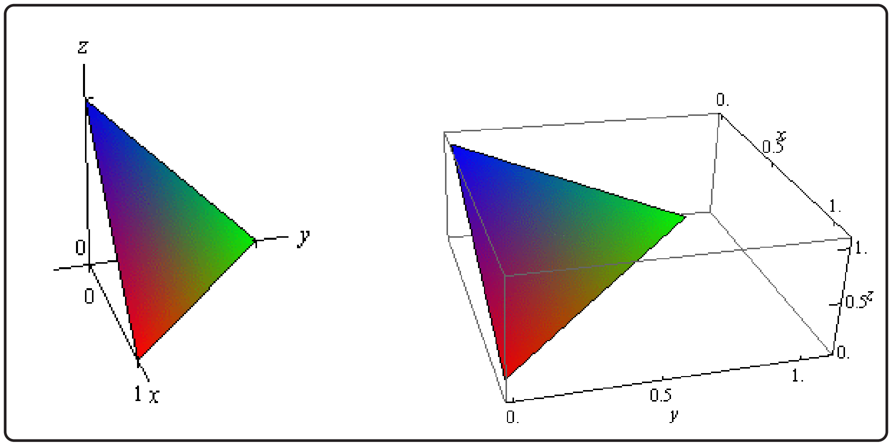
Evaluate $\displaystyle\iint\limits_S 6xy\, dS$ where S is the portion of the plane $x + y + z = 1$ that lies in the 1^{st} octant and is in front of the yz-plane.

Solution

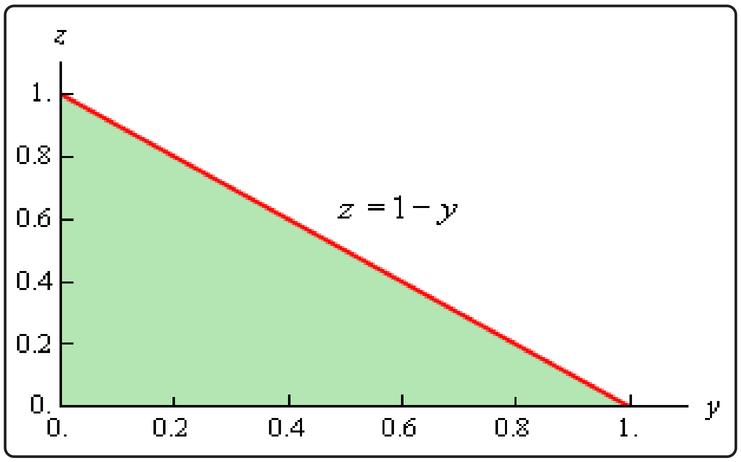
Okay, since we are looking for the portion of the plane that lies in front of the yz-plane we are going to need to write the equation of the surface in the form $x = g(y, z)$. This is easy enough to do.

$$x = 1 - y - z$$

Next, we need to determine just what D is. Here is a sketch of the surface S.



Here is a sketch of the region D.



Notice that the axes are labeled differently than we are used to seeing in the sketch of D. This was to keep the sketch consistent with the sketch of the surface. We arrived at the equation of the hypotenuse by setting x equal to zero in the equation of the plane and solving for z. Here are the ranges for y and z.

$$0 \le y \le 1 \qquad 0 \le z \le 1 - y$$

Now, because the surface is not in the form $z = g(x, y)$ we can't use the formula above. However, as noted above we can modify this formula to get one that will work for us. Here it is,

$$\iint\limits_S f(x, y, z)\, dS = \iint\limits_D f\left(g(y, z), y, z\right) \sqrt{1 + \left(\frac{\partial g}{\partial y}\right)^2 + \left(\frac{\partial g}{\partial z}\right)^2}\, dA$$

The changes made to the formula should be the somewhat obvious changes. So, let's do the integral.

$$\iint\limits_S 6xy\,dS = \iint\limits_D 6\,(1-y-z)\,y\sqrt{1+(-1)^2+(-1)^2}\,dA$$

Notice that we plugged in the equation of the plane for the *x* in the integrand. At this point we've got a fairly simple double integral to do. Here is that work.

$$\begin{aligned}\iint\limits_S 6xy\,dS &= \sqrt{3}\iint\limits_D 6\left(y-y^2-zy\right)\,dA \\ &= 6\sqrt{3}\int_0^1\int_0^{1-y} y-y^2-zy\,dz\,dy \\ &= 6\sqrt{3}\int_0^1 \left(yz-zy^2-\frac{1}{2}z^2y\right)\bigg|_0^{1-y}\,dy \\ &= 6\sqrt{3}\int_0^1 \frac{1}{2}y-y^2+\frac{1}{2}y^3\,dy \\ &= 6\sqrt{3}\left(\frac{1}{4}y^2-\frac{1}{3}y^3+\frac{1}{8}y^4\right)\bigg|_0^1 = \frac{\sqrt{3}}{4}\end{aligned}$$

Example 2

Evaluate $\iint\limits_S z\,dS$ where S is the upper half of a sphere of radius 2.

Solution

We gave the parameterization of a sphere in the previous section. Here is the parameterization for this sphere.

$$\vec{r}\,(\theta,\varphi) = 2\sin(\varphi)\cos(\theta)\,\vec{i} + 2\sin(\varphi)\sin(\theta)\,\vec{j} + 2\cos(\varphi)\,\vec{k}$$

Since we are working on the upper half of the sphere here are the limits on the parameters.

$$0 \le \theta \le 2\pi \qquad\qquad 0 \le \varphi \le \frac{\pi}{2}$$

Next, we need to determine $\vec{r}_\theta \times \vec{r}_\varphi$. Here are the two individual vectors.

$$\begin{aligned}\vec{r}_\theta\,(\theta,\varphi) &= -2\sin(\varphi)\sin(\theta)\,\vec{i} + 2\sin(\varphi)\cos(\theta)\,\vec{j} \\ \vec{r}_\varphi\,(\theta,\varphi) &= 2\cos(\varphi)\cos(\theta)\,\vec{i} + 2\cos(\varphi)\sin(\theta)\,\vec{j} - 2\sin(\varphi)\,\vec{k}\end{aligned}$$

Now let's take the cross product.

$$\begin{aligned}\vec{r}_\theta \times \vec{r}_\varphi &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ -2\sin(\varphi)\sin(\theta) & 2\sin(\varphi)\cos(\theta) & 0 \\ 2\cos(\varphi)\cos(\theta) & 2\cos(\varphi)\sin(\theta) & -2\sin(\varphi) \end{vmatrix} \\ &= -4\sin^2(\varphi)\cos(\theta)\,\vec{i} - 4\sin(\varphi)\cos(\varphi)\sin^2(\theta)\,\vec{k} - 4\sin^2(\varphi)\sin(\theta)\,\vec{j} - \\ &\qquad 4\sin(\varphi)\cos(\varphi)\cos^2(\theta)\,\vec{k} \\ &= -4\sin^2(\varphi)\cos(\theta)\,\vec{i} - 4\sin^2(\varphi)\sin(\theta)\,\vec{j} - 4\sin(\varphi)\cos(\varphi)\left(\sin^2(\theta) + \cos^2(\theta)\right)\vec{k} \\ &= -4\sin^2(\varphi)\cos(\theta)\,\vec{i} - 4\sin^2(\varphi)\sin(\theta)\,\vec{j} - 4\sin(\varphi)\cos(\varphi)\,\vec{k}\end{aligned}$$

Finally, we need the magnitude of this,

$$\begin{aligned}\|\vec{r}_\theta \times \vec{r}_\varphi\| &= \sqrt{16\sin^4(\varphi)\cos^2(\theta) + 16\sin^4(\varphi)\sin^2(\theta) + 16\sin^2(\varphi)\cos^2(\varphi)} \\ &= \sqrt{16\sin^4(\varphi)\left(\cos^2(\theta) + \sin^2(\theta)\right) + 16\sin^2(\varphi)\cos^2(\varphi)} \\ &= \sqrt{16\sin^2(\varphi)\left(\sin^2(\varphi) + \cos^2(\varphi)\right)} \\ &= 4\sqrt{\sin^2(\varphi)} \\ &= 4\left|\sin(\varphi)\right| \\ &= 4\sin(\varphi)\end{aligned}$$

We can drop the absolute value bars in the sine because sine is positive in the range of φ that we are working with. The surface integral is then,

$$\iint\limits_S z\,dS = \iint\limits_D 2\cos(\varphi)\left(4\sin(\varphi)\right)\,dA$$

Don't forget that we need to plug in for x, y and/or z in these as well, although in this case we just needed to plug in z. Here is the evaluation for the double integral.

$$\begin{aligned}\iint\limits_S z\,dS &= \int_0^{2\pi}\int_0^{\frac{\pi}{2}} 4\sin\left(2\varphi\right)\,d\varphi\,d\theta \\ &= \int_0^{2\pi} \left.\left(-2\cos\left(2\varphi\right)\right)\right|_0^{\frac{\pi}{2}}\,d\theta \\ &= \int_0^{2\pi} 4\,d\theta \\ &= 8\pi\end{aligned}$$

Example 3

Evaluate $\displaystyle\iint\limits_S y\,dS$ where S is the portion of the cylinder $x^2+y^2=3$ that lies between $z=0$ and $z=6$.

Solution

We parameterized up a cylinder in the previous section. Here is the parameterization of this cylinder.

$$\vec{r}\,(z,\theta)=\sqrt{3}\cos(\theta)\,\vec{i}+\sqrt{3}\sin(\theta)\,\vec{j}+z\,\vec{k}$$

The ranges of the parameters are,

$$0\le z\le 6 \qquad 0\le\theta\le 2\pi$$

Now we need $\vec{r}_z\times\vec{r}_\theta$. Here are the two vectors.

$$\begin{aligned}\vec{r}_z\,(z,\theta)&=\vec{k}\\ \vec{r}_\theta\,(z,\theta)&=-\sqrt{3}\sin(\theta)\,\vec{i}+\sqrt{3}\cos(\theta)\,\vec{j}\end{aligned}$$

Here is the cross product.

$$\begin{aligned}\vec{r}_z\times\vec{r}_\theta&=\begin{vmatrix}\vec{i} & \vec{j} & \vec{k}\\ 0 & 0 & 1\\ -\sqrt{3}\sin(\theta) & \sqrt{3}\cos(\theta) & 0\end{vmatrix}\\ &=-\sqrt{3}\cos(\theta)\,\vec{i}-\sqrt{3}\sin(\theta)\,\vec{j}\end{aligned}$$

The magnitude of this vector is,

$$\|\vec{r}_z\times\vec{r}_\theta\|=\sqrt{3\cos^2(\theta)+3\sin^2(\theta)}=\sqrt{3}$$

The surface integral is then,

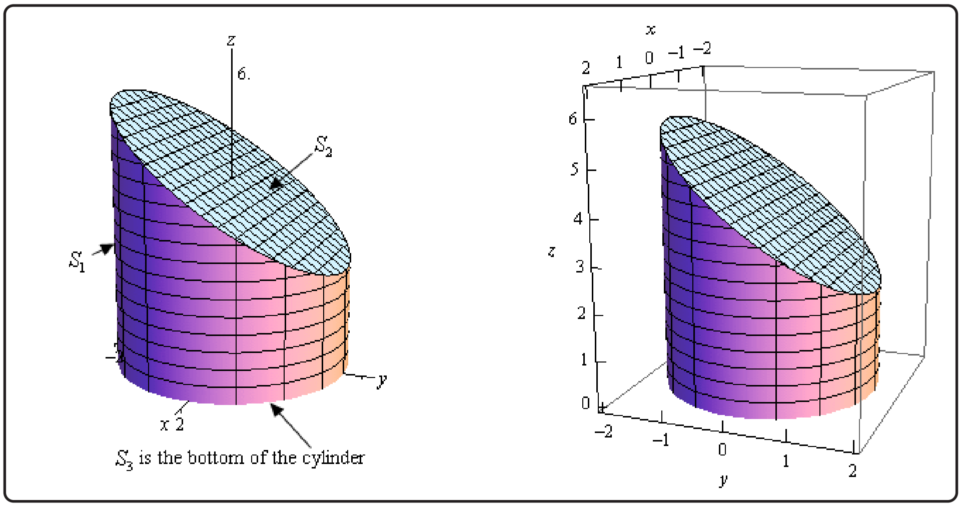
$$\begin{aligned}\iint\limits_S y\,dS&=\iint\limits_D \sqrt{3}\sin(\theta)\left(\sqrt{3}\right)\,dA\\ &=3\int_0^{2\pi}\int_0^6\sin(\theta)\,dz\,d\theta\\ &=3\int_0^{2\pi}6\sin(\theta)\,d\theta=-18\cos(\theta)\Big|_0^{2\pi}=0\end{aligned}$$

Example 4

Evaluate $\displaystyle\iint\limits_S y + z\,dS$ where S is the surface whose side is the cylinder $x^2 + y^2 = 3$, whose bottom is the disk $x^2 + y^2 \le 3$ in the xy-plane and whose top is the plane $z = 4 - y$.

Solution

There is a lot of information that we need to keep track of here. First, we are using pretty much the same surface (the integrand is different however) as the previous example. However, unlike the previous example we are putting a top and bottom on the surface this time. Let's first start out with a sketch of the surface.



We need to be careful here. There is more to this sketch than the actual surface itself. We're going to let S_1 be the portion of the cylinder that goes from the xy-plane to the plane. In other words, the top of the cylinder will be at an angle. We'll call the portion of the plane that lies inside (*i.e.* the cap on the cylinder) S_2. Finally, the bottom of the cylinder (not shown here) is the disk of radius $\sqrt{3}$ in the xy-plane and is denoted by S_3.

In order to do this integral we'll need to note that just like the standard double integral, if the surface is split up into pieces we can also split up the surface integral. So, for our example we will have,

$$\iint\limits_S y + z\,dS = \iint\limits_{S_1} y + z\,dS + \iint\limits_{S_2} y + z\,dS + \iint\limits_{S_3} y + z\,dS$$

We're going to need to do three integrals here. However, we've done most of the work for the first one in the previous example so let's start with that.

S_1 : The Cylinder

The parameterization of the cylinder and $\|\vec{r}_z \times \vec{r}_\theta\|$ is,

$$\vec{r}(z,\theta) = \sqrt{3}\cos(\theta)\,\vec{i} + \sqrt{3}\sin(\theta)\,\vec{j} + z\,\vec{k} \qquad\qquad \|\vec{r}_z \times \vec{r}_\theta\| = \sqrt{3}$$

The difference between this problem and the previous one is the limits on the parameters. Here they are.

$$0 \le \theta \le 2\pi$$
$$0 \le z \le 4 - y = 4 - \sqrt{3}\sin(\theta)$$

The upper limit for the z's is the plane so we can just plug that in. However, since we are on the cylinder we know what y is from the parameterization so we will also need to plug that in.

Here is the integral for the cylinder.

$$\begin{aligned}
\iint\limits_{S_1} y + z\,dS &= \iint\limits_{D} \left(\sqrt{3}\sin(\theta) + z\right)\left(\sqrt{3}\right)\,dA \\
&= \sqrt{3}\int_0^{2\pi}\int_0^{4-\sqrt{3}\sin(\theta)} \sqrt{3}\sin(\theta) + z\,dz\,d\theta \\
&= \sqrt{3}\int_0^{2\pi} \sqrt{3}\sin(\theta)\left(4 - \sqrt{3}\sin(\theta)\right) + \frac{1}{2}\left(4 - \sqrt{3}\sin(\theta)\right)^2\,d\theta \\
&= \sqrt{3}\int_0^{2\pi} 8 - \frac{3}{2}\sin^2(\theta)\,d\theta \\
&= \sqrt{3}\int_0^{2\pi} 8 - \frac{3}{4}\left(1 - \cos(2\theta)\right)d\theta \\
&= \sqrt{3}\left(\frac{29}{4}\theta + \frac{3}{8}\sin(2\theta)\right)\Big|_0^{2\pi} \\
&= \frac{29\sqrt{3}\,\pi}{2}
\end{aligned}$$

S_2 : Plane on Top of the Cylinder

In this case we don't need to do any parameterization since it is set up to use the formula that we gave at the start of this section. Remember that the plane is given by $z = 4 - y$. Also note that, for this surface, D is the disk of radius $\sqrt{3}$ centered at the origin.

Here is the integral for the plane.

$$\begin{aligned}
\iint\limits_{S_2} y + z\,dS &= \iint\limits_{D} (y + 4 - y)\sqrt{(0)^2 + (-1)^2 + 1}\,dA \\
&= \sqrt{2}\iint\limits_{D} 4\,dA
\end{aligned}$$

Don't forget that we need to plug in for z! Now at this point we can proceed in one of two ways. Either we can proceed with the integral or we can recall that $\iint\limits_D dA$ is nothing more than the area of D and we know that D is the disk of radius $\sqrt{3}$ and so there is no reason to do the integral.

Here is the remainder of the work for this problem.

$$\begin{aligned}\iint\limits_{S_2} y + z\,dS &= 4\sqrt{2}\iint\limits_D dA \\ &= 4\sqrt{2}\left(\pi\left(\sqrt{3}\right)^2\right) \\ &= 12\sqrt{2}\,\pi\end{aligned}$$

S_3 : Bottom of the Cylinder

Again, this is set up to use the initial formula we gave in this section once we realize that the equation for the bottom is given by $g(x, y) = 0$ and D is the disk of radius $\sqrt{3}$ centered at the origin. Also, don't forget to plug in for z.

Here is the work for this integral.

$$\begin{aligned}\iint\limits_{S_3} y + z\,dS &= \iint\limits_D (y + 0)\sqrt{(0)^2 + (0)^2 + 1}\,dA \\ &= \iint\limits_D y\,dA \\ &= \int_0^{2\pi}\int_0^{\sqrt{3}} r^2 \sin(\theta)\,dr\,d\theta \\ &= \int_0^{2\pi}\left(\frac{1}{3}r^3\sin(\theta)\right)\Bigg|_0^{\sqrt{3}} d\theta \\ &= \int_0^{2\pi}\sqrt{3}\sin(\theta)\,d\theta \\ &= -\sqrt{3}\cos(\theta)\Big|_0^{2\pi} \\ &= 0\end{aligned}$$

We can now get the value of the integral that we are after.

$$\begin{aligned}\iint\limits_S y + z\,dS &= \iint\limits_{S_1} y + z\,dS + \iint\limits_{S_2} y + z\,dS + \iint\limits_{S_3} y + z\,dS \\ &= \frac{29\sqrt{3}\,\pi}{2} + 12\sqrt{2}\,\pi + 0 \\ &= \frac{\pi}{2}\left(29\sqrt{3} + 24\sqrt{2}\right)\end{aligned}$$

17.4 Surface Integrals of Vector Fields

Just as we did with line integrals we now need to move on to surface integrals of vector fields. Recall that in line integrals the orientation of the curve we were integrating along could change the answer. The same thing will hold true with surface integrals. So, before we really get into doing surface integrals of vector fields we first need to introduce the idea of an **oriented surface**.

Let's start off with a surface that has two sides (while this may seem strange, recall that the Mobius Strip is a surface that only has one side!) that has a tangent plane at every point (except possibly along the boundary). Making this assumption means that every point will have two unit normal vectors, $\vec{n}_1$ and $\vec{n}_2 = -\vec{n}_1$. This means that every surface will have two sets of normal vectors. The set that we choose will give the surface an orientation.

There is one convention that we will make in regard to certain kinds of oriented surfaces. First, we need to define a **closed surface**. A surface S is closed if it is the boundary of some solid region E. A good example of a closed surface is the surface of a sphere. We say that the closed surface S has a **positive** orientation if we choose the set of unit normal vectors that point outward from the region E while the **negative** orientation will be the set of unit normal vectors that point in towards the region E.

Note that this convention is only used for closed surfaces.

In order to work with surface integrals of vector fields we will need to be able to write down a formula for the unit normal vector corresponding to the orientation that we've chosen to work with. We have two ways of doing this depending on how the surface has been given to us.

First, let's suppose that the function is given by $z = g(x, y)$. In this case we first define a new function,

$$f(x, y, z) = z - g(x, y)$$

In terms of our new function the surface is then given by the equation $f(x, y, z) = 0$. Now, recall that ∇f will be orthogonal (or normal) to the surface given by $f(x, y, z) = 0$. This means that we have a normal vector to the surface. The only potential problem is that it might not be a unit normal vector. That isn't a problem since we also know that we can turn any vector into a unit vector by dividing the vector by its length. In our case this is,

$$\vec{n} = \frac{\nabla f}{\|\nabla f\|}$$

In this case it will be convenient to actually compute the gradient vector and plug this into the formula for the normal vector. Doing this gives,

$$\vec{n} = \frac{\nabla f}{\|\nabla f\|} = \frac{-g_x\,\vec{i} - g_y\,\vec{j} + \vec{k}}{\sqrt{(g_x)^2 + (g_y)^2 + 1}}$$

Now, from a notational standpoint this might not have been so convenient, but it does allow us to make a couple of additional comments.

First, notice that the component of the normal vector in the z-direction (identified by the $\vec{k}$ in the normal vector) is always positive and so this normal vector will generally point upwards. It may not point directly up, but it will have an upwards component to it.

This will be important when we are working with a closed surface and we want the positive orientation. If we know that we can then look at the normal vector and determine if the "positive" orientation should point upwards or downwards. Remember that the "positive" orientation must point out of the region and this may mean downwards in places. Of course, if it turns out that we need the downward orientation we can always take the negative of this unit vector and we'll get the one that we need. Again, remember that we always have that option when choosing the unit normal vector.

Before we move onto the second method of giving the surface we should point out that we only did this for surfaces in the form $z = g(x, y)$. We could just as easily done the above work for surfaces in the form $y = g(x, z)$ (so $f(x, y, z) = y - g(x, z)$) or for surfaces in the form $x = g(y, z)$ (so $f(x, y, z) = x - g(y, z)$).

Now, we need to discuss how to find the unit normal vector if the surface is given parametrically as,

$$\vec{r}(u, v) = x(u, v)\,\vec{i} + y(u, v)\,\vec{j} + z(u, v)\,\vec{k}$$

In this case recall that the vector $\vec{r}_u \times \vec{r}_v$ will be normal to the tangent plane at a particular point. But if the vector is normal to the tangent plane at a point then it will also be normal to the surface at that point. So, this is a normal vector. In order to guarantee that it is a unit normal vector we will also need to divide it by its magnitude.

So, in the case of parametric surfaces one of the unit normal vectors will be,

$$\vec{n} = \frac{\vec{r}_u \times \vec{r}_v}{\|\vec{r}_u \times \vec{r}_v\|}$$

As with the first case we will need to look at this once it's computed and determine if it points in the correct direction or not. If it doesn't then we can always take the negative of this vector and that will point in the correct direction.

Finally, remember that we can always parameterize any surface given by $z = g(x, y)$ (or $y = g(x, z)$ or $x = g(y, z)$) easily enough and so if we want to we can always use the parameterization formula to find the unit normal vector.

Okay, now that we've looked at oriented surfaces and their associated unit normal vectors we can actually give a formula for evaluating surface integrals of vector fields.

Given a vector field $\vec{F}$ with unit normal vector $\vec{n}$ then the surface integral of $\vec{F}$ over the surface S is given by,

Surface Integral

$$\iint\limits_S \vec{F} \cdot d\vec{S} = \iint\limits_S \vec{F} \cdot \vec{n}\, dS$$

where the right hand integral is a standard surface integral. This is sometimes called the **flux of $\vec{F}$ across S.**

Before we work any examples let's notice that we can substitute in for the unit normal vector to get a *somewhat* easier formula to use. We will need to be careful with each of the following formulas however as each will assume a certain orientation and we may have to change the normal vector to match the given orientation.

Let's first start by assuming that the surface is given by $z = g(x, y)$. In this case let's also assume that the vector field is given by $\vec{F} = P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}$ and that the orientation that we are after is the "upwards" orientation. Under all of these assumptions the surface integral of $\vec{F}$ over S is,

$$\begin{aligned}
\iint\limits_S \vec{F} \cdot d\vec{S} &= \iint\limits_S \vec{F} \cdot \vec{n}\,dS \\
&= \iint\limits_D \left(P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}\right) \cdot \left(\frac{-g_x\,\vec{i} - g_y\,\vec{j} + \vec{k}}{\sqrt{(g_x)^2 + (g_y)^2 + 1}}\right)\sqrt{(g_x)^2 + (g_y)^2 + 1}\,dA \\
&= \iint\limits_D \left(P\,\vec{i} + Q\,\vec{j} + R\,\vec{k}\right) \cdot \left(-g_x\,\vec{i} - g_y\,\vec{j} + \vec{k}\right) dA \\
&= \iint\limits_D -Pg_x - Qg_y + R\,dA
\end{aligned}$$

Now, remember that this assumed the "upward" orientation. If we'd needed the "downward" orientation, then we would need to change the signs on the normal vector. This would in turn change the signs on the integrand as well. So, we really need to be careful here when using this formula. In general, it is best to rederive this formula as you need it.

When we've been given a surface that is not in parametric form there are in fact 6 possible integrals here. Two for each form of the surface $z = g(x, y)$, $y = g(x, z)$ and $x = g(y, z)$. Given each form of the surface there will be two possible unit normal vectors and we'll need to choose the correct one to match the given orientation of the surface. However, the derivation of each formula is similar to that given here and so shouldn't be too bad to do as you need to.

Notice as well that because we are using the unit normal vector the messy square root will always drop out. This means that when we do need to derive the formula we won't really need to put this in. All we'll need to work with is the numerator of the unit vector. We will see at least one more of these derived in the examples below. It should also be noted that the square root is nothing more than,

$$\sqrt{(g_x)^2 + (g_y)^2 + 1} = \|\nabla f\|$$

so in the following work we will probably just use this notation in place of the square root when we can to make things a little simpler.

Let's now take a quick look at the formula for the surface integral when the surface is given parametrically by $\vec{r}(u,v)$. In this case the surface integral is,

$$\begin{aligned}\iint\limits_S \vec{F}\cdot d\vec{S} &= \iint\limits_S \vec{F}\cdot\vec{n}\,dS\\ &= \iint\limits_D \vec{F}\cdot\left(\frac{\vec{r}_u\times\vec{r}_v}{\|\vec{r}_u\times\vec{r}_v\|}\right)\|\vec{r}_u\times\vec{r}_v\|\,dA\\ &= \iint\limits_D \vec{F}\cdot(\vec{r}_u\times\vec{r}_v)\,dA\end{aligned}$$

Again, note that we may have to change the sign on $\vec{r}_u\times\vec{r}_v$ to match the orientation of the surface and so there is once again really two formulas here. Also note that again the magnitude cancels in this case and so we won't need to worry that in these problems either.

Note as well that there are even times when we will use the definition, $\iint\limits_S \vec{F}\cdot d\vec{S} = \iint\limits_S \vec{F}\cdot\vec{n}\,dS$, directly. We will see an example of this below.

Let's now work a couple of examples.

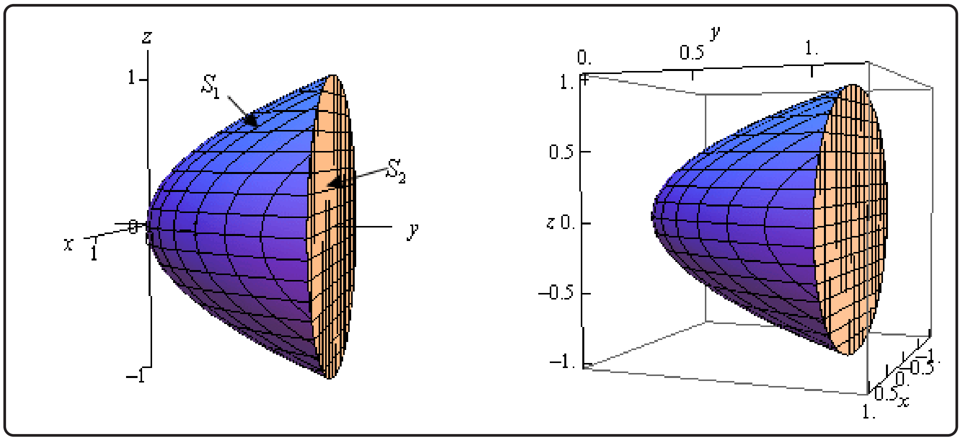
Example 1

Evaluate $\iint\limits_S \vec{F}\cdot d\vec{S}$ where $\vec{F} = y\,\vec{j} - z\,\vec{k}$ and S is the surface given by the paraboloid $y = x^2+z^2$, $0\le y\le 1$ and the disk $x^2+z^2\le 1$ at $y=1$. Assume that S has positive orientation.

Solution

Okay, first let's notice that the disk is really nothing more than the cap on the paraboloid. This means that we have a closed surface. This is important because we've been told that the surface has a positive orientation and by convention this means that all the unit normal vectors will need to point outwards from the region enclosed by S.

Let's first get a sketch of S so we can get a feel for what is going on and in which direction we will need to unit normal vectors to point.



As noted in the sketch we will denote the paraboloid by S_1 and the disk by S_2. Also note that in order for unit normal vectors on the paraboloid to point away from the region they will all need to point generally in the negative y direction. On the other hand, unit normal vectors on the disk will need to point in the positive y direction in order to point away from the region.

Since S is composed of the two surfaces we'll need to do the surface integral on each and then add the results to get the overall surface integral. Let's start with the paraboloid. In this case we have the surface in the form $y = g(x, z)$ so we will need to derive the correct formula since the one given initially wasn't for this kind of function. This is easy enough to do however. First define,

$$f(x, y, z) = y - g(x, z) = y - x^2 - z^2$$

We will next need the gradient vector of this function.

$$\nabla f = \langle -2x, 1, -2z \rangle$$

Now, the y component of the gradient is positive and so this vector will generally point in the positive y direction. However, as noted above we need the normal vector point in the negative y direction to make sure that it will be pointing away from the enclosed region. This means that we will need to use

$$\vec{n} = \frac{-\nabla f}{\|-\nabla f\|} = \frac{\langle 2x, -1, 2z \rangle}{\|\nabla f\|}$$

Let's note a couple of things here before we proceed. We don't really need to divide this by the magnitude of the gradient since this will just cancel out once we actually do the integral. So, because of this we didn't bother computing it. Also, the dropping of the minus sign is not a typo. When we compute the magnitude we are going to square each of the components and so the minus sign will drop out.

S_1 : The Paraboloid

Okay, here is the surface integral in this case.

$$\begin{aligned}\iint\limits_{S_1} \vec{F}\cdot d\vec{S} &= \iint\limits_{D} \left(y\,\vec{j} - z\,\vec{k}\right)\cdot\left(\frac{\langle 2x,-1,2z\rangle}{\|\nabla f\|}\right)\|\nabla f\|\,dA\\ &= \iint\limits_{D} -y-2z^2\,dA\\ &= \iint\limits_{D} -\left(x^2+z^2\right)-2z^2\,dA\\ &= -\iint\limits_{D} x^2+3z^2\,dA\end{aligned}$$

Don't forget that we need to plug in the equation of the surface for y before we actually compute the integral. In this case D is the disk of radius 1 in the xz-plane and so it makes sense to use polar coordinates to complete this integral. Here are polar coordinates for this region.

$$\begin{array}{ll} x = r\cos(\theta) & z = r\sin(\theta)\\ 0\le\theta\le 2\pi & 0\le r\le 1\end{array}$$

Note that we kept the x conversion formula the same as the one we are used to using for x and let z be the formula that used the sine. We could have done it any order, however in this way we are at least working with one of them as we are used to working with.

Here is the evaluation of this integral.

$$\begin{aligned}\iint\limits_{S_1} \vec{F}\cdot d\vec{S} &= -\iint\limits_{D} x^2+3z^2\,dA\\ &= -\int_0^{2\pi}\int_0^1 \left(r^2\cos^2(\theta)+3r^2\sin^2(\theta)\right)r\,dr\,d\theta\\ &= -\int_0^{2\pi}\int_0^1 \left(\cos^2(\theta)+3\sin^2(\theta)\right)r^3\,dr\,d\theta\\ &= -\int_0^{2\pi}\left(\frac{1}{2}\left(1+\cos\left(2\theta\right)\right)+\frac{3}{2}\left(1-\cos\left(2\theta\right)\right)\right)\left.\left(\frac{1}{4}r^4\right)\right|_0^1 d\theta\\ &= -\frac{1}{8}\int_0^{2\pi} 4-2\cos\left(2\theta\right)\,d\theta\\ &= -\left.\frac{1}{8}\left(4\theta-\sin\left(2\theta\right)\right)\right|_0^{2\pi}\\ &= -\pi\end{aligned}$$

S_2 : The Cap of the Paraboloid

We can now do the surface integral on the disk (cap on the paraboloid). This one is actually fairly easy to do and in fact we can use the definition of the surface integral directly. First let's notice that the disk is really just the portion of the plane $y = 1$ that is in front of the disk of radius 1 in the xz-plane.

Now we want the unit normal vector to point away from the enclosed region and since it must also be orthogonal to the plane $y = 1$ then it must point in a direction that is parallel to the y-axis, but we already have a unit vector that does this. Namely,

$$\vec{n} = \vec{j}$$

the standard unit basis vector. It also points in the correct direction for us to use. Because we have the vector field and the normal vector we can plug directly into the definition of the surface integral to get,

$$\iint\limits_{S_2} \vec{F} \cdot d\vec{S} = \iint\limits_{S_2} \left(y\,\vec{j} - z\,\vec{k}\right) \cdot \left(\vec{j}\right)\, dS = \iint\limits_{S_2} y\, dS$$

At this point we need to plug in for y (since S_2 is a portion of the plane $y = 1$ we do know what it is) and we'll also need the square root this time when we convert the surface integral over to a double integral. In this case since we are using the definition directly we won't get the canceling of the square root that we saw with the first portion. To get the square root well need to acknowledge that

$$y = 1 = g(x, z)$$

and so the square root is,

$$\sqrt{\left(g_x\right)^2 + 1 + \left(g_z\right)^2}$$

The surface integral is then,

$$\begin{aligned}\iint\limits_{S_2} \vec{F} \cdot d\vec{S} &= \iint\limits_{S_2} y\, dS \\ &= \iint\limits_{D} 1\sqrt{0 + 1 + 0}\, dA = \iint\limits_{D} dA\end{aligned}$$

At this point we can acknowledge that D is a disk of radius 1 and this double integral is nothing more than the double integral that will give the area of the region D so there is no reason to compute the integral. Here is the value of the surface integral.

$$\iint\limits_{S_2} \vec{F} \cdot d\vec{S} = \pi$$

Finally, to finish this off we just need to add the two parts up. Here is the surface integral that we were actually asked to compute.

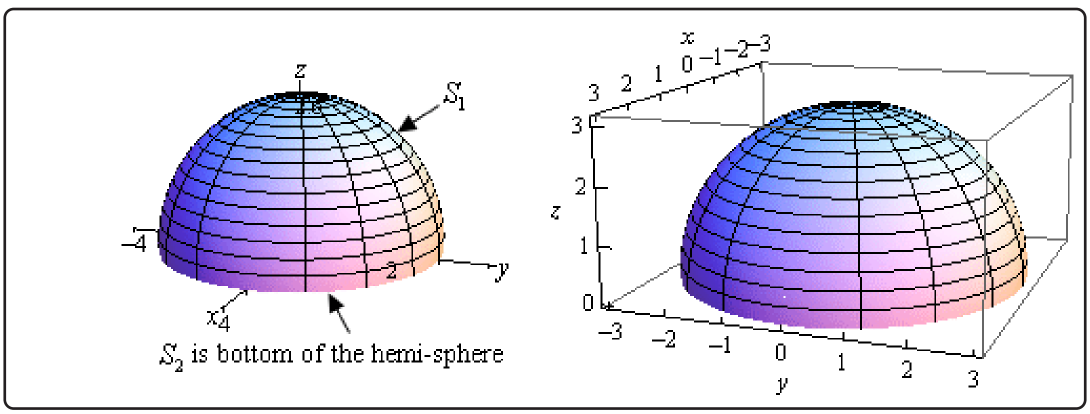
$$\iint\limits_{S} \vec{F} \cdot d\vec{S} = \iint\limits_{S_1} \vec{F} \cdot d\vec{S} + \iint\limits_{S_2} \vec{F} \cdot d\vec{S} = -\pi + \pi = 0$$

Example 2

Evaluate $\displaystyle\iint\limits_{S} \vec{F} \cdot d\vec{S}$ where $\vec{F} = x\,\vec{i} + y\,\vec{j} + z^4\,\vec{k}$ and S is the upper half the sphere $x^2 + y^2 + z^2 = 9$ and the disk $x^2 + y^2 \le 9$ in the plane $z = 0$. Assume that S has the positive orientation.

Solution

So, as with the previous problem we have a closed surface and since we are also told that the surface has a positive orientation all the unit normal vectors must point away from the enclosed region. To help us visualize this here is a sketch of the surface.



We will call S_1 the hemisphere and S_2 will be the bottom of the hemisphere (which isn't shown on the sketch). Now, in order for the unit normal vectors on the sphere to point away from enclosed region they will all need to have a positive z component. Remember that the vector must be normal to the surface and if there is a positive z component and the vector is normal it will have to be pointing away from the enclosed region.

On the other hand, the unit normal on the bottom of the disk must point in the negative z direction in order to point away from the enclosed region.

S_1 : The Sphere

Let's do the surface integral on S_1 first. In this case since the surface is a sphere we will need to use the parametric representation of the surface. This is,

$$\vec{r}(\theta,\varphi) = 3\sin(\varphi)\cos(\theta)\,\vec{i} + 3\sin(\varphi)\sin(\theta)\,\vec{j} + 3\cos(\varphi)\,\vec{k}$$

Since we are working on the hemisphere here are the limits on the parameters that we'll need to use.

$$0 \le \theta \le 2\pi \qquad\qquad 0 \le \varphi \le \frac{\pi}{2}$$

Next, we need to determine $\vec{r}_\theta \times \vec{r}_\varphi$. Here are the two individual vectors and the cross product.

$$\begin{aligned}\vec{r}_\theta(\theta,\varphi) &= -3\sin(\varphi)\sin(\theta)\,\vec{i} + 3\sin(\varphi)\cos(\theta)\,\vec{j}\\ \vec{r}_\varphi(\theta,\varphi) &= 3\cos(\varphi)\cos(\theta)\,\vec{i} + 3\cos(\varphi)\sin(\theta)\,\vec{j} - 3\sin(\varphi)\,\vec{k}\end{aligned}$$

$$\begin{aligned}\vec{r}_\theta \times \vec{r}_\varphi &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ -3\sin(\varphi)\sin(\theta) & 3\sin(\varphi)\cos(\theta) & 0 \\ 3\cos(\varphi)\cos(\theta) & 3\cos(\varphi)\sin(\theta) & -3\sin(\varphi) \end{vmatrix}\\ &= -9\sin^2(\varphi)\cos(\theta)\,\vec{i} - 9\sin(\varphi)\cos(\varphi)\sin^2(\theta)\,\vec{k} - 9\sin^2(\varphi)\sin(\theta)\,\vec{j}\\ &\qquad - 9\sin(\varphi)\cos(\varphi)\cos^2(\theta)\,\vec{k}\\ &= -9\sin^2(\varphi)\cos(\theta)\,\vec{i} - 9\sin^2(\varphi)\sin(\theta)\,\vec{j} - 9\sin(\varphi)\cos(\varphi)\left(\sin^2(\theta) + \cos^2(\theta)\right)\vec{k}\\ &= -9\sin^2(\varphi)\cos(\theta)\,\vec{i} - 9\sin^2(\varphi)\sin(\theta)\,\vec{j} - 9\sin(\varphi)\cos(\varphi)\,\vec{k}\end{aligned}$$

Note that we won't need the magnitude of the cross product since that will cancel out once we start doing the integral.

Notice that for the range of φ that we've got both sine and cosine are positive and so this vector will have a negative z component and as we noted above in order for this to point away from the enclosed area we will need the z component to be positive. Therefore, we will need to use the following vector for the unit normal vector.

$$\vec{n} = -\frac{\vec{r}_\theta \times \vec{r}_\varphi}{\|\vec{r}_\theta \times \vec{r}_\varphi\|} = \frac{9\sin^2(\varphi)\cos(\theta)\,\vec{i} + 9\sin^2(\varphi)\sin(\theta)\,\vec{j} + 9\sin(\varphi)\cos(\varphi)\,\vec{k}}{\|\vec{r}_\theta \times \vec{r}_\varphi\|}$$

Again, we will drop the magnitude once we get to actually doing the integral since it will just cancel in the integral.

Okay, next we'll need

$$\vec{F}(\vec{r}(\theta,\varphi)) = 3\sin(\varphi)\cos(\theta)\,\vec{i} + 3\sin(\varphi)\sin(\theta)\,\vec{j} + 81\cos^4(\varphi)\,\vec{k}$$

Remember that in this evaluation we are just plugging in the x component of $\vec{r}(\theta,\varphi)$ into the vector field *etc*.

We also may as well get the dot product out of the way that we know we are going to need.

$$\begin{aligned}\vec{F}(\vec{r}(\theta,\varphi))\cdot(-\vec{r}_\theta\times\vec{r}_\varphi) &= 27\sin^3(\varphi)\cos^2(\theta)+27\sin^3(\varphi)\sin^2(\theta)+729\sin(\varphi)\cos^5(\varphi)\\ &= 27\sin^3(\varphi)+729\sin(\varphi)\cos^5(\varphi)\end{aligned}$$

Now we can do the integral.

$$\begin{aligned}\iint\limits_{S_1}\vec{F}\cdot d\vec{S} &= \iint\limits_{D}\vec{F}\cdot\left(\frac{\vec{r}_\theta\times\vec{r}_\varphi}{\|\vec{r}_\theta\times\vec{r}_\varphi\|}\right)\|\vec{r}_\theta\times\vec{r}_\varphi\|\;dA\\ &= \int_0^{2\pi}\int_0^{\frac{\pi}{2}} 27\sin^3(\varphi)+729\sin(\varphi)\cos^5(\varphi)\,d\varphi\,d\theta\\ &= \int_0^{2\pi}\int_0^{\frac{\pi}{2}} 27\sin(\varphi)\left(1-\cos^2(\varphi)\right)+729\sin(\varphi)\cos^5(\varphi)\,d\varphi\,d\theta\\ &= -\int_0^{2\pi}\left(27\left(\cos(\varphi)-\frac{1}{3}\cos^3(\varphi)\right)+\frac{729}{6}\cos^6(\varphi)\right)\Bigg|_0^{\frac{\pi}{2}}\,d\theta\\ &= \int_0^{2\pi}\frac{279}{2}\,d\theta\\ &= 279\pi\end{aligned}$$

S_2 : The Bottom of the Hemi-Sphere

Now, we need to do the integral over the bottom of the hemisphere. In this case we are looking at the disk $x^2+y^2\le 9$ that lies in the plane $z=0$ and so the equation of this surface is actually $z=0$. The disk is really the region D that tells us how much of the surface we are going to use. This also means that we can use the definition of the surface integral here with

$$\vec{n}=-\vec{k}$$

We need the negative since it must point away from the enclosed region.

The surface integral in this case is,

$$\begin{aligned}\iint\limits_{S_2}\vec{F}\cdot d\vec{S} &= \iint\limits_{S_2}\left(x\,\vec{i}+y\,\vec{j}+z^4\,\vec{k}\right)\cdot\left(-\vec{k}\right)\,dS\\ &= \iint\limits_{S_2}-z^4\,dS\end{aligned}$$

Remember, however, that we are in the plane given by $z=0$ and so the surface integral becomes,

$$\iint\limits_{S_2}\vec{F}\cdot d\vec{S}=\iint\limits_{S_2}-z^4\,dS=\iint\limits_{S_2}0\,dS=0$$

The last step is to then add the two pieces up. Here is surface integral that we were asked to look at.

$$\iint\limits_{S} \vec{F} \cdot d\vec{S} = \iint\limits_{S_1} \vec{F} \cdot d\vec{S} + \iint\limits_{S_2} \vec{F} \cdot d\vec{S} = 279\pi + 0 = 279\pi$$

We will leave this section with a quick interpretation of a surface integral over a vector field. If $\vec{v}$ is the velocity field of a fluid then the surface integral

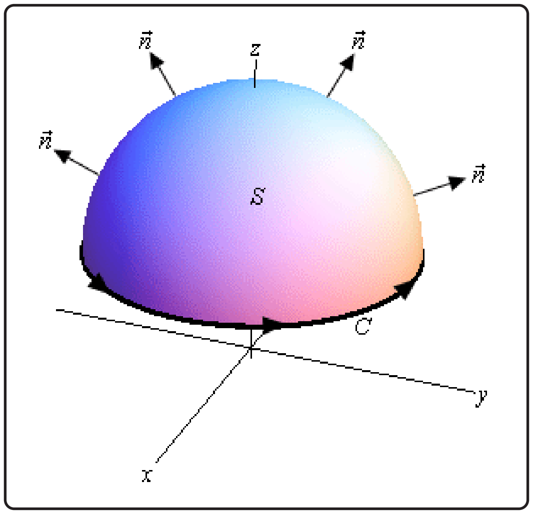
$$\iint\limits_{S} \vec{v} \cdot d\vec{S}$$

represents the volume of fluid flowing through S per time unit (*i.e.* per second, per minute, or whatever time unit you are using).

17.5 Stokes' Theorem

In this section we are going to take a look at a theorem that is a higher dimensional version of Green's Theorem. In Green's Theorem we related a line integral to a double integral over some region. In this section we are going to relate a line integral to a surface integral. However, before we give the theorem we first need to define the curve that we're going to use in the line integral.

Let's start off with the following surface with the indicated orientation.



Around the edge of this surface we have a curve C. This curve is called the **boundary curve**. The orientation of the surface S will induce the **positive orientation of** C. To get the positive orientation of C think of yourself as walking along the curve. While you are walking along the curve if your head is pointing in the same direction as the unit normal vectors while the surface is on the left then you are walking in the positive direction on C.

Now that we have this curve definition out of the way we can give Stokes' Theorem.

Stokes' Theorem

Let S be an oriented smooth surface that is bounded by a simple, closed, smooth boundary curve C with positive orientation. Also let $\vec{F}$ be a vector field then,

$$\int\limits_C \vec{F} \cdot d\vec{r} = \iint\limits_S \text{curl}\vec{F} \cdot d\vec{S}$$

In this theorem note that the surface S can actually be any surface so long as its boundary curve is given by C. This is something that can be used to our advantage to simplify the surface integral on occasion.

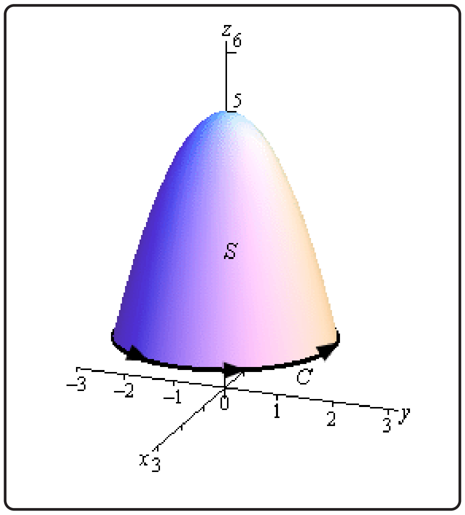
Let's take a look at a couple of examples.

Example 1

Use Stokes' Theorem to evaluate $\displaystyle\iint\limits_S \text{curl}\vec{F} \cdot d\vec{S}$ where $\vec{F} = z^2\,\vec{i} - 3xy\,\vec{j} + x^3y^3\,\vec{k}$ and S is the part of $z = 5 - x^2 - y^2$ above the plane $z = 1$. Assume that S is oriented upwards.

Solution

Let's start this off with a sketch of the surface.



In this case the boundary curve C will be where the surface intersects the plane $z = 1$ and so will be the curve

$$\begin{aligned} 1 &= 5 - x^2 - y^2 \\ x^2 + y^2 &= 4 \qquad \text{at } z = 1 \end{aligned}$$

So, the boundary curve will be the circle of radius 2 that is in the plane $z = 1$. The parameterization of this curve is,

$$\vec{r}(t) = 2\cos(t)\,\vec{i} + 2\sin(t)\,\vec{j} + \vec{k}, \quad 0 \le t \le 2\pi$$

The first two components give the circle and the third component makes sure that it is in the plane $z = 1$.

Using Stokes' Theorem we can write the surface integral as the following line integral.

$$\iint\limits_S \text{curl}\vec{F} \cdot d\vec{S} = \int\limits_C \vec{F} \cdot d\vec{r} = \int_0^{2\pi} \vec{F}(\vec{r}(t)) \cdot \vec{r}\,'(t)\;dt$$

So, it looks like we need a couple of quantities before we do this integral. Let's first get the vector field evaluated on the curve. Remember that this is simply plugging the components of the parameterization into the vector field.

$$\begin{aligned}\vec{F}\left(\vec{r}\left(t\right)\right) &= (1)^2\,\vec{i} - 3\left(2\cos(t)\right)\left(2\sin(t)\right)\vec{j} + \left(2\cos(t)\right)^3\left(2\sin(t)\right)^3\vec{k} \\ &= \vec{i} - 12\cos(t)\sin(t)\,\vec{j} + 64\cos^3(t)\sin^3(t)\,\vec{k}\end{aligned}$$

Next, we need the derivative of the parameterization and the dot product of this and the vector field.

$$\begin{aligned}\vec{r}'\left(t\right) &= -2\sin(t)\,\vec{i} + 2\cos(t)\,\vec{j} \\ \vec{F}\left(\vec{r}\left(t\right)\right)\cdot\vec{r}'\left(t\right) &= -2\sin(t) - 24\sin(t)\cos^2(t)\end{aligned}$$

We can now do the integral.

$$\begin{aligned}\iint\limits_S \operatorname{curl}\vec{F}\cdot d\vec{S} &= \int_0^{2\pi} -2\sin(t) - 24\sin(t)\cos^2(t)\,dt \\ &= \left(2\cos(t) + 8\cos^3(t)\right)\Big|_0^{2\pi} \\ &= 0\end{aligned}$$

Example 2

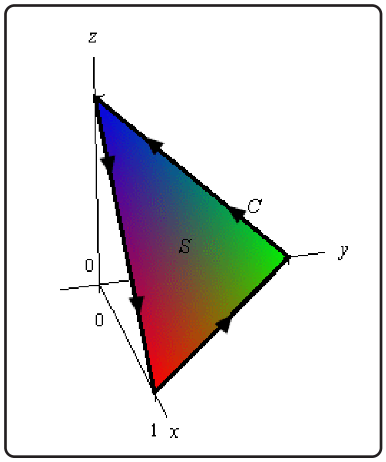
Use Stokes' Theorem to evaluate $\displaystyle\int\limits_C \vec{F}\cdot d\vec{r}$ where $\vec{F} = z^2\,\vec{i} + y^2\,\vec{j} + x\,\vec{k}$ and C is the triangle with vertices $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$ with counter-clockwise rotation.

Solution

We are going to need the curl of the vector field eventually so let's get that out of the way first.

$$\operatorname{curl}\vec{F} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ z^2 & y^2 & x \end{vmatrix} = 2z\,\vec{j} - \vec{j} = (2z-1)\,\vec{j}$$

Now, all we have is the boundary curve for the surface that we'll need to use in the surface integral. However, as noted above all we need is any surface that has this as its boundary curve. So, let's use the following plane with upwards orientation for the surface.



Since the plane is oriented upwards this induces the positive direction on C as shown. The equation of this plane is,

$$x + y + z = 1 \qquad \Rightarrow \qquad z = g(x, y) = 1 - x - y$$

Now, let's use Stokes' Theorem and get the surface integral set up.

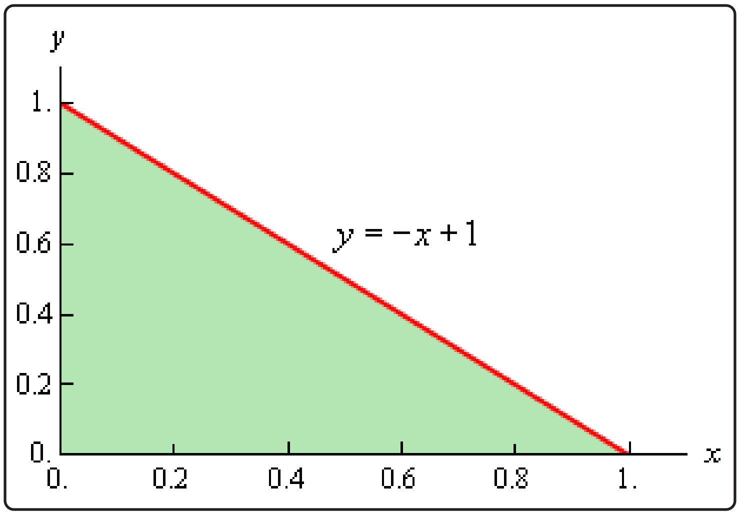
$$\begin{aligned}\int_C \vec{F} \cdot d\vec{r} &= \iint_S \text{curl}\vec{F} \cdot d\vec{S} \\ &= \iint_S (2z - 1)\,\vec{j} \cdot d\vec{S} \\ &= \iint_D (2z - 1)\,\vec{j} \cdot \frac{\nabla f}{\|\nabla f\|} \; \|\nabla f\| \, dA\end{aligned}$$

Okay, we now need to find a couple of quantities. First let's get the gradient. Recall that this comes from the function of the surface.

$$\begin{aligned}f(x, y, z) &= z - g(x, y) = z - 1 + x + y \\ \nabla f &= \vec{i} + \vec{j} + \vec{k}\end{aligned}$$

Note as well that this also points upwards and so we have the correct direction.

Now, D is the region in the xy-plane shown below,



We get the equation of the line by plugging in $z = 0$ into the equation of the plane. So based on this the ranges that define D are,

$$0 \le x \le 1 \qquad 0 \le y \le -x + 1$$

The integral is then,

$$\begin{aligned}\int_C \vec{F} \cdot d\,\vec{r} &= \iint_D (2z - 1)\,\vec{j} \cdot \left(\vec{i} + \vec{j} + \vec{k}\right)\,dA \\ &= \int_0^1 \int_0^{-x+1} 2\,(1 - x - y) - 1\,dy\,dx\end{aligned}$$

Don't forget to plug in for z since we are doing the surface integral on the plane. Finishing this out gives,

$$\begin{aligned}\int_C \vec{F} \cdot d\,\vec{r} &= \int_0^1 \int_0^{-x+1} 1 - 2x - 2y\,dy\,dx \\ &= \int_0^1 \left(y - 2xy - y^2\right)\Big|_0^{-x+1}\,dx \\ &= \int_0^1 x^2 - x\,dx \\ &= \left(\frac{1}{3}x^3 - \frac{1}{2}x^2\right)\Big|_0^1 \\ &= -\frac{1}{6}\end{aligned}$$

In both of these examples we were able to take an integral that would have been somewhat unpleasant to deal with and by the use of Stokes' Theorem we were able to convert it into an integral that wasn't too bad.

17.6 Divergence Theorem

In this section we are going to relate surface integrals to triple integrals. We will do this with the Divergence Theorem.

Divergence Theorem

Let E be a simple solid region and S is the boundary surface of E with positive orientation. Let $\vec{F}$ be a vector field whose components have continuous first order partial derivatives. Then,

$$\iint\limits_S \vec{F} \cdot d\vec{S} = \iiint\limits_E \text{div}\vec{F}\, dV$$

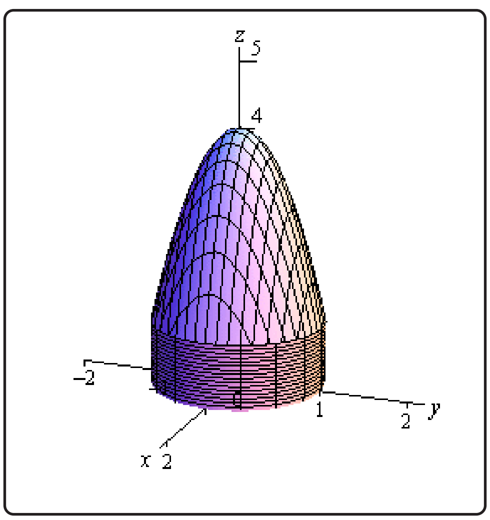
Let's see an example of how to use this theorem.

Example 1

Use the divergence theorem to evaluate $\iint\limits_S \vec{F} \cdot d\vec{S}$ where $\vec{F} = xy\,\vec{i} - \frac{1}{2}y^2\,\vec{j} + z\,\vec{k}$ and the surface consists of the three surfaces, $z = 4 - 3x^2 - 3y^2$, $1 \le z \le 4$ on the top, $x^2 + y^2 = 1$, $0 \le z \le 1$ on the sides and $z = 0$ on the bottom.

Solution

Let's start this off with a sketch of the surface.



The region E for the triple integral is then the region enclosed by these surfaces. Note that cylindrical coordinates would be a perfect coordinate system for this region. If we do that here are the limits for the ranges.

$$\begin{gathered} 0 \le z \le 4 - 3r^2 \\ 0 \le r \le 1 \\ 0 \le \theta \le 2\pi \end{gathered}$$

We'll also need the divergence of the vector field so let's get that.

$$\text{div}\vec{F} = y - y + 1 = 1$$

The integral is then,

$$\begin{aligned} \iint\limits_S \vec{F} \cdot d\vec{S} &= \iiint\limits_E \text{div}\vec{F}\, dV \\ &= \int_0^{2\pi} \int_0^1 \int_0^{4-3r^2} r\, dz\, dr\, d\theta \\ &= \int_0^{2\pi} \int_0^1 4r - 3r^3\, dr\, d\theta \\ &= \int_0^{2\pi} \left(2r^2 - \frac{3}{4}r^4\right)\Big|_0^1\, d\theta \\ &= \int_0^{2\pi} \frac{5}{4}\, d\theta \\ &= \frac{5}{2}\pi \end{aligned}$$

Appendix A : Calculus I Extras

This appendix is for material that didn't fit into other sections from the material typically taught in a Calculus I course for a variety of reasons.

Most of the sections in this appendix give proofs of most, if not all, of the various facts and properties from those sections. The proofs were not included in sections themselves because they are often lengthy proofs and in almost all cases they don't actually help to understand the material being discussed at the time. That is not to say that the proofs aren't important. They are. It's just that, for example, the proof of the Chain Rule will in no way help a student understand and actually be able to apply Chain Rule and, in fact, may actually just end up confusing the student given that it is a little hard to follow at times and fairly long.

So, to avoid causing confusion for students in a typical Calculus I course the proofs have all been put into this appendix. Therefore, those that are interested in seeing them can go through the proofs but those that don't need them to not need to worry about them.

Any proofs from the material typically taught in Calculus I material that are not given here, or in the text, are skipped due to the complexity of them and often due to the fact that they regularly involve material from upper level math courses that students don't have yet.

Note that proofs of facts/properties for material typically taught in a Calculus II or Calculus III course are either skipped outright or included in the sections themselves. The proofs that are skipped are skipped because they are often really unpleasant and will fairly regularly involve material from upper level math courses that students won't have yet. The remaining proofs are fairly short and, in most cases, there are at most one per section and so are easily included in the section, often at the very end of the section.

In addition to the proofs given here in this appendix there is also a discussion on the concept of infinity (or more appropriately a discussion on types of infinity), a discussion of the constant of integration and a quick introduction/overview of summation notation.

A.1 Proof of Various Limit Properties

In this section we are going to prove some of the basic properties and facts about limits that we saw in the Limits chapter. Before proceeding with any of the proofs we should note that many of the proofs use the precise definition of the limit and it is assumed that not only have you read that section but that you have a fairly good feel for doing that kind of proof. If you're not very comfortable using the definition of the limit to prove limits you'll find many of the proofs in this section difficult to follow.

The proofs that we'll be doing here will not be quite as detailed as those in the precise definition of the limit section. The "proofs" that we did in that section first did some work to get a guess for the δ and then we verified the guess. The reality is that often the work to get the guess is not shown and the guess for δ is just written down and then verified. For the proofs in this section where a δ is actually chosen we'll do it that way. To make matters worse, in some of the proofs in this section work very differently from those that were in the limit definition section.

So, with that out of the way, let's get to the proofs.

Limit Properties

In the Limit Properties section we gave several properties of limits. We'll prove most of them here. First, let's recall the properties here so we have them in front of us. We'll also be making a small change to the notation to make the proofs go a little easier. Here are the properties for reference purposes.

Limit Properties

Assume that $\lim\limits_{x \to a} f(x) = K$ and $\lim\limits_{x \to a} g(x) = L$ exist and that c is any constant. Then,

1. $\lim\limits_{x \to a} \left[c f(x) \right] = c \lim\limits_{x \to a} f(x) = cK$

2. $\lim\limits_{x \to a} \left[f(x) \pm g(x) \right] = \lim\limits_{x \to a} f(x) \pm \lim\limits_{x \to a} g(x) = K \pm L$

3. $\lim\limits_{x \to a} \left[f(x) g(x) \right] = \lim\limits_{x \to a} f(x) \, \lim\limits_{x \to a} g(x) = KL$

4. $\lim\limits_{x \to a} \left[\dfrac{f(x)}{g(x)} \right] = \dfrac{\lim\limits_{x \to a} f(x)}{\lim\limits_{x \to a} g(x)} = \dfrac{K}{L}$, provided $L = \lim\limits_{x \to a} g(x) \neq 0$

5. $\lim\limits_{x \to a} \left[f(x) \right]^n = \left[\lim\limits_{x \to a} f(x) \right]^n = K^n$, where n is any real number

6. $\lim\limits_{x \to a} \left[\sqrt[n]{f(x)} \right] = \sqrt[n]{\lim\limits_{x \to a} f(x)}$

7. $\lim\limits_{x \to a} c = c$

8. $\lim\limits_{x \to a} x = a$

9. $\lim\limits_{x \to a} x^n = a^n$

Note that we added values (K, L, *etc.*) to each of the limits to make the proofs much easier. In these proofs we'll be using the fact that we know $\lim\limits_{x \to a} f(x) = K$ and $\lim\limits_{x \to a} g(x) = L$ we'll use the definition of the limit to make a statement about $|f(x) - K|$ and $|g(x) - L|$ which will then be used to prove what we actually want to prove. When you see these statements do not worry too much about why we chose them as we did. The reason will become apparent once the proof is done.

Also, we're not going to be doing the proofs in the order they are written above. Some of the proofs will be easier if we've got some of the others proved first.

Proof of 7

This is a very simple proof. To make the notation a little clearer let's define the function $f(x) = c$ then what we're being asked to prove is that $\lim\limits_{x \to a} f(x) = c$. So let's do that.

Let $\varepsilon > 0$ and we need to show that we can find a $\delta > 0$ so that

$$|f(x) - c| < \varepsilon \qquad \text{whenever} \qquad 0 < |x - a| < \delta$$

The left inequality is trivially satisfied for any x however because we defined $f(x) = c$. So simply choose $\delta > 0$ to be any number you want (you generally can't do this with these proofs). Then,

$$|f(x) - c| = |c - c| = 0 < \varepsilon$$

Proof of 1

There are several ways to prove this part. If you accept **3** And **7** then all you need to do is let $g(x) = c$ and then this is a direct result of **3** and **7**. However, we'd like to do a more rigorous mathematical proof. So here is that proof.

First, note that if $c = 0$ then $cf(x) = 0$ and so,

$$\lim_{x \to a} [0f(x)] = \lim_{x \to a} 0 = 0 = 0f(x)$$

The limit evaluation is a special case of **7** (with $c = 0$) which we just proved Therefore we know **1** is true for $c = 0$ and so we can assume that $c \neq 0$ for the remainder of this proof.

Let $\varepsilon > 0$ then because $\lim\limits_{x \to a} f(x) = K$ by the definition of the limit there is a $\delta_1 > 0$ such that,

$$|f(x) - K| < \frac{\varepsilon}{|c|} \qquad \text{whenever} \qquad 0 < |x - a| < \delta_1$$

Now choose $\delta = \delta_1$ and we need to show that

$$|cf(x) - cK| < \varepsilon \qquad \text{whenever} \qquad 0 < |x - a| < \delta$$

and we'll be done. So, assume that $0 < |x - a| < \delta$ and then,

$$|cf(x) - cK| = |c|\,|f(x) - K| < |c|\frac{\varepsilon}{|c|} = \varepsilon$$

Proof of 2

Note that we'll need something called the **triangle inequality** in this proof. The triangle inequality states that,

$$|a + b| \le |a| + |b|$$

Here's the actual proof.

Proof of 2

We'll be doing this proof in two parts. First let's prove $\lim\limits_{x \to a} [f(x) + g(x)] = K + L$.

Let $\varepsilon > 0$ then because $\lim\limits_{x \to a} f(x) = K$ and $\lim\limits_{x \to a} g(x) = L$ there is a $\delta_1 > 0$ and a $\delta_2 > 0$ such that,

$$\begin{aligned} |f(x) - K| &< \frac{\varepsilon}{2} \qquad \text{whenever} \qquad 0 < |x - a| < \delta_1 \\ |g(x) - L| &< \frac{\varepsilon}{2} \qquad \text{whenever} \qquad 0 < |x - a| < \delta_2 \end{aligned}$$

Now choose $\delta = \min\{\delta_1, \delta_2\}$. Then we need to show that

$$|f(x) + g(x) - (K + L)| < \varepsilon \qquad \text{whenever} \qquad 0 < |x - a| < \delta$$

Assume that $0 < |x - a| < \delta$. We then have,

$$\begin{aligned} |f(x) + g(x) - (K + L)| &= |(f(x) - K) + (g(x) - L)| \\ &\le |f(x) - K| + |g(x) - L| \qquad \text{by the triangle inequality} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon \end{aligned}$$

In the third step we used the fact that, by our choice of δ, we also have $0 < |x - a| < \delta_1$ and $0 < |x - a| < \delta_2$ and so we can use the initial statements in our proof.

Next, we need to prove $\lim\limits_{x \to a} [f(x) - g(x)] = K - L$. We could do a similar proof as we did above for the sum of two functions. However, we might as well take advantage of the fact that we've proven this for a sum and that we've also proven **1**

$$\begin{aligned}
\lim_{x \to a} [f(x) - g(x)] &= \lim_{x \to a} [f(x) + (-1) g(x)] \\
&= \lim_{x \to a} f(x) + \lim_{x \to a} (-1) g(x) && \text{by first part of } \mathbf{2}. \\
&= \lim_{x \to a} f(x) + (-1) \lim_{x \to a} g(x) && \text{by } \mathbf{1}. \\
&= K + (-1) L \\
&= K - L
\end{aligned}$$

Proof of 3

This one is a little tricky. First, let's note that because $\lim\limits_{x \to a} f(x) = K$ and $\lim\limits_{x \to a} g(x) = L$ we can use **2** and **7** to prove the following two limits.

$$\begin{aligned}
\lim_{x \to a} [f(x) - K] &= \lim_{x \to a} f(x) - \lim_{x \to a} K = K - K = 0 \\
\lim_{x \to a} [g(x) - L] &= \lim_{x \to a} g(x) - \lim_{x \to a} L = L - L = 0
\end{aligned}$$

Now, let $\varepsilon > 0$. Then there is a $\delta_1 > 0$ and a $\delta_2 > 0$ such that,

$$\begin{aligned}
|(f(x) - K) - 0| &< \sqrt{\varepsilon} && \text{whenever} && 0 < |x - a| < \delta_1 \\
|(g(x) - L) - 0| &< \sqrt{\varepsilon} && \text{whenever} && 0 < |x - a| < \delta_2
\end{aligned}$$

Choose $\delta = \min\{\delta_1, \delta_2\}$. If $0 < |x - a| < \delta$ we then get,

$$\begin{aligned}
|[f(x) - K][g(x) - L] - 0| &= |f(x) - K| \, |g(x) - L| \\
&< \sqrt{\varepsilon}\sqrt{\varepsilon} \\
&= \varepsilon
\end{aligned}$$

So, we've managed to prove that,

$$\lim_{x \to a} [f(x) - K][g(x) - L] = 0$$

This apparently has nothing to do with what we actually want to prove, but as you'll see in a bit it is needed.

Before launching into the actual proof of **3** let's do a little Algebra. First, expand the following product.

$$[f(x) - K][g(x) - L] = f(x) g(x) - L f(x) - K g(x) + KL$$

Rearranging this gives the following way to write the product of the two functions.

$$f(x)\,g(x) = [f(x) - K]\,[g(x) - L] + Lf(x) + Kg(x) - KL$$

With this we can now proceed with the proof of **3**.

$$\begin{aligned}
\lim_{x\to a} f(x)\,g(x) &= \lim_{x\to a} [[f(x) - K]\,[g(x) - L] + Lf(x) + Kg(x) - KL] \\
&= \lim_{x\to a} [f(x) - K]\,[g(x) - L] + \lim_{x\to a} Lf(x) + \lim_{x\to a} Kg(x) - \lim_{x\to a} KL \\
&= 0 + \lim_{x\to a} Lf(x) + \lim_{x\to a} Kg(x) - \lim_{x\to a} KL \\
&= LK + KL - KL \\
&= LK
\end{aligned}$$

Fairly simple proof really, once you see all the steps that you have to take before you even start. The second step made multiple uses of property **2**. In the third step we used the limit we initially proved. In the fourth step we used properties **1** and **7**. Finally, we just did some simplification.

Proof of 4

This one is also a little tricky. First, we'll start of by proving,

$$\lim_{x\to a} \frac{1}{g(x)} = \frac{1}{L}$$

Let $\varepsilon > 0$. We'll not need this right away, but these proofs always start off with this statement. Now, because $\lim\limits_{x\to a} g(x) = L$ there is a $\delta_1 > 0$ such that,

$$|g(x) - L| < \frac{|L|}{2} \qquad \text{whenever} \qquad 0 < |x - a| < \delta_1$$

Now, assuming that $0 < |x - a| < \delta_1$ we have,

$$\begin{aligned}
|L| &= |L - g(x) + g(x)| && \text{just adding zero to } L \\
&\le |L - g(x)| + |g(x)| && \text{using the triangle inequality} \\
&= |g(x) - L| + |g(x)| && |L - g(x)| = |g(x) - L| \\
&< \frac{|L|}{2} + |g(x)| && \text{assuming that } 0 < |x - a| < \delta_1
\end{aligned}$$

Rearranging this gives,

$$|L| < \frac{|L|}{2} + |g(x)| \qquad \Rightarrow \qquad \frac{|L|}{2} < |g(x)| \qquad \Rightarrow \qquad \frac{1}{|g(x)|} < \frac{2}{|L|}$$

Now, there is also a $\delta_2 > 0$ such that,

$$|g(x) - L| < \frac{|L|^2}{2}\varepsilon \qquad \text{whenever} \qquad 0 < |x - a| < \delta_2$$

Choose $\delta = \min\{\delta_1, \delta_2\}$. If $0 < |x - a| < \delta$ we have,

$$\begin{aligned}
\left|\frac{1}{g(x)} - \frac{1}{L}\right| &= \left|\frac{L - g(x)}{Lg(x)}\right| && \text{common denominators} \\
&= \frac{1}{|Lg(x)|}|L - g(x)| && \text{doing a little rewriting} \\
&= \frac{1}{|L|}\frac{1}{|g(x)|}|g(x) - L| && \text{doing a little more rewriting} \\
&< \frac{1}{|L|}\frac{2}{|L|}|g(x) - L| && \text{assuming that } 0 < |x - a| < \delta \le \delta_1 \\
&< \frac{2}{|L|^2}\frac{|L|^2}{2}\varepsilon && \text{assuming that } 0 < |x - a| < \delta \le \delta_2 \\
&= \varepsilon
\end{aligned}$$

Now that we've proven $\lim\limits_{x \to a} \frac{1}{g(x)} = \frac{1}{L}$ the more general fact is easy.

$$\begin{aligned}
\lim_{x \to a}\left[\frac{f(x)}{g(x)}\right] &= \lim_{x \to a}\left[f(x)\frac{1}{g(x)}\right] \\
&= \lim_{x \to a} f(x) \lim_{x \to a}\frac{1}{g(x)} && \text{using property } \mathbf{3}. \\
&= K\frac{1}{L} = \frac{K}{L}
\end{aligned}$$

Proof of 5. for n an integer

As noted we're only going to prove **5** for integer exponents. This will also involve proof by induction so if you aren't familiar with induction proofs you can skip this proof.

So, we're going to prove,

$$\lim_{x \to a}[f(x)]^n = \left[\lim_{x \to a} f(x)\right]^n = K^n, \qquad n \ge 2,\ n \text{ is an integer.}$$

For $n = 2$ we have nothing more than a special case of property **3.**

$$\lim_{x \to a}[f(x)]^2 = \lim_{x \to a} f(x)\, f(x) = \lim_{x \to a} f(x) \lim_{x \to a} f(x) = KK = K^2$$

So, **5** is proven for $n = 2$. Now assume that **5** is true for $n - 1$, or $\lim\limits_{x \to a}[f(x)]^{n-1} = K^{n-1}$.

Then, again using property **3** we have,

$$\begin{aligned}\lim_{x \to a} [f(x)]^n &= \lim_{x \to a} \left([f(x)]^{n-1} f(x) \right) \\ &= \lim_{x \to a} [f(x)]^{n-1} \lim_{x \to a} f(x) \\ &= K^{n-1} K \\ &= K^n\end{aligned}$$

Proof of 6

As pointed out in the Limit Properties section this is nothing more than a special case of the full version of **5** and the proof is given there and so is the proof is not give here.

Proof of 8

This is a simple proof. If we define $f(x) = x$ to make the notation a little easier, we're being asked to prove that $\lim\limits_{x \to a} f(x) = a$.

Let $\varepsilon > 0$ and let $\delta = \varepsilon$. Then, if $0 < |x - a| < \delta = \varepsilon$ we have,

$$|f(x) - a| = |x - a| < \delta = \varepsilon$$

So, we've proved that $\lim\limits_{x \to a} x = a$.

Proof of 9

This is just a special case of property **5** with $f(x) = x$ and so we won't prove it here.

Facts, Infinite Limits

Given the functions $f(x)$ and $g(x)$ suppose we have,

$$\lim_{x \to c} f(x) = \infty \qquad\qquad \lim_{x \to c} g(x) = L$$

for some real numbers c and L. Then,

1. $\lim\limits_{x \to c} [f(x) \pm g(x)] = \infty$
2. If $L > 0$ then $\lim\limits_{x \to c} [f(x)\, g(x)] = \infty$
3. If $L < 0$ then $\lim\limits_{x \to c} [f(x)\, g(x)] = -\infty$
4. $\lim\limits_{x \to c} \frac{g(x)}{f(x)} = 0$

Partial Proof of 1

We will prove $\lim\limits_{x \to c} [f(x) + g(x)] = \infty$ here. The proof of $\lim\limits_{x \to c} [f(x) - g(x)] = \infty$ is nearly identical and is left to you.

Let $M > 0$ then because we know $\lim\limits_{x \to c} f(x) = \infty$ there exists a $\delta_1 > 0$ such that if $0 < |x - c| < \delta_1$ we have,

$$f(x) > M - L + 1$$

Also, because we know $\lim\limits_{x \to c} g(x) = L$ there exists a $\delta_2 > 0$ such that if $0 < |x - c| < \delta_2$ we have,

$$0 < |g(x) - L| < 1 \qquad \to \qquad -1 < g(x) - L < 1 \qquad \to \qquad L - 1 < g(x) < L + 1$$

Now, let $\delta = \min\{\delta_1, \delta_2\}$ and so if $0 < |x - c| < \delta$ we know from the above statements that we will have both,

$$f(x) > M - L + 1 \qquad\qquad g(x) > L - 1$$

This gives us,

$$\begin{aligned} f(x) + g(x) &> M - L + 1 + L - 1 \\ &= M \qquad\qquad \Rightarrow \qquad\qquad f(x) + g(x) > M \end{aligned}$$

So, we've proved that $\lim\limits_{x \to c} [f(x) + g(x)] = \infty$.

Proof of 2

Let $M > 0$ then because we know $\lim\limits_{x \to c} f(x) = \infty$ there exists a $\delta_1 > 0$ such that if $0 < |x - c| < \delta_1$ we have,

$$f(x) > \frac{2M}{L}$$

Also, because we know $\lim\limits_{x \to c} g(x) = L$ there exists a $\delta_2 > 0$ such that if $0 < |x - c| < \delta_2$ we have,

$$0 < |g(x) - L| < \frac{L}{2} \qquad \to \qquad -\frac{L}{2} < g(x) - L < \frac{L}{2} \qquad \to \qquad \frac{L}{2} < g(x) < \frac{3L}{2}$$

Note that because we know that $L > 0$ choosing $\frac{L}{2}$ in the first inequality above is a valid choice because it will also be positive as required by the definition of the limit.

Now, let $\delta = \min\{\delta_1, \delta_2\}$ and so if $0 < |x - c| < \delta$ we know from the above statements that we will have both,

$$f(x) > \frac{2M}{L} \qquad\qquad g(x) > \frac{L}{2}$$

This gives us,

$$\begin{aligned} f(x)\,g(x) &> \left(\frac{2M}{L}\right)\left(\frac{L}{2}\right) \\ &= M \qquad\qquad \Rightarrow \qquad\qquad f(x)\,g(x) > M \end{aligned}$$

So, we've proved that $\lim\limits_{x \to c} f(x)\,g(x) = \infty$.

Proof of 3

Let $M > 0$ then because we know $\lim\limits_{x \to c} f(x) = \infty$ there exists a $\delta_1 > 0$ such that if $0 < |x - c| < \delta_1$ we have,

$$f(x) > \frac{-2M}{L}$$

Note that because $L < 0$ in the case we will have $\frac{-2M}{L} > 0$ here.

Next, because we know $\lim\limits_{x \to c} g(x) = L$ there exists a $\delta_2 > 0$ such that if $0 < |x - c| < \delta_2$ we have,

$$0 < |g(x) - L| < -\frac{L}{2} \qquad \to \qquad \frac{L}{2} < g(x) - L < -\frac{L}{2} \qquad \to \qquad \frac{3L}{2} < g(x) < \frac{L}{2}$$

Again, because we know that $L < 0$ we will have $-\frac{L}{2} > 0$. Also, for reasons that will shortly be apparent, multiply the final inequality by a minus sign to get,

$$-\frac{L}{2} < -g(x) < -\frac{3L}{2}$$

Now, let $\delta = \min\{\delta_1, \delta_2\}$ and so if $0 < |x-c| < \delta$ we know from the above statements that we will have both,

$$f(x) > \frac{-2M}{L} \qquad\qquad -g(x) > -\frac{L}{2}$$

This gives us,

$$\begin{aligned} -f(x)\,g(x) &= f(x)\,[-g(x)] \\ &> \left(-\frac{2M}{L}\right)\left(-\frac{L}{2}\right) \\ &= M \qquad\qquad \Rightarrow \qquad\qquad -f(x)\,g(x) > M \end{aligned}$$

This may seem to not be what we needed however multiplying this by a minus sign gives,

$$f(x)\,g(x) < -M$$

and because we originally chose $M > 0$ we have now proven that $\lim\limits_{x\to c} f(x)\,g(x) = -\infty$.

Proof of 4

We'll need to do this in three cases. Let's start with the easiest case.

Case 1 : $L = 0$

Let $\varepsilon > 0$ then because we know $\lim\limits_{x\to c} f(x) = \infty$ there exists a $\delta_1 > 0$ such that if $0 < |x-c| < \delta_1$ we have,

$$f(x) > \frac{1}{\sqrt{\varepsilon}} > 0$$

Next, because we know $\lim\limits_{x\to c} g(x) = 0$ there exists a $\delta_2 > 0$ such that if $0 < |x-c| < \delta_2$ we have,

$$0 < |g(x)| < \sqrt{\varepsilon}$$

Now, let $\delta = \min\{\delta_1, \delta_2\}$ and so if $0 < |x-c| < \delta$ we know from the above statements that we will have both,

$$f(x) > \frac{1}{\sqrt{\varepsilon}} \qquad\qquad |g(x)| < \sqrt{\varepsilon}$$

This gives us,

$$\left|\frac{g(x)}{f(x)}\right| = \frac{|g(x)|}{f(x)} < \frac{\sqrt{\varepsilon}}{f(x)} < \frac{\sqrt{\varepsilon}}{{}^{1}/_{\sqrt{\varepsilon}}} = \varepsilon$$

In the second step we could remove the absolute value bars from $f(x)$ because we know it is positive.

So, we proved that $\lim\limits_{x\to c} \frac{g(x)}{f(x)} = 0$ if $L = 0$.

Case 2 : $L > 0$

Let $\varepsilon > 0$ then because we know $\lim\limits_{x\to c} f(x) = \infty$ there exists a $\delta_1 > 0$ such that if $0 < |x-c| < \delta_1$ we have,

$$f(x) > \frac{3L}{2\varepsilon} > 0$$

Next, because we know $\lim\limits_{x\to c} g(x) = L$ there exists a $\delta_2 > 0$ such that if $0 < |x-c| < \delta_2$ we have,

$$0 < |g(x) - L| < \frac{L}{2} \qquad \rightarrow \qquad -\frac{L}{2} < g(x) - L < \frac{L}{2} \qquad \rightarrow \qquad \frac{L}{2} < g(x) < \frac{3L}{2}$$

Also, because we are assuming that $L > 0$ it is safe to assume that for $0 < |x-c| < \delta_2$ we have $g(x) > 0$.

Now, let $\delta = \min\{\delta_1, \delta_2\}$ and so if $0 < |x-c| < \delta$ we know from the above statements that we will have both,

$$f(x) > \frac{3L}{2\varepsilon} \qquad\qquad g(x) < \frac{3L}{2}$$

This gives us,

$$\left|\frac{g(x)}{f(x)}\right| = \frac{g(x)}{f(x)} < \frac{{}^{3L}/_{2}}{|f(x)|} < \frac{{}^{3L}/_{2}}{{}^{3L}/_{2\varepsilon}} = \varepsilon$$

In the second step we could remove the absolute value bars because we know or can safely assume (as noted above) that both functions were positive.

So, we proved that $\lim\limits_{x\to c} \frac{g(x)}{f(x)} = 0$ if $L > 0$.

Case 3 : $L < 0$.

Let $\varepsilon > 0$ then because we know $\lim\limits_{x\to c} f(x) = \infty$ there exists a $\delta_1 > 0$ such that if $0 < |x-c| < \delta_1$ we have,

$$f(x) > \frac{-3L}{2\varepsilon} > 0$$

Next, because we know $\lim\limits_{x\to c} g(x) = L$ there exists a $\delta_2 > 0$ such that if $0 < |x-c| < \delta_2$ we have,

$$0 < |g(x) - L| < -\frac{L}{2} \qquad \rightarrow \qquad \frac{L}{2} < g(x) - L < -\frac{L}{2} \qquad \rightarrow \qquad \frac{3L}{2} < g(x) < \frac{L}{2}$$

Next, multiply this be a negative sign to get,

$$-\frac{L}{2} < -g(x) < -\frac{3L}{2}$$

Also, because we are assuming that $L < 0$ it is safe to assume that for $0 < |x - c| < \delta_2$ we have $g(x) < 0$.

Now, let $\delta = \min\{\delta_1, \delta_2\}$ and so if $0 < |x - c| < \delta$ we know from the above statements that we will have both,

$$f(x) > \frac{-3L}{2\varepsilon} \qquad\qquad -g(x) < \frac{-3L}{2}$$

This gives us,

$$\left|\frac{g(x)}{f(x)}\right| = \frac{-g(x)}{f(x)} < \frac{{}^{-3L}/_{2}}{|f(x)|} < \frac{{}^{-3L}/_{2}}{{}^{-3L}/_{2\varepsilon}} = \varepsilon$$

In the second step we could remove the absolute value bars by adding in the negative because we know that $f(x) > 0$ and can safely assume that $g(x) < 0$ (as noted above).

So, we proved that $\lim\limits_{x \to c} \frac{g(x)}{f(x)} = 0$ if $L < 0$.

Fact 1, Limits At Infinity, Part 1

1. If r is a positive rational number and c is any real number then,

$$\lim_{x \to \infty} \frac{c}{x^r} = 0$$

2. If r is a positive rational number, c is any real number and x^r is defined for $x < 0$ then,

$$\lim_{x \to -\infty} \frac{c}{x^r} = 0$$

Proof of 1

This is actually a fairly simple proof but we'll need to do three separate cases.

Case 1 : Assume that $c > 0$. Next, let $\varepsilon > 0$ and define

$$M = \sqrt[r]{\frac{c}{\varepsilon}}$$

Note that because c and ε are both positive we know that this root will exist. Now, assume that we have

$$x > M = \sqrt[r]{\frac{c}{\varepsilon}}$$

Given this assumption we have,

$$\begin{aligned} x &> \sqrt[r]{\frac{c}{\varepsilon}} \\ x^r &> \frac{c}{\varepsilon} && \text{get rid of the root} \\ \frac{c}{x^r} &< \varepsilon && \text{rearrange things a little} \\ \left|\frac{c}{x^r} - 0\right| &< \varepsilon && \text{everything is positive so we can add absolute value bars} \end{aligned}$$

So, provided $c > 0$ we've proven that $\lim\limits_{x\to\infty} \frac{c}{x^r} = 0$.

Case 2 : Assume that $c = 0$. Here all we need to do is the following,

$$\lim_{x\to\infty} \frac{c}{x^r} = \lim_{x\to\infty} \frac{0}{x^r} = \lim_{x\to\infty} 0 = 0$$

Case 3 : Finally, assume that $c < 0$. In this case we can then write $c = -k$ where $k > 0$. Then using Case 1 and the fact that we can factor constants out of a limit we get,

$$\lim_{x\to\infty} \frac{c}{x^r} = \lim_{x\to\infty} \frac{-k}{x^r} = -\lim_{x\to\infty} \frac{k}{x^r} = -0 = 0$$

Proof of 2

This is very similar to the proof of **1** so we'll just do the first case (as it's the hardest) and leave the other two cases up to you to prove.

Case 1 : Assume that $c > 0$. Next, let $\varepsilon > 0$ and define

$$N = -\sqrt[r]{\frac{c}{\varepsilon}}$$

Note that because c and ε are both positive we know that this root will exist. Now, assume that we have

$$x < N = -\sqrt[r]{\frac{c}{\varepsilon}}$$

Note that this assumption also tells us that x will be negative. Give this assumption we have,

$$\begin{aligned} x &< -\sqrt[r]{\frac{c}{\varepsilon}} && \\ |x| &> \left|\sqrt[r]{\frac{c}{\varepsilon}}\right| && \text{take absolute value of both sides} \\ |x^r| &> \left|\frac{c}{\varepsilon}\right| && \text{get rid of the root} \\ \left|\frac{c}{x^r}\right| &< |\varepsilon| = \varepsilon && \text{rearrange things a little and use the fact that } \varepsilon > 0 \\ \left|\frac{c}{x^r} - 0\right| &< \varepsilon && \text{rewrite things a little} \end{aligned}$$

So, provided $c > 0$ we've proven that $\lim\limits_{x \to -\infty} \frac{c}{x^r} = 0$. Note that the main difference here is that we need to take the absolute value first to deal with the minus sign. Because both sides are negative we know that when we take the absolute value of both sides the direction of the inequality will have to switch as well.

Case 2, Case 3 : As noted above these are identical to the proof of the corresponding cases in the first proof and so are omitted here.

Fact 2, Limits At Infinity, Part I

If $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ is a polynomial of degree n (*i.e.* $a_n \neq 0$) then,

$$\lim_{x \to \infty} p(x) = \lim_{x \to \infty} a_n x^n \qquad\qquad \lim_{x \to -\infty} p(x) = \lim_{x \to -\infty} a_n x^n$$

Proof of $\lim\limits_{x \to \infty} p(x) = \lim\limits_{x \to \infty} a_n x^n$

We're going to prove this in an identical fashion to the problems that we worked in this section involving polynomials. We'll first factor out $a_n x^n$ from the polynomial and then make a giant use of **Fact 1** (which we just proved above) and the basic properties of limits.

$$\begin{aligned} \lim_{x \to \infty} p(x) &= \lim_{x \to \infty} \left(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0\right) \\ &= \lim_{x \to \infty} \left[a_n x^n \left(1 + \frac{a_{n-1}}{a_n x} + \cdots + \frac{a_1}{a_n x^{n-1}} + \frac{a_0}{a_n x^n}\right)\right] \end{aligned}$$

Now, clearly the limit of the second term is one and the limit of the first term will be either ∞ or $-\infty$ depending upon the sign of a_n. Therefore by the Facts from the Infinite Limits section we can see that the limit of the whole polynomial will be the same as the limit of the first term or,

$$\lim_{x\to\infty} p(x) = \lim_{x\to\infty} \left(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0\right) = \lim_{x\to\infty} a_n x^n$$

Proof of $\lim\limits_{x\to-\infty} p(x) = \lim\limits_{x\to-\infty} a_n x^n$

The proof of this part is literally identical to the proof of the first part, with the exception that all ∞'s are changed to $-\infty$, and so is omitted here.

Fact 2, Continuity

If $f(x)$ is continuous at $x = b$ and $\lim\limits_{x\to a} g(x) = b$ then,

$$\lim_{x\to a} f(g(x)) = f\left(\lim_{x\to a} g(x)\right) = f(b)$$

Proof

Let $\varepsilon > 0$ then we need to show that there is a $\delta > 0$ such that,

$$|f(g(x)) - f(b)| < \varepsilon \qquad \text{whenever} \qquad 0 < |x-a| < \delta$$

Let's start with the fact that $f(x)$ is continuous at $x = b$. Recall that this means that $\lim\limits_{x\to b} f(x) = f(b)$ and so there must be a $\delta_1 > 0$ so that,

$$|f(x) - f(b)| < \varepsilon \qquad \text{whenever} \qquad 0 < |x-b| < \delta_1$$

Now, let's recall that $\lim\limits_{x\to a} g(x) = b$. This means that there must be a $\delta > 0$ so that,

$$|g(x) - b| < \delta_1 \qquad \text{whenever} \qquad 0 < |x-a| < \delta$$

But all this means that we're done.

Let's summarize up. First assume that $0 < |x-a| < \delta$. This then tells us that,

$$|g(x) - b| < \delta_1$$

But, we also know that if $0 < |x-b| < \delta_1$ then we must also have $|f(x) - f(b)| < \varepsilon$. What this is telling us is that if a number is within a distance of δ_1 of b then we can plug that number into $f(x)$ and we'll be within a distance of ε of $f(b)$.

So,

$$|g(x) - b| < \delta_1$$

is telling us that $g(x)$ is within a distance of δ_1 of b and so if we plug it into $f(x)$ we'll get,

$$|f(g(x)) - f(b)| < \varepsilon$$

and this is exactly what we wanted to show.

A.2 Proof of Various Derivative Properties

In this section we're going to prove many of the various derivative facts, formulas and/or properties that we encountered in the early part of the Derivatives chapter. Not all of them will be proved here and some will only be proved for special cases, but at least you'll see that some of them aren't just pulled out of the air.

Theorem, from Definition of Derivative

If $f(x)$ is differentiable at $x = a$ then $f(x)$ is continuous at $x = a$.

Proof

Because $f(x)$ is differentiable at $x = a$ we know that

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists. We'll need this in a bit.

If we next assume that $x \neq a$ we can write the following,

$$f(x) - f(a) = \frac{f(x) - f(a)}{x - a}(x - a)$$

Then basic properties of limits tells us that we have,

$$\begin{aligned} \lim_{x \to a}(f(x) - f(a)) &= \lim_{x \to a}\left[\frac{f(x) - f(a)}{x - a}(x - a)\right] \\ &= \lim_{x \to a}\frac{f(x) - f(a)}{x - a} \quad \lim_{x \to a}(x - a) \end{aligned}$$

The first limit on the right is just $f'(a)$ as we noted above and the second limit is clearly zero and so,

$$\lim_{x \to a}(f(x) - f(a)) = f'(a) \cdot 0 = 0$$

Okay, we've managed to prove that $\lim\limits_{x \to a}(f(x) - f(a)) = 0$. But just how does this help us to prove that $f(x)$ is continuous at $x = a$?

Let's start with the following.

$$\lim_{x \to a} f(x) = \lim_{x \to a}[f(x) + f(a) - f(a)]$$

Note that we've just added in zero on the right side. A little rewriting and the use of limit properties gives,

$$\begin{aligned}\lim_{x\to a} f(x) &= \lim_{x\to a}\left[f(a) + f(x) - f(a)\right]\\ &= \lim_{x\to a} f(a) + \lim_{x\to a}\left[f(x) - f(a)\right]\end{aligned}$$

Now, we just proved above that $\lim\limits_{x\to a}\left(f(x) - f(a)\right) = 0$ and because $f(a)$ is a constant we also know that $\lim\limits_{x\to a} f(a) = f(a)$ and so this becomes,

$$\lim_{x\to a} f(x) = \lim_{x\to a} f(a) + 0 = f(a)$$

Or, in other words,

$$\lim_{x\to a} f(x) = f(a)$$

but this is exactly what it means for $f(x)$ is continuous at $x = a$ and so we're done.

Proof of Sum/Difference of Two Functions : $(f(x) \pm g(x))' = f'(x) \pm g'(x)$

This is easy enough to prove using the definition of the derivative. We'll start with the sum of two functions. First plug the sum into the definition of the derivative and rewrite the numerator a little.

$$\begin{aligned}(f(x) + g(x))' &= \lim_{h\to 0}\frac{f(x+h) + g(x+h) - (f(x) + g(x))}{h}\\ &= \lim_{h\to 0}\frac{f(x+h) - f(x) + g(x+h) - g(x)}{h}\end{aligned}$$

Now, break up the fraction into two pieces and recall that the limit of a sum is the sum of the limits. Using this fact we see that we end up with the definition of the derivative for each of the two functions.

$$\begin{aligned}(f(x) + g(x))' &= \lim_{h\to 0}\frac{f(x+h) - f(x)}{h} + \lim_{h\to 0}\frac{g(x+h) - g(x)}{h}\\ &= f'(x) + g'(x)\end{aligned}$$

The proof of the difference of two functions in nearly identical so we'll give it here without any explanation.

$$\begin{aligned}(f(x)-g(x))' &= \lim_{h\to 0}\frac{f(x+h)-g(x+h)-(f(x)-g(x))}{h}\\ &= \lim_{h\to 0}\frac{f(x+h)-f(x)-(g(x+h)-g(x))}{h}\\ &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}-\frac{g(x+h)-g(x)}{h}\\ &= f'(x)-g'(x)\end{aligned}$$

Proof of Constant Times a Function : $(cf(x))' = cf'(x)$

This is property is very easy to prove using the definition provided you recall that we can factor a constant out of a limit. Here's the work for this property.

$$(cf(x))' = \lim_{h\to 0}\frac{cf(x+h)-cf(x)}{h} = c\lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = cf'(x)$$

Proof of the Derivative of a Constant : $\dfrac{d}{dx}(c) = 0$

This is very easy to prove using the definition of the derivative so define $f(x) = c$ and the use the definition of the derivative.

$$f'(x) = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} = \lim_{h\to 0}\frac{c-c}{h} = \lim_{h\to 0}0 = 0$$

Power Rule

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

There are actually three proofs that we can give here and we're going to go through all three here so you can see all of them. However, having said that, for the first two we will need to restrict n to be a positive integer. At the time that the Power Rule was introduced only enough information has been given to allow the proof for only integers. So, the first two proofs are really to be read at that point.

The third proof will work for any real number n. However, it does assume that you've read most of the Derivatives chapter and so should only be read after you've gone through the whole chapter.

Proof 1

In this case as noted above we need to assume that n is a positive integer. We'll use the definition of the derivative and the Binomial Theorem in this theorem. The Binomial Theorem tells us that,

$$\begin{aligned}(a+b)^n &= \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k \\ &= a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \binom{n}{3}a^{n-3}b^3 + \\ &\qquad \cdots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n \\ &= a^n + na^{n-1}b + \frac{n(n-1)}{2!}a^{n-2}b^2 + \frac{n(n-1)(n-2)}{3!}a^{n-3}b^3 + \cdots + nab^{n-1} + b^n\end{aligned}$$

where,

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

are called the binomial coefficients and $n! = n(n-1)(n-2)\cdots(2)(1)$ is the factorial.

So, let's go through the details of this proof. First, plug $f(x) = x^n$ into the definition of the derivative and use the Binomial Theorem to expand out the first term.

$$\begin{aligned}f'(x) &= \lim_{h\to 0}\frac{(x+h)^n - x^n}{h} \\ &= \lim_{h\to 0}\frac{\left(x^n + nx^{n-1}h + \frac{n(n-1)}{2!}x^{n-2}h^2 + \cdots + nxh^{n-1} + h^n\right) - x^n}{h}\end{aligned}$$

Now, notice that we can cancel an x^n and then each term in the numerator will have an h in them that can be factored out and then canceled against the h in the denominator. At this point we can evaluate the limit.

$$\begin{aligned}f'(x) &= \lim_{h\to 0}\frac{nx^{n-1}h + \frac{n(n-1)}{2!}x^{n-2}h^2 + \cdots + nxh^{n-1} + h^n}{h} \\ &= \lim_{h\to 0} nx^{n-1} + \frac{n(n-1)}{2!}x^{n-2}h + \cdots + nxh^{n-2} + h^{n-1} \\ &= nx^{n-1}\end{aligned}$$

Proof 2

For this proof we'll again need to restrict n to be a positive integer. In this case if we define $f(x) = x^n$ we know from the alternate limit form of the definition of the derivative that the derivative $f'(a)$ is given by,

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a} \frac{x^n - a^n}{x - a}$$

Now we have the following formula,

$$x^n - a^n = (x - a)\left(x^{n-1} + ax^{n-2} + a^2x^{n-3} + \cdots + a^{n-3}x^2 + a^{n-2}x + a^{n-1}\right)$$

You can verify this if you'd like by simply multiplying the two factors together. Also, notice that there are a total of n terms in the second factor (this will be important in a bit).

If we plug this into the formula for the derivative we see that we can cancel the $x - a$ and then compute the limit.

$$\begin{aligned} f'(a) &= \lim_{x \to a} \frac{(x - a)\left(x^{n-1} + ax^{n-2} + a^2x^{n-3} + \cdots + a^{n-3}x^2 + a^{n-2}x + a^{n-1}\right)}{x - a} \\ &= \lim_{x \to a} x^{n-1} + ax^{n-2} + a^2x^{n-3} + \cdots + a^{n-3}x^2 + a^{n-2}x + a^{n-1} \\ &= a^{n-1} + aa^{n-2} + a^2a^{n-3} + \cdots + a^{n-3}a^2 + a^{n-2}a + a^{n-1} \\ &= na^{n-1} \end{aligned}$$

After combining the exponents in each term we can see that we get the same term. So, then recalling that there are n terms in second factor we can see that we get what we claimed it would be.

To completely finish this off we simply replace the a with an x to get,

$$f'(x) = nx^{n-1}$$

Proof 3

In this proof we no longer need to restrict n to be a positive integer. It can now be any real number. However, this proof also assumes that you've read all the way through the Derivative chapter. In particular it needs both Implicit Differentiation and Logarithmic Differentiation. If you've not read, and understand, these sections then this proof will not make any sense to you.

So, to get set up for logarithmic differentiation let's first define $y = x^n$ then take the log of both sides, simplify the right side using logarithm properties and then differentiate using implicit differentiation.

$$\ln(y) = \ln(x^n)$$
$$\ln(y) = n\ln(x)$$
$$\frac{y'}{y} = n\frac{1}{x}$$

Finally, all we need to do is solve for y' and then substitute in for y.

$$y' = y\frac{n}{x} = x^n\left(\frac{n}{x}\right) = nx^{n-1}$$

Before moving onto the next proof, let's notice that in all three proofs we did require that the exponent, n, be a number (integer in the first two, any real number in the third). In the first proof we couldn't have used the Binomial Theorem if the exponent wasn't a positive integer. In the second proof we couldn't have factored $x^n - a^n$ if the exponent hadn't been a positive integer. Finally, in the third proof we would have gotten a much different derivative if n had not been a constant.

This is important because people will often misuse the power rule and use it even when the exponent is not a number and/or the base is not a variable.

Product Rule

$$\big(f(x)\,g(x)\big)' = f'(x)\,g(x) + f(x)\,g'(x)$$

As with the Power Rule above, the Product Rule can be proved either by using the definition of the derivative or it can be proved using Logarithmic Differentiation. We'll show both proofs here.

Proof 1

This proof can be a little tricky when you first see it so let's be a little careful here. We'll first use the definition of the derivative on the product.

$$\big(f(x)\,g(x)\big)' = \lim_{h\to 0}\frac{f(x+h)\,g(x+h) - f(x)\,g(x)}{h}$$

On the surface this appears to do nothing for us. We'll first need to manipulate things a little to get the proof going. What we'll do is subtract out and add in $f(x+h)\,g(x)$ to the numerator. Note that we're really just adding in a zero here since these two terms will cancel. This will give us,

$$\left(f(x)\,g(x)\right)' = \lim_{h \to 0} \frac{f(x+h)\,g(x+h) - f(x+h)\,g(x) + f(x+h)\,g(x) - f(x)\,g(x)}{h}$$

Notice that we added the two terms into the middle of the numerator. As written we can break up the limit into two pieces. From the first piece we can factor a $f(x+h)$ out and we can factor a $g(x)$ out of the second piece. Doing this gives,

$$\begin{aligned}\left(f(x)\,g(x)\right)' &= \lim_{h \to 0} \frac{f(x+h)\left(g(x+h) - g(x)\right)}{h} + \lim_{h \to 0} \frac{g(x)\left(f(x+h) - f(x)\right)}{h} \\ &= \lim_{h \to 0} f(x+h)\,\frac{g(x+h) - g(x)}{h} + \lim_{h \to 0} g(x)\,\frac{f(x+h) - f(x)}{h}\end{aligned}$$

At this point we can use limit properties to write,

$$\begin{aligned}\left(f(x)\,g(x)\right)' = \left(\lim_{h \to 0} f(x+h)\right)\left(\lim_{h \to 0} \frac{g(x+h) - g(x)}{h}\right) + \\ \left(\lim_{h \to 0} g(x)\right)\left(\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}\right)\end{aligned}$$

The individual limits in here are,

$$\lim_{h \to 0} \frac{g(x+h) - g(x)}{h} = g'(x) \qquad \lim_{h \to 0} g(x) = g(x)$$

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x) \qquad \lim_{h \to 0} f(x+h) = f(x)$$

The two limits on the left are nothing more than the definition the derivative for $g(x)$ and $f(x)$ respectively. The upper limit on the right seems a little tricky but remember that the limit of a constant is just the constant. In this case since the limit is only concerned with allowing h to go to zero. The key here is to recognize that changing h will not change x and so as far as this limit is concerned $g(x)$ is a constant. Note that the function is probably not a constant, however as far as the limit is concerned the function can be treated as a constant. We get the lower limit on the right we get simply by plugging $h = 0$ into the function

Plugging all these into the last step gives us,

$$\left(f(x)\,g(x)\right)' = f(x)\,g'(x) + g(x)\,f'(x)$$

Proof 2

This is a much quicker proof but does presuppose that you've read and understood the Implicit Differentiation and Logarithmic Differentiation sections. If you haven't then this proof will not make a lot of sense to you.

First write call the product y and take the log of both sides and use a property of logarithms on the right side.

$$\begin{aligned} y &= f(x)\, g(x) \\ \ln(y) &= \ln(f(x)\, g(x)) = \ln f(x) + \ln g(x) \end{aligned}$$

Next, we take the derivative of both sides and solve for y'.

$$\frac{y'}{y} = \frac{f'(x)}{f(x)} + \frac{g'(x)}{g(x)} \qquad \Rightarrow \qquad y' = y\left(\frac{f'(x)}{f(x)} + \frac{g'(x)}{g(x)}\right)$$

Finally, all we need to do is plug in for y and then multiply this through the parenthesis and we get the Product Rule.

$$y' = f(x)\, g(x)\left(\frac{f'(x)}{f(x)} + \frac{g'(x)}{g(x)}\right) \qquad \Rightarrow \qquad \big(f(x)\, g(x)\big)' = g(x)\, f'(x) + f(x)\, g'(x)$$

Quotient Rule

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)\, g(x) - f(x)\, g'(x)}{\left[g(x)\right]^2}$$

Again, we can do this using the definition of the derivative or with Logarithmic Definition.

Proof 1

First plug the quotient into the definition of the derivative and rewrite the quotient a little.

$$\begin{aligned} \left(\frac{f(x)}{g(x)}\right)' &= \lim_{h\to 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h} \\ &= \lim_{h\to 0} \frac{1}{h}\, \frac{f(x+h)\, g(x) - f(x)\, g(x+h)}{g(x+h)\, g(x)} \end{aligned}$$

To make our life a little easier we moved the h in the denominator of the first step out to the front as a $\frac{1}{h}$. We also wrote the numerator as a single rational expression. This step is required to make this proof work.

Now, for the next step will need to subtract out and add in $f(x)\,g(x)$ to the numerator.

$$\left(\frac{f(x)}{g(x)}\right)' = \lim_{h\to 0}\frac{1}{h}\,\frac{f(x+h)\,g(x) - f(x)\,g(x) + f(x)\,g(x) - f(x)\,g(x+h)}{g(x+h)\,g(x)}$$

The next step is to rewrite things a little,

$$\left(\frac{f(x)}{g(x)}\right)' = \lim_{h\to 0}\frac{1}{g(x+h)\,g(x)}\,\frac{f(x+h)\,g(x) - f(x)\,g(x) + f(x)\,g(x) - f(x)\,g(x+h)}{h}$$

Note that all we did was interchange the two denominators. Since we are multiplying the fractions we can do this.

Next, the larger fraction can be broken up as follows.

$$\left(\frac{f(x)}{g(x)}\right)' = \lim_{h\to 0}\frac{1}{g(x+h)\,g(x)}\left(\frac{f(x+h)\,g(x) - f(x)\,g(x)}{h} + \frac{f(x)\,g(x) - f(x)\,g(x+h)}{h}\right)$$

In the first fraction we will factor a $g(x)$ out and in the second we will factor a $-f(x)$ out. This gives,

$$\left(\frac{f(x)}{g(x)}\right)' = \lim_{h\to 0}\frac{1}{g(x+h)\,g(x)}\left(g(x)\,\frac{f(x+h) - f(x)}{h} - f(x)\,\frac{g(x+h) - g(x)}{h}\right)$$

We can now use the basic properties of limits to write this as,

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{1}{\lim\limits_{h\to 0} g(x+h)\,\lim\limits_{h\to 0} g(x)}\left(\left(\lim_{h\to 0} g(x)\right)\left(\lim_{h\to 0}\frac{f(x+h) - f(x)}{h}\right) - \left(\lim_{h\to 0} f(x)\right)\left(\lim_{h\to 0}\frac{g(x+h) - g(x)}{h}\right)\right)$$

The individual limits are,

$$\lim_{h\to 0}\frac{g(x+h) - g(x)}{h} = g'(x) \qquad \lim_{h\to 0} g(x+h) = g(x) \qquad \lim_{h\to 0} g(x) = g(x)$$

$$\lim_{h\to 0}\frac{f(x+h) - f(x)}{h} = f'(x) \qquad \lim_{h\to 0} f(x) = f(x)$$

The first two limits in each row are nothing more than the definition the derivative for $g(x)$ and $f(x)$ respectively. The middle limit in the top row we get simply by plugging in $h = 0$. The final limit in each row may seem a little tricky. Recall that the limit of a constant is just the constant. Well since the limit is only concerned with allowing h to go to zero as far as its concerned $g(x)$ and $f(x)$ are constants since changing h will not change x. Note that the function is probably not a constant, however as far as the limit is concerned the function can be treated as a constant.

Plugging in the limits and doing some rearranging gives,

$$\begin{aligned}\left(\frac{f(x)}{g(x)}\right)' &= \frac{1}{g(x)\,g(x)}\left(g(x)\,f'(x) - f(x)\,g'(x)\right)\\ &= \frac{f'(x)\,g(x) - f(x)\,g'(x)}{\left[g(x)\right]^2}\end{aligned}$$

There's the quotient rule.

Proof 2

Now let's do the proof using Logarithmic Differentiation. We'll first call the quotient y, take the log of both sides and use a property of logs on the right side.

$$\begin{aligned}y &= \frac{f(x)}{g(x)}\\ \ln y &= \ln\left(\frac{f(x)}{g(x)}\right) = \ln f(x) - \ln g(x)\end{aligned}$$

Next, we take the derivative of both sides and solve for y'.

$$\frac{y'}{y} = \frac{f'(x)}{f(x)} - \frac{g'(x)}{g(x)} \qquad \Rightarrow \qquad y' = y\left(\frac{f'(x)}{f(x)} - \frac{g'(x)}{g(x)}\right)$$

Next, plug in y and do some simplification to get the quotient rule.

$$\begin{aligned}y' &= \frac{f(x)}{g(x)}\left(\frac{f'(x)}{f(x)} - \frac{g'(x)}{g(x)}\right)\\ &= \frac{f'(x)}{g(x)} - \frac{g'(x)\,f(x)}{\left[g(x)\right]^2}\\ &= \frac{f'(x)\,g(x)}{\left[g(x)\right]^2} - \frac{f(x)\,g'(x)}{\left[g(x)\right]^2} = \frac{f'(x)\,g(x) - f(x)\,g'(x)}{\left[g(x)\right]^2}\end{aligned}$$

Chain Rule

If $f(x)$ and $g(x)$ are both differentiable functions and we define $F(x) = (f \circ g)(x)$ then the derivative of *F(x)* is $F'(x) = f'(g(x))\ \ g'(x)$.

Proof

We'll start off the proof by defining $u = g(x)$ and noticing that in terms of this definition what we're being asked to prove is,

$$\frac{d}{dx}[f(u)] = f'(u)\frac{du}{dx}$$

Let's take a look at the derivative of $u(x)$ (again, remember we've defined $u = g(x)$ and so u really is a function of x) which we know exists because we are assuming that $g(x)$ is differentiable. By definition we have,

$$u'(x) = \lim_{h \to 0}\frac{u(x+h) - u(x)}{h}$$

Note as well that,

$$\lim_{h \to 0}\left(\frac{u(x+h) - u(x)}{h} - u'(x)\right) = \lim_{h \to 0}\frac{u(x+h) - u(x)}{h} - \lim_{h \to 0}u'(x) = u'(x) - u'(x) = 0$$

Now, define,

$$v(h) = \begin{cases} \dfrac{u(x+h) - u(x)}{h} - u'(x) & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

and notice that $\lim\limits_{h \to 0} v(h) = 0 = v(0)$ and so $v(h)$ is continuous at $h = 0$

Now if we assume that $h \neq 0$ we can rewrite the definition of $v(h)$ to get,

$$u(x+h) = u(x) + h\left(v(h) + u'(x)\right) \tag{A.1}$$

Now, notice that Equation A.1 is in fact valid even if we let $h = 0$ and so is valid for any value of h.

Next, since we also know that $f(x)$ is differentiable we can do something similar. However, we're going to use a different set of letters/variables here for reasons that will be apparent in a bit. So, define,

$$w(k) = \begin{cases} \dfrac{f(z+k) - f(z)}{k} - f'(z) & \text{if } k \neq 0 \\ 0 & \text{if } k = 0 \end{cases}$$

we can go through a similar argument that we did above so show that $w(k)$ is continuous at $k = 0$ and that,

$$f(z+k) = f(z) + k\left(w(k) + f'(z)\right) \tag{A.2}$$

Do not get excited about the different letters here all we did was use k instead of h and let $x = z$. Nothing fancy here, but the change of letters will be useful down the road.

Okay, to this point it doesn't look like we've really done anything that gets us even close to proving the chain rule. The work above will turn out to be very important in our proof however so let's get going on the proof.

What we need to do here is use the definition of the derivative and evaluate the following limit.

$$\frac{d}{dx}\left[f\left[u(x)\right]\right] = \lim_{h\to 0}\frac{f\left[u(x+h)\right] - f\left[u(x)\right]}{h} \tag{A.3}$$

Note that even though the notation is more than a little messy if we use $u(x)$ instead of u we need to remind ourselves here that u really is a function of x.

Let's now use Equation A.1 to rewrite the $u(x+h)$ and yes the notation is going to be unpleasant but we're going to have to deal with it. By using Equation A.1, the numerator in the limit above becomes,

$$f\left[u(x+h)\right] - f\left[u(x)\right] = f\left[u(x) + h\left(v(h) + u'(x)\right)\right] - f\left[u(x)\right]$$

If we then define $z = u(x)$ and $k = h\left(v(h) + u'(x)\right)$ we can use Equation A.2 to further write this as,

$$\begin{aligned} f\left[u(x+h)\right] - f\left[u(x)\right] &= f\left[u(x) + h\left(v(h) + u'(x)\right)\right] - f\left[u(x)\right] \\ &= f\left[u(x)\right] + h\left(v(h) + u'(x)\right)\left(w(k) + f'\left[u(x)\right]\right) - f\left[u(x)\right] \\ &= h\left(v(h) + u'(x)\right)\left(w(k) + f'\left[u(x)\right]\right) \end{aligned}$$

Notice that we were able to cancel a $f\left[u(x)\right]$ to simplify things up a little. Also, note that the $w(k)$ was intentionally left that way to keep the mess to a minimum here, just remember that $k = h\left(v(h) + u'(x)\right)$ here as that will be important here in a bit. Let's now go back and remember that all this was the numerator of our limit, Equation A.3. Plugging this into Equation A.3 gives,

$$\begin{aligned} \frac{d}{dx}\left[f\left[u(x)\right]\right] &= \lim_{h\to 0}\frac{h\left(v(h) + u'(x)\right)\left(w(k) + f'\left[u(x)\right]\right)}{h} \\ &= \lim_{h\to 0}\left(v(h) + u'(x)\right)\left(w(k) + f'\left[u(x)\right]\right) \end{aligned}$$

Notice that the h's canceled out. Next, recall that $k = h\left(v(h) + u'(x)\right)$ and so,

$$\lim_{h\to 0} k = \lim_{h\to 0} h\left(v(h) + u'(x)\right) = 0$$

But, if $\lim\limits_{h \to 0} k = 0$, as we've defined k anyway, then by the definition of w and the fact that we know $w(k)$ is continuous at $k = 0$ we also know that,

$$\lim_{h \to 0} w(k) = w\left(\lim_{h \to 0} k\right) = w(0) = 0$$

Also, recall that $\lim\limits_{h \to 0} v(h) = 0$. Using all of these facts our limit becomes,

$$\begin{aligned}
\frac{d}{dx}\left[f\left[u(x)\right]\right] &= \lim_{h \to 0} \big(v(h) + u'(x)\big)\big(w(k) + f'\left[u(x)\right]\big) \\
&= u'(x)\, f'\left[u(x)\right] \\
&= f'\left[u(x)\right]\frac{du}{dx}
\end{aligned}$$

This is exactly what we needed to prove and so we're done.

A.3 Proof of Trig Limits

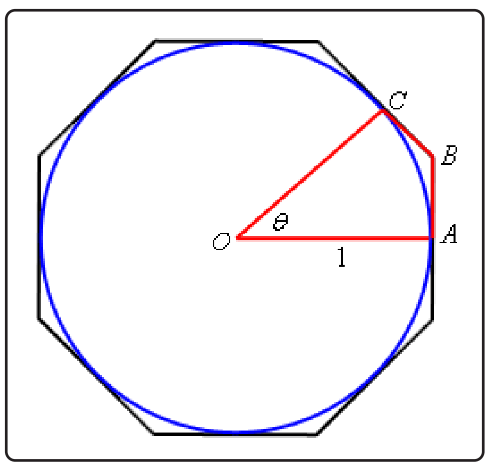
In this section we're going to provide the proof of the two limits that are used in the derivation of the derivative of sine and cosine in the Derivatives of Trig Functions section of the Derivatives chapter.

Proof of : $\lim\limits_{\theta \to 0} \dfrac{\sin(\theta)}{\theta} = 1$

This proof of this limit uses the Squeeze Theorem. However, getting things set up to use the Squeeze Theorem can be a somewhat complex geometric argument that can be difficult to follow so we'll try to take it fairly slow.

Let's start by assuming that $0 \le \theta \le \frac{\pi}{2}$. Since we are proving a limit that has $\theta \to 0$ it's okay to assume that θ is not too large (*i.e.* $\theta \le \frac{\pi}{2}$). Also, by assuming that θ is positive we're actually going to first prove that the above limit is true if it is the right-hand limit. As you'll see if we can prove this then the proof of the limit will be easy.

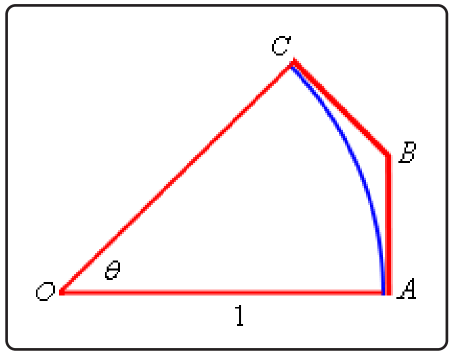
So, now that we've got our assumption on θ taken care of let's start off with the unit circle circumscribed by an octagon with a small slice marked out as shown below.



Points A and C are the midpoints of their respective sides on the octagon and are in fact tangent to the circle at that point. We'll call the point where these two sides meet B.

From this figure we can see that the circumference of the circle is less than the length of the octagon. This also means that if we look at the slice of the figure marked out above then the length of the portion of the circle included in the slice must be less than the length of the portion of the octagon included in the slice.

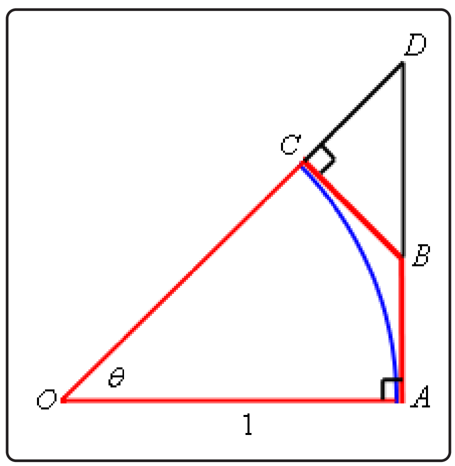
Because we're going to be doing most of our work on just the slice of the figure let's strip that out and look at just it. Here is a sketch of just the slice.



Now denote the portion of the circle by arc AC and the lengths of the two portion of the octagon shown by $|AB|$ and $|BC|$. Then by the observation about lengths we made above we must have,

$$\text{arc } AC < |AB| + |BC| \tag{A.4}$$

Next, extend the lines AB and OC as shown below and call the point that they meet D. The triangle now formed by AOD is a right triangle. All this is shown in the figure below.



The triangle BCD is a right triangle with hypotenuse BD and so we know $|BC| < |BD|$. Also notice that $|AB|+|BD| = |AD|$. If we use these two facts in Equation A.4 we get,

$$\begin{aligned} \text{arc } AC &< |AB| + |BC| \\ &< |AB| + |BD| \\ &= |AD| \end{aligned} \tag{A.5}$$

Next, as noted already the triangle AOD is a right triangle and so we can use a little right triangle trigonometry to write $|AD| = |AO| \tan(\theta)$. Also note that $|AO| = 1$ since it is nothing more than the radius of the unit circle. Using this information in Equation A.5 gives,

$$\begin{aligned}\text{arc } AC &< |AD| \\ &< |AO| \tan(\theta) \\ &= \tan\theta \end{aligned} \tag{A.6}$$

The next thing that we need to recall is that the length of a portion of a circle is given by the radius of the circle times the angle (in radians!) that traces out the portion of the circle we're trying to measure. For our portion this means that,

$$\text{arc } AC = |AO|\,\theta = \theta$$

Before proceeding a quick note. Students often ask why we always use radians in a Calculus class. This is the reason why! The formula for the length of a portion of a circle used above assumed that the angle is in radians. The formula for angles in degrees is different and if we used that we would get a different answer. So, remember to always use radians.

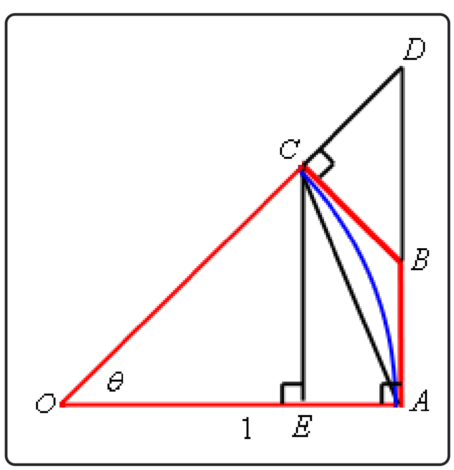
So, putting this into Equation A.6 we see that,

$$\theta = \text{arc } AC < \tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}$$

or, if we do a little rearranging we get,

$$\cos(\theta) < \frac{\sin(\theta)}{\theta} \tag{A.7}$$

We'll be coming back to Equation A.7 in a bit. Let's now add in a couple more lines into our figure above. Let's connect A and C with a line and drop a line straight down from C until it intersects AO at a right angle and let's call the intersection point E. This is all shown in the figure below.



Okay, the first thing to notice here is that,

$$|CE| < |AC| < \text{arc } AC \tag{A.8}$$

Also note that triangle EOC is a right triangle with a hypotenuse of $|CO| = 1$. Using some right triangle trig we can see that,

$$|CE| = |CO| \sin(\theta) = \sin(\theta)$$

Plugging this into Equation A.8 and recalling that arc $AC = \theta$ we get,

$$\sin(\theta) = |CE| < \text{arc } AC = \theta$$

and with a little rewriting we get,

$$\frac{\sin(\theta)}{\theta} < 1 \tag{A.9}$$

Okay, we're almost done here. Putting Equation A.7 and Equation A.9 together we see that,

$$\cos(\theta) < \frac{\sin(\theta)}{\theta} < 1$$

provided $0 \le \theta \le \frac{\pi}{2}$. Let's also note that,

$$\lim_{\theta \to 0} \cos(\theta) = 1 \qquad\qquad \lim_{\theta \to 0} 1 = 1$$

We are now set up to use the Squeeze Theorem. The only issue that we need to worry about is that we are staying to the right of $\theta = 0$ in our assumptions and so the best that the Squeeze Theorem will tell us is,

$$\lim_{\theta \to 0^+} \frac{\sin(\theta)}{\theta} = 1$$

So, we know that the limit is true if we are only working with a right-hand limit. However we know that $\sin(\theta)$ is an odd function and so,

$$\frac{\sin(-\theta)}{-\theta} = \frac{-\sin(\theta)}{-\theta} = \frac{\sin(\theta)}{\theta}$$

In other words, if we approach zero from the left (*i.e.* negative θ's) then we'll get the same values in the function as if we'd approached zero from the right (*i.e.* positive θ's) and so,

$$\lim_{\theta \to 0^-} \frac{\sin(\theta)}{\theta} = 1$$

We have now shown that the two one-sided limits are the same and so we must also have,

$$\lim_{\theta \to 0} \frac{\sin(\theta)}{\theta} = 1$$

That was a somewhat long proof and if you're not really good at geometric arguments it can be kind of daunting and confusing. Nicely, the second limit is very simple to prove, provided you've already proved the first limit.

Proof of : $\lim\limits_{\theta \to 0} \dfrac{\cos(\theta) - 1}{\theta} = 0$

We'll start by doing the following,

$$\lim_{\theta \to 0} \frac{\cos(\theta) - 1}{\theta} = \lim_{\theta \to 0} \frac{(\cos(\theta) - 1)(\cos(\theta) + 1)}{\theta (\cos(\theta) + 1)} = \lim_{\theta \to 0} \frac{\cos^2\theta - 1}{\theta (\cos(\theta) + 1)} \tag{A.10}$$

Now, let's recall that,

$$\cos^2\theta + \sin^2\theta = 1 \qquad \Rightarrow \qquad \cos^2\theta - 1 = -\sin^2\theta$$

Using this in Equation A.10 gives us,

$$\begin{aligned} \lim_{\theta \to 0} \frac{\cos(\theta) - 1}{\theta} &= \lim_{\theta \to 0} \frac{-\sin^2\theta}{\theta (\cos(\theta) + 1)} \\ &= \lim_{\theta \to 0} \frac{\sin(\theta)}{\theta} \frac{-\sin(\theta)}{\cos(\theta) + 1} \\ &= \lim_{\theta \to 0} \frac{\sin(\theta)}{\theta} \lim_{\theta \to 0} \frac{-\sin(\theta)}{\cos(\theta) + 1} \end{aligned}$$

At this point, because we just proved the first limit and the second can be taken directly we're pretty much done. All we need to do is take the limits.

$$\lim_{\theta \to 0} \frac{\cos(\theta) - 1}{\theta} = \lim_{\theta \to 0} \frac{\sin(\theta)}{\theta} \lim_{\theta \to 0} \frac{-\sin(\theta)}{\cos(\theta) + 1} = (1)(0) = 0$$

A.4 Proofs of Derivative Applications Facts

In this section we'll be proving some of the facts and/or theorems from the Applications of Derivatives chapter. Not all of the facts and/or theorems will be proved here.

Fermat's Theorem

If $f(x)$ has a relative extrema at $x = c$ and $f'(c)$ exists then $x = c$ is a critical point of $f(x)$. In fact, it will be a critical point such that $f'(c) = 0$.

Proof

This is a fairly simple proof. We'll assume that $f(x)$ has a relative maximum to do the proof. The proof for a relative minimum is nearly identical. So, if we assume that we have a relative maximum at $x = c$ then we know that $f(c) \geq f(x)$ for all x that are sufficiently close to $x = c$. In particular for all h that are sufficiently close to zero (positive or negative) we must have,

$$f(c) \geq f(c+h)$$

or, with a little rewrite we must have,

$$f(c+h) - f(c) \leq 0 \tag{A.11}$$

Now, at this point assume that $h > 0$ and divide both sides of Equation A.11 by h. This gives,

$$\frac{f(c+h) - f(c)}{h} \leq 0$$

Because we're assuming that $h > 0$ we can now take the right-hand limit of both sides of this.

$$\lim_{h \to 0^+} \frac{f(c+h) - f(c)}{h} \leq \lim_{h \to 0^+} 0 = 0$$

We are also assuming that $f'(c)$ exists and recall that if a normal limit exists then it must be equal to both one-sided limits. We can then say that,

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = \lim_{h \to 0^+} \frac{f(c+h) - f(c)}{h} \leq 0$$

If we put this together we have now shown that $f'(c) \leq 0$.

Okay, now let's turn things around and assume that $h < 0$ and divide both sides of Equation A.11 by h. This gives,

$$\frac{f(c+h) - f(c)}{h} \geq 0$$

Remember that because we're assuming $h < 0$ we'll need to switch the inequality when we divide by a negative number. We can now do a similar argument as above to get that,

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = \lim_{h \to 0^-} \frac{f(c+h) - f(c)}{h} \geq \lim_{h \to 0^-} 0 = 0$$

The difference here is that this time we're going to be looking at the left-hand limit since we're assuming that $h < 0$. This argument shows that $f'(c) \geq 0$.

We've now shown that $f'(c) \leq 0$ and $f'(c) \geq 0$. Then only way both of these can be true at the same time is to have $f'(c) = 0$ and this in turn means that $x = c$ must be a critical point.

As noted above, if we assume that $f(x)$ has a relative minimum then the proof is nearly identical and so isn't shown here. The main differences are simply some inequalities need to be switched.

Fact, The Shape of a Graph, Part I

1. If $f'(x) > 0$ for every x on some interval I, then $f(x)$ is increasing on the interval.
2. If $f'(x) < 0$ for every x on some interval I, then $f(x)$ is decreasing on the interval.
3. If $f'(x) = 0$ for every x on some interval I, then $f(x)$ is constant on the interval.

The proof of this fact uses the Mean Value Theorem which, if you're following along in my notes has actually not been covered yet. The Mean Value Theorem can be covered at any time and for whatever the reason I decided to put where it is. Before reading through the proof of this fact you should take a quick look at the Mean Value Theorem section. You really just need the conclusion of the Mean Value Theorem for this proof however.

Proof of 1

Let x_1 and x_2 be in I and suppose that $x_1 < x_2$. Now, using the Mean Value Theorem on $[x_1, x_2]$ means there is a number c such that $x_1 < c < x_2$ and,

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1)$$

Because $x_1 < c < x_2$ we know that c must also be in I and so we know that $f'(c) > 0$ we also know that $x_2 - x_1 > 0$. So, this means that we have,

$$f(x_2) - f(x_1) > 0$$

Rewriting this gives,

$$f(x_1) < f(x_2)$$

and so, by definition, since x_1 and x_2 were two arbitrary numbers in I, $f(x)$ must be increasing on I.

Proof of 2

This proof is nearly identical to the previous part.

Let x_1 and x_2 be in I and suppose that $x_1 < x_2$. Now, using the Mean Value Theorem on $[x_1, x_2]$ means there is a number c such that $x_1 < c < x_2$ and,

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1)$$

Because $x_1 < c < x_2$ we know that c must also be in I and so we know that $f'(c) < 0$ we also know that $x_2 - x_1 > 0$. So, this means that we have,

$$f(x_2) - f(x_1) < 0$$

Rewriting this gives,

$$f(x_1) > f(x_2)$$

and so, by definition, since x_1 and x_2 were two arbitrary numbers in I, $f(x)$ must be decreasing on I.

Proof of 3

Again, this proof is nearly identical to the previous two parts, but in this case is actually somewhat easier.

Let x_1 and x_2 be in I. Now, using the Mean Value Theorem on $[x_1, x_2]$ there is a number c such that c is between x_1 and x_2 and,

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1)$$

Note that for this part we didn't need to assume that $x_1 < x_2$ and so all we know is that c is between x_1 and x_2 and so, more importantly, c is also in I. and this means that $f'(c) = 0$. So, this means that we have,

$$f(x_2) - f(x_1) = 0$$

Rewriting this gives,

$$f(x_1) = f(x_2)$$

and so, since x_1 and x_2 were two arbitrary numbers in I, $f(x)$ must be constant on I.

Fact, The Shape of a Graph, Part II

Given the function $f(x)$ then,

1. If $f''(x) > 0$ for all x in some interval I then $f(x)$ is concave up on I.
2. If $f''(x) < 0$ for all x in some interval I then $f(x)$ is concave down on I.

The proof of this fact uses the Mean Value Theorem which, if you're following along in my notes has actually not been covered yet. The Mean Value Theorem can be covered at any time and for whatever the reason I decided to put it after the section this fact is in. Before reading through the proof of this fact you should take a quick look at the Mean Value Theorem section. You really just need the conclusion of the Mean Value Theorem for this proof however.

Proof of 1

Let a be any number in the interval I. The tangent line to $f(x)$ at $x = a$ is,

$$y = f(a) + f'(a)(x - a)$$

To show that $f(x)$ is concave up on I then we need to show that for any x, $x \neq a$, in I that,

$$f(x) > f(a) + f'(a)(x - a)$$

or in other words, the tangent line is always below the graph of $f(x)$ on I. Note that we require $x \neq a$ because at that point we know that $f(x) = f(a)$ since we are talking about the tangent line.

Let's start the proof off by first assuming that $x > a$. Using the Mean Value Theorem on $[a, x]$ means there is a number c such that $a < c < x$ and,

$$f(x) - f(a) = f'(c)(x - a)$$

With some rewriting this is,

$$f(x) = f(a) + f'(c)(x - a) \tag{A.12}$$

Next, let's use the fact that $f''(x) > 0$ for every x on I. This means that the first derivative, $f'(x)$, must be increasing (because its derivative, $f''(x)$, is positive). Now, we know from the Mean Value Theorem that $a < c$ and so because $f'(x)$ is increasing we must have,

$$f'(a) < f'(c) \tag{A.13}$$

Recall as well that we are assuming $x > a$ and so $x-a > 0$. If we now multiply Equation A.13 by $x - a$ (which is positive and so the inequality stays the same) we get,

$$f'(a)(x-a) < f'(c)(x-a)$$

Next, add $f(a)$ to both sides of this to get,

$$f(a) + f'(a)(x-a) < f(a) + f'(c)(x-a)$$

However, by Equation A.12, the right side of this is nothing more than $f(x)$ and so we have,

$$f(a) + f'(a)(x-a) < f(x)$$

but this is exactly what we wanted to show.

So, provided $x > a$ the tangent line is in fact below the graph of $f(x)$.

We now need to assume $x < a$. Using the Mean Value Theorem on $[x, a]$ means there is a number c such that $x < c < a$ and,

$$f(a) - f(x) = f'(c)(a-x)$$

If we multiply both sides of this by -1 and then adding $f(a)$ to both sides and we again arise at Equation A.12.

Now, from the Mean Value Theorem we know that $c < a$ and because $f''(x) > 0$ for every x on I we know that the derivative is still increasing and so we have,

$$f'(c) < f'(a)$$

Let's now multiply this by $x - a$, which is now a negative number since $x < a$. This gives,

$$f'(c)(x-a) > f'(a)(x-a)$$

Notice that we had to switch the direction of the inequality since we were multiplying by a negative number. If we now add $f(a)$ to both sides of this and then substitute Equation A.12 into the results we arrive at,

$$\begin{aligned} f(a) + f'(c)(x-a) &> f(a) + f'(a)(x-a) \\ f(x) &> f(a) + f'(a)(x-a) \end{aligned}$$

So, again we've shown that the tangent line is always below the graph of $f(x)$.

We've now shown that if x is any number in I, with $x \neq a$ the tangent lines are always below the graph of $f(x)$ on I and so $f(x)$ is concave up on I.

Proof of 2

This proof is nearly identical to the proof of 1 and since that proof is fairly long we're going to just get things started and then leave the rest of it to you to go through.

Let a be any number in I . To show that $f(x)$ is concave down we need to show that for any x in I, $x \neq a$, that the tangent line is always above the graph of $f(x)$ or,

$$f(x) < f(a) + f'(a)(x-a)$$

From this point on the proof is almost identical to the proof of 1 except that you'll need to use the fact that the derivative in this case is decreasing since $f''(x) < 0$. We'll leave it to you to fill in the details of this proof.

Second Derivative Test

Suppose that $x = c$ is a critical point of $f'(c)$ such that $f'(c) = 0$ and that $f''(x)$ is continuous in a region around $x = c$. Then,

1. If $f''(c) < 0$ then $x = c$ is a relative maximum.
2. If $f''(c) > 0$ then $x = c$ is a relative minimum.
3. If $f''(c) = 0$ then $x = c$ can be a relative maximum, relative minimum or neither.

The proof of this fact uses the Mean Value Theorem which, if you're following along in my notes has actually not been covered yet. The Mean Value Theorem can be covered at any time and for whatever the reason I decided to put it after the section this fact is in. Before reading through the proof of this fact you should take a quick look at the Mean Value Theorem section. You really just need the conclusion of the Mean Value Theorem for this proof however.

Proof of 1

First since we are assuming that $f''(x)$ is continuous in a region around $x = c$ then we can assume that in fact $f''(c) < 0$ is also true in some open region, say (a, b) around $x = c$, *i.e.* $a < c < b$.

Now let x be any number such that $a < x < c$, we're going to use the Mean Value Theorem on $[x, c]$. However, instead of using it on the function itself we're going to use it on the first derivative. So, the Mean Value Theorem tells us that there is a number $x < d < c$ such that,

$$f'(c) - f'(x) = f''(d)(c - x)$$

Now, because $a < x < d < c$ we know that $f''(d) < 0$ and we also know that $c - x > 0$ so we then get that,

$$f'(c) - f'(x) < 0$$

However, we also assumed that $f'(c) = 0$ and so we have,

$$-f'(x) < 0 \qquad \Rightarrow \qquad f'(x) > 0$$

Or, in other words to the left of $x = c$ the function is increasing.

Let's now turn things around and let x be any number such that $c < x < b$ and use the Mean Value Theorem on $[c, x]$ and the first derivative. The Mean Value Theorem tells us that there is a number $c < d < x$ such that,

$$f'(x) - f'(c) = f''(d)(x - c)$$

Now, because $c < d < x < b$ we know that $f''(d) < 0$ and we also know that $x - c > 0$ so we then get that,

$$f'(x) - f'(c) < 0$$

Again, use the fact that we also assumed that $f'(c) = 0$ to get,

$$f'(x) < 0$$

We now know that to the right of $x = c$ the function is decreasing.

So, to the left of $x = c$ the function is increasing and to the right of $x = c$ the function is decreasing so by the first derivative test this means that $x = c$ must be a relative maximum.

Proof of 2

This proof is nearly identical to the proof of 1 and since that proof is somewhat long we're going to leave the proof to you to do. In this case the only difference is that now we are going to assume that $f''(x) < 0$ and that will give us the opposite signs of the first derivative on either side of $x = c$ which gives us the conclusion we were after. We'll leave it to you to fill in all the details of this.

Proof of 3

There isn't really anything to prove here. All this statement says is that any of the three cases are possible and to "prove" this all one needs to do is provide an example of each of the three cases. This was done in The Shape of a Graph, Part II section where this test was presented so we'll leave it to you to go back to that section to see those graphs to verify that all three possibilities really can happen.

Rolle's Theorem

Suppose $f(x)$ is a function that satisfies all of the following.

1. $f(x)$ is continuous on the closed interval $[a, b]$.
2. $f(x)$ is differentiable on the open interval (a, b).
3. $f(a) = f(b)$

Then there is a number c such that $a < c < b$ and $f'(x) = 0$. Or, in other words, $f(x)$ has a critical point in (a, b).

Proof

We'll need to do this with 3 cases.

Case 1 : $f(x) = k$ on $[a, b]$ where k is a constant.

In this case $f'(x) = 0$ for all x in $[a, b]$ and so we can take c to be any number in $[a, b]$.

Case 2 : There is some number d in (a, b) such that $f(d) > f(a)$.

Because $f(x)$ is continuous on $[a, b]$ by the Extreme Value Theorem we know that $f(x)$ will have a maximum somewhere in $[a, b]$. Also, because $f(a) = f(b)$ and $f(d) > f(a)$ we know that in fact the maximum value will have to occur at some c that is in the open interval (a, b), or $a < c < b$. Because c occurs in the interior of the interval this means that $f(x)$ will actually have a relative maximum at $x = c$ and by the second hypothesis above we also know that $f'(c)$ exists. Finally, by Fermat's Theorem we then know that in fact $x = c$ must be a critical point and because we know that $f'(c)$ exists we must have $f'(c) = 0$ (as opposed to $f'(c)$ not existing...).

Case 3 : There is some number d in (a, b) such that $f(d) < f(a)$.

This is nearly identical to Case 2 so we won't put in quite as much detail. By the Extreme Value Theorem $f(x)$ will have minimum in $[a, b]$ and because $f(a) = f(b)$ and $f(d) < f(a)$ we know that the minimum must occur at $x = c$ where $a < c < b$. Finally, by Fermat's Theorem we know that $f'(c) = 0$.

The Mean Value Theorem

Suppose $f(x)$ is a function that satisfies both of the following.

1. $f(x)$ is continuous on the closed interval $[a, b]$.
2. $f(x)$ is differentiable on the open interval (a, b).

Then there is a number c such that $a < c < b$ and

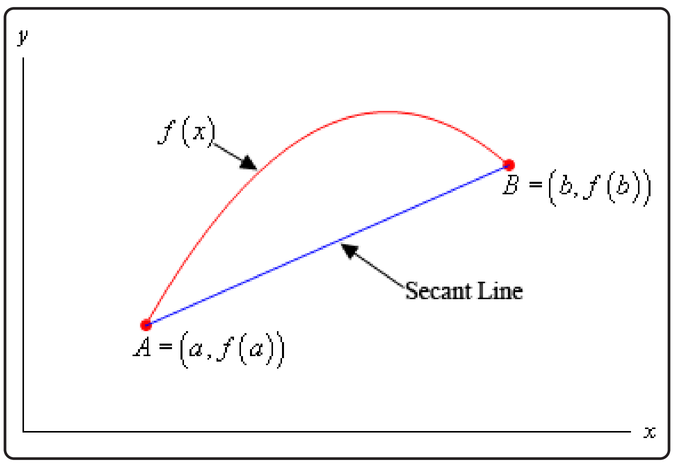
$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

Or,

$$f(b) - f(a) = f'(c)(b - a)$$

Proof

For illustration purposes let's suppose that the graph of $f(x)$ is,



Note of course that it may not look like this, but we just need a quick sketch to make it easier to see what we're talking about here.

The first thing that we need is the equation of the secant line that goes through the two points A and B as shown above. This is,

$$y = f(a) + \frac{f(b) - f(a)}{b - a}(x - a)$$

Let's now define a new function, $g(x)$, as to be the difference between $f(x)$ and the equation of the secant line or,

$$g(x) = f(x) - \left(f(a) + \frac{f(b) - f(a)}{b - a}(x - a)\right) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a)$$

Next, let's notice that because $g(x)$ is the sum of $f(x)$, which is assumed to be continuous on $[a,b]$, and a linear polynomial, which we know to be continuous everywhere, we know that $g(x)$ must also be continuous on $[a,b]$.

Also, we can see that $g(x)$ must be differentiable on (a,b) because it is the sum of $f(x)$, which is assumed to be differentiable on (a,b), and a linear polynomial, which we know to be differentiable.

We could also have just computed the derivative as follows,

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}$$

at which point we can see that it exists on (a,b) because we assumed that $f'(x)$ exists on (a,b) and the last term is just a constant.

Finally, we have,

$$g(a) = f(a) - f(a) - \frac{f(b) - f(a)}{b - a}(a - a) = f(a) - f(a) = 0$$

$$g(b) = f(b) - f(a) - \frac{f(b) - f(a)}{b - a}(b - a) = f(b) - f(a) - (f(b) - f(a)) = 0$$

In other words, $g(x)$ satisfies the three conditions of Rolle's Theorem and so we know that there must be a number c such that $a < c < b$ and that,

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} \qquad \Rightarrow \qquad f'(c) = \frac{f(b) - f(a)}{b - a}$$

A.5 Proof of Various Integral Properties

In this section we've got the proof of several of the properties we saw in the Integrals Chapter as well as a couple from the Applications of Integrals Chapter.

Proof of : $\int k\, f(x)\, dx = k \int f(x)\, dx$ where k is any number.

This is a very simple proof. Suppose that $F(x)$ is an anti-derivative of $f(x)$, *i.e.* $F'(x) = f(x)$. Then by the basic properties of derivatives we also have that,

$$\left(k\, F(x)\right)' = k\, F'(x) = k\, f(x)$$

and so $k\, F(x)$ is an anti-derivative of $k\, f(x)$, *i.e.* $(k\, F(x))' = k\, f(x)$. In other words,

$$\int k\, f(x)\, dx = k\, F(x) + c = k \int f(x)\, dx$$

Proof of : $\int f(x) \pm g(x)\, dx = \int f(x)\, dx \pm \int g(x)\, dx$

This is also a very simple proof Suppose that $F(x)$ is an anti-derivative of $f(x)$ and that $G(x)$ is an anti-derivative of $g(x)$. So we have that $F'(x) = f(x)$ and $G'(x) = g(x)$. Basic properties of derivatives also tell us that

$$\left(F(x) \pm G(x)\right)' = F'(x) \pm G'(x) = f(x) \pm g(x)$$

and so $F(x) + G(x)$ is an anti-derivative of $f(x) + g(x)$ and $F(x) - G(x)$ is an anti-derivative of $f(x) - g(x)$. In other words,

$$\int f(x) \pm g(x)\, dx = F(x) \pm G(x) + c = \int f(x)\, dx \pm \int g(x)\, dx$$

Proof of : $\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$

From the definition of the definite integral we have,

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\,\Delta x \qquad \Delta x = \frac{b-a}{n}$$

and we also have,

$$\int_b^a f(x)\,dx = \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\,\Delta x \qquad \Delta x = \frac{a-b}{n}$$

Therefore,

$$\begin{aligned}
\int_a^b f(x)\,dx &= \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\,\frac{b-a}{n} \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\,\frac{-(a-b)}{n} \\
&= \lim_{n\to\infty} \left(-\sum_{i=1}^{n} f(x_i^*)\,\frac{a-b}{n}\right) \\
&= -\lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\,\frac{a-b}{n} = -\int_b^a f(x)\,dx
\end{aligned}$$

Proof of : $\int_a^a f(x)\,dx = 0$

From the definition of the definite integral we have,

$$\begin{aligned}
\int_a^a f(x)\,dx &= \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\,\Delta x \qquad \Delta x = \frac{a-a}{n} = 0 \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)\,(0) \\
&= \lim_{n\to\infty} 0 \\
&= 0
\end{aligned}$$

Proof of : $\displaystyle\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx$

From the definition of the definite integral we have,

$$\begin{aligned}\int_a^b c\,f(x)\,dx &= \lim_{n\to\infty}\sum_{i=1}^{n} c\,f(x_i^*)\,\Delta x\\ &= \lim_{n\to\infty} c\sum_{i=1}^{n} f(x_i^*)\,\Delta x\\ &= c\lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\,\Delta x\\ &= c\int_a^b f(x)\,dx\end{aligned}$$

Remember that we can pull constants out of summations and out of limits.

Proof of : $\displaystyle\int_a^b f(x)\pm g(x)\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,d$

First we'll prove the formula for "$+$". From the definition of the definite integral we have,

$$\begin{aligned}\int_a^b f(x)+g(x)\,dx &= \lim_{n\to\infty}\sum_{i=1}^{n}\left(f(x_i^*)+g(x_i^*)\right)\Delta x\\ &= \lim_{n\to\infty}\left(\sum_{i=1}^{n} f(x_i^*)\,\Delta x + \sum_{i=1}^{n} g(x_i^*)\,\Delta x\right)\\ &= \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\,\Delta x + \lim_{n\to\infty}\sum_{i=1}^{n} g(x_i^*)\,\Delta x\\ &= \int_a^b f(x)\,dx + \int_a^b g(x)\,dx\end{aligned}$$

To prove the formula for "$-$" we can either redo the above work with a minus sign instead of a plus sign or we can use the fact that we now know this is true with a plus and using the properties proved above as follows.

$$\begin{aligned}\int_a^b f(x) - g(x)\,dx &= \int_a^b f(x) + \left(-g(x)\right)dx \\ &= \int_a^b f(x)\,dx + \int_a^b \left(-g(x)\right)dx \\ &= \int_a^b f(x)\,dx - \int_a^b g(x)\,dx\end{aligned}$$

Proof of : $\int_a^b c\,dx = c(b-a)$, c is any number.

If we define $f(x) = c$ then from the definition of the definite integral we have,

$$\begin{aligned}\int_a^b c\,dx &= \int_a^b f(x)\,dx \\ &= \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\,\Delta x \qquad \Delta x = \frac{b-a}{n} \\ &= \lim_{n\to\infty}\left(\sum_{i=1}^{n} c\right)\frac{b-a}{n} \\ &= \lim_{n\to\infty}(cn)\frac{b-a}{n} \\ &= \lim_{n\to\infty} c(b-a) \\ &= c(b-a)\end{aligned}$$

Proof of : If $f(x) \geq 0$ for $a \leq x \leq b$ then $\int_a^b f(x)\,dx \geq 0$.

From the definition of the definite integral we have,

$$\int_a^b f(x)\,dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\,\Delta x \qquad \Delta x = \frac{b-a}{n}$$

Now, by assumption $f(x) \geq 0$ and we also have $\Delta x > 0$ and so we know that

$$\sum_{i=1}^{n} f(x_i^*)\,\Delta x \geq 0$$

So, from the basic properties of limits we then have,

$$\lim_{n \to \infty} \sum_{i=1}^{n} f(x_i^*)\,\Delta x \geq \lim_{n \to \infty} 0 = 0$$

But the left side is exactly the definition of the integral and so we have,

$$\int_a^b f(x)\,dx = \lim_{n \to \infty} \sum_{i=1}^{n} f(x_i^*)\,\Delta x \geq 0$$

Proof of : If $f(x) \geq g(x)$ for $a \leq x \leq b$ then $\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx$.

Since we have $f(x) \geq g(x)$ then we know that $f(x) - g(x) \geq 0$ on $a \leq x \leq b$ and so by Property 8 proved above we know that,

$$\int_a^b f(x) - g(x)\,dx \geq 0$$

We also know from Property 4 that,

$$\int_a^b f(x) - g(x)\,dx = \int_a^b f(x)\,dx - \int_a^b g(x)\,dx$$

So, we then have,

$$\begin{aligned} \int_a^b f(x)\,dx - \int_a^b g(x)\,dx &\geq 0 \\ \int_a^b f(x)\,dx &\geq \int_a^b g(x)\,dx \end{aligned}$$

Proof of : If $m \le f(x) \le M$ for $a \le x \le b$ then $m(b-a) \le \int_a^b f(x)\,dx \le M(b-a)$.

Given $m \le f(x) \le M$ we can use Property 9 on each inequality to write,

$$\int_a^b m\,dx \le \int_a^b f(x)\,dx \le \int_a^b M\,dx$$

Then by Property 7 on the left and right integral to get,

$$m(b-a) \le \int_a^b f(x)\,dx \le M(b-a)$$

Proof of : $\left|\int_a^b f(x)\,dx\right| \le \int_a^b |f(x)|\,dx$

First let's note that we can say the following about the function and the absolute value,

$$-|f(x)| \le f(x) \le |f(x)|$$

If we now use Property 9 on each inequality we get,

$$\int_a^b -|f(x)|\,dx \le \int_a^b f(x)\,dx \le \int_a^b |f(x)|\,dx$$

We know that we can factor the minus sign out of the left integral to get,

$$-\int_a^b |f(x)|\,dx \le \int_a^b f(x)\,dx \le \int_a^b |f(x)|\,dx$$

Finally, recall that if $|p| \le b$ then $-b \le p \le b$ and of course this works in reverse as well so we then must have,

$$\left|\int_a^b f(x)\,dx\right| \le \int_a^b |f(x)|\,dx$$

Fundamental Theorem of Calculus, Part I

If $f(x)$ is continuous on $[a,b]$ then,

$$g(x) = \int_a^x f(t)\ dt$$

is continuous on $[a,b]$ and it is differentiable on (a,b) and that,

$$g'(x) = f(x)$$

Proof

Suppose that x and $x+h$ are in (a,b). We then have,

$$g(x+h) - g(x) = \int_a^{x+h} f(t)\ dt - \int_a^x f(t)\ dt$$

Now, using Property 5 of the Integral Properties we can rewrite the first integral and then do a little simplification as follows.

$$\begin{aligned} g(x+h) - g(x) &= \left(\int_a^x f(t)\ dt + \int_x^{x+h} f(t)\ dt\right) - \int_a^x f(t)\ dt \\ &= \int_x^{x+h} f(t)\ dt \end{aligned}$$

Finally assume that $h \neq 0$ and we get,

$$\frac{g(x+h) - g(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t)\ dt \tag{A.14}$$

Let's now assume that $h > 0$ and since we are still assuming that $x+h$ are in (a,b) we know that $f(x)$ is continuous on $[x, x+h]$ and so by the Extreme Value Theorem we know that there are numbers c and d in $[x, x+h]$ so that $f(c) = m$ is the absolute minimum of $f(x)$ in $[x, x+h]$ and that $f(d) = M$ is the absolute maximum of $f(x)$ in $[x, x+h]$.

So, by Property 10 of the Integral Properties we then know that we have,

$$mh \le \int_x^{x+h} f(t)\ dt \le Mh$$

Or,

$$f(c)\,h \le \int_x^{x+h} f(t)\ dt \le f(d)\,h$$

Now divide both sides of this by h to get,

$$f(c) \le \frac{1}{h}\int_x^{x+h} f(t)\ dt \le f(d)$$

and then use Equation A.14 to get,

$$f(c) \le \frac{g(x+h) - g(x)}{h} \le f(d) \tag{A.15}$$

Next, if $h < 0$ we can go through the same argument above except we'll be working on $[x+h, x]$ to arrive at exactly the same inequality above. In other words, Equation A.15 is true provided $h \ne 0$.

Now, if we take $h \to 0$ we also have $c \to x$ and $d \to x$ because both c and d are between x and $x+h$. This means that we have the following two limits.

$$\lim_{h \to 0} f(c) = \lim_{c \to x} f(c) = f(x) \qquad\qquad \lim_{h \to 0} f(d) = \lim_{d \to x} f(d) = f(x)$$

The Squeeze Theorem then tells us that,

$$\lim_{h \to 0} \frac{g(x+h) - g(x)}{h} = f(x) \tag{A.16}$$

but the left side of this is exactly the definition of the derivative of $g(x)$ and so we get that,

$$g'(x) = f(x)$$

So, we've shown that $g(x)$ is differentiable on (a, b).

Now, the Theorem at the end of the Definition of the Derivative section tells us that $g(x)$ is also continuous on (a, b). Finally, if we take $x = a$ or $x = b$ we can go through a similar argument we used to get Equation A.16 using one-sided limits to get the same result and so the theorem at the end of the Definition of the Derivative section will also tell us that $g(x)$ is continuous at $x = a$ or $x = b$ and so in fact $g(x)$ is also continuous on $[a, b]$.

Fundamental Theorem of Calculus, Part II

Suppose $f(x)$ is a continuous function on $[a, b]$ and also suppose that $F(x)$ is any anti-derivative for $f(x)$. Then,

$$\int_a^b f(x)\,dx = F(x)\Big|_a^b = F(b) - F(a)$$

Proof

First let $g(x) = \int_a^x f(t)\ dt$ and then we know from Part I of the Fundamental Theorem of Calculus that $g'(x) = f(x)$ and so $g(x)$ is an anti-derivative of $f(x)$ on $[a, b]$. Further suppose that $F(x)$ is any anti-derivative of $f(x)$ on $[a, b]$ that we want to choose. So, this means that we must have,

$$g'(x) = F'(x)$$

Then, by Fact 2 in the Mean Value Theorem section we know that $g(x)$ and $F(x)$ can differ by no more than an additive constant on (a, b). In other words, for $a < x < b$ we have,

$$F(x) = g(x) + c$$

Now because $g(x)$ and $F(x)$ are continuous on $[a, b]$, if we take the limit of this as $x \to a^+$ and $x \to b^-$ we can see that this also holds if $x = a$ and $x = b$.

So, for $a \le x \le b$ we know that $F(x) = g(x) + c$. Let's use this and the definition of $g(x)$ to do the following.

$$\begin{aligned} F(b) - F(a) &= \big(g(b) + c\big) - \big(g(a) + c\big) \\ &= g(b) - g(a) \\ &= \int_a^b f(t)\ dt + \int_a^a f(t)\ dt \\ &= \int_a^b f(t)\ dt + 0 \\ &= \int_a^b f(x)\ dx \end{aligned}$$

Note that in the last step we used the fact that the variable used in the integral does not matter and so we could change the t's to x's.

Average Function Value

The average value of a continuous function $f(x)$ over the interval $[a, b]$ is given by,

$$f_{avg} = \frac{1}{b-a}\int_a^b f(x)\ dx$$

Proof

We know that the average value of n numbers is simply the sum of all the numbers divided by n so let's start off with that. Let's take the interval $[a, b]$ and divide it into n subintervals each of length,

$$\Delta x = \frac{b-a}{n}$$

Now from each of these intervals choose the points $x_1^*, x_2^*, \dots, x_n^*$ and note that it doesn't really matter how we choose each of these numbers as long as they come from the appropriate interval. We can then compute the average of the values $f(x_1^*), f(x_2^*), \dots, f(x_n^*)$ by computing,

$$\frac{f(x_1^*) + f(x_2^*) + \cdots + f(x_n^*)}{n} \tag{A.17}$$

Now, from our definition of Δx we can get the following formula for n.

$$n = \frac{b-a}{\Delta x}$$

and we can plug this into Equation A.17 to get,

$$\begin{aligned} \frac{f(x_1^*) + f(x_2^*) + \cdots + f(x_n^*)}{\frac{b-a}{\Delta x}} &= \frac{\left[f(x_1^*) + f(x_2^*) + \cdots + f(x_n^*)\right]\Delta x}{b-a} \\ &= \frac{1}{b-a}\left[f(x_1^*)\,\Delta x + f(x_2^*)\,\Delta x + \cdots + f(x_n^*)\,\Delta x\right] \\ &= \frac{1}{b-a}\sum_{i=1}^{n} f(x_i^*)\,\Delta x \end{aligned}$$

Let's now increase n. Doing this will mean that we're taking the average of more and more function values in the interval and so the larger we chose n the better this will approximate the average value of the function.

If we then take the limit as n goes to infinity we should get the average function value. Or,

$$f_{avg} = \lim_{n\to\infty} \frac{1}{b-a}\sum_{i=1}^{n} f(x_i^*)\,\Delta x = \frac{1}{b-a}\lim_{n\to\infty}\sum_{i=1}^{n} f(x_i^*)\,\Delta x$$

We can factor the $\frac{1}{b-a}$ out of the limit as we've done and now the limit of the sum should look familiar as that is the definition of the definite integral. So, putting in definite integral we get the formula that we were after.

$$f_{avg} = \frac{1}{b-a}\int_a^b f(x)\,dx$$

The Mean Value Theorem for Integrals

If $f(x)$ is a continuous function on $[a,b]$ then there is a number c in $[a,b]$ such that,

$$\int_a^b f(x)\,dx = f(c)(b-a)$$

Proof

Let's start off by defining,

$$F(x) = \int_a^x f(t)\,dt$$

Since $f(x)$ is continuous we know from the Fundamental Theorem of Calculus, Part I that $F(x)$ is continuous on $[a,b]$, differentiable on (a,b) and that $F'(x) = f(x)$.

Now, from the Mean Value Theorem we know that there is a number c such that $a < c < b$ and that,

$$F(b) - F(a) = F'(c)(b-a)$$

However, we know that $F'(c) = f(c)$ and,

$$F(b) = \int_a^b f(t)\,dt = \int_a^b f(x)\,dx \qquad F(a) = \int_a^a f(t)\,dt = 0$$

So, we then have,

$$\int_a^b f(x)\,dx = f(c)(b-a)$$

Work

The work done by the force $F(x)$ (assuming that $F(x)$ is continuous) over the range $a \le x \le b$ is,

$$W = \int_a^b F(x)\,dx$$

Proof

Let's start off by dividing the range $a \le x \le b$ into n subintervals of width Δx and from each of these intervals choose the points $x_1^*, x_2^*, \ldots, x_n^*$.

Now, if n is large and because $F(x)$ is continuous we can assume that $F(x)$ won't vary by much over each interval and so in the i^{th} interval we can assume that the force is approximately constant with a value of $F(x) \approx F(x_i^*)$. The work on each interval is then approximately,

$$W_i \approx F(x_i^*)\,\Delta x$$

The total work over $a \le x \le b$ is then approximately,

$$W \approx \sum_{i=1}^{n} W_i = \sum_{i=0}^{n} F(x_i^*)\,\Delta x$$

Finally, if we take the limit of this as n goes to infinity we'll get the exact work done. So,

$$W = \lim_{n\to\infty} \sum_{i=0}^{n} F(x_i^*)\,\Delta x$$

This is, however, nothing more than the definition of the definite integral and so the work done by the force $F(x)$ over $a \le x \le b$ is,

$$W = \int_a^b F(x)\,dx$$

A.6 Area and Volume Formulas

In this section we will derive the formulas used to get the area between two curves and the volume of a solid of revolution.

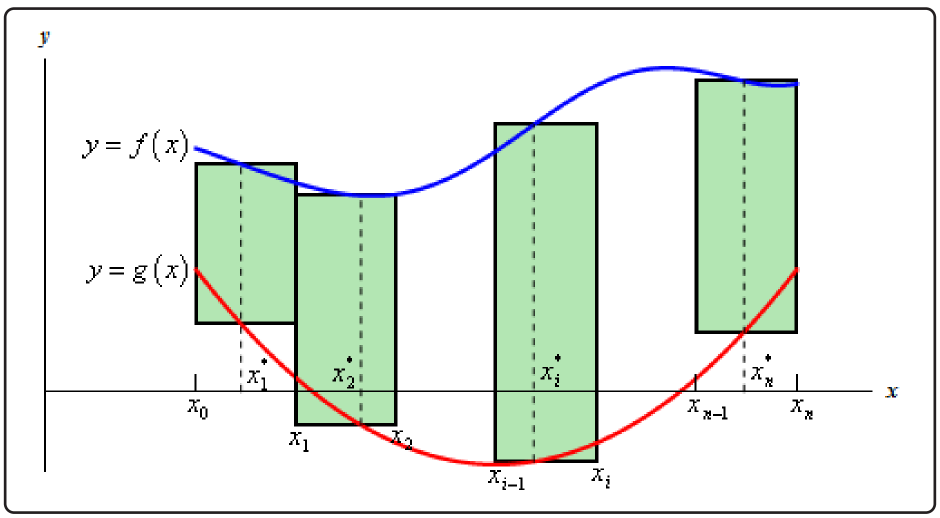
Area Between Two Curves

We will start with the formula for determining the area between $y = f(x)$ and $y = g(x)$ on the interval $[a, b]$. We will also assume that $f(x) \geq g(x)$ on $[a, b]$.

We will now proceed much as we did when we looked that the Area Problem in the Integrals Chapter. We will first divide up the interval into n equal subintervals each with length,

$$\Delta x = \frac{b-a}{n}$$

Next, pick a point in each subinterval, x_i^*, and we can then use rectangles on each interval as follows.



The height of each of these rectangles is given by,

$$f(x_i^*) - g(x_i^*)$$

and the area of each rectangle is then,

$$\big(f(x_i^*) - g(x_i^*)\big)\Delta x$$

So, the area between the two curves is then approximated by,

$$A \approx \sum_{i=1}^{n} \left(f(x_i^*) - g(x_i^*)\right)\Delta x$$

The exact area is,

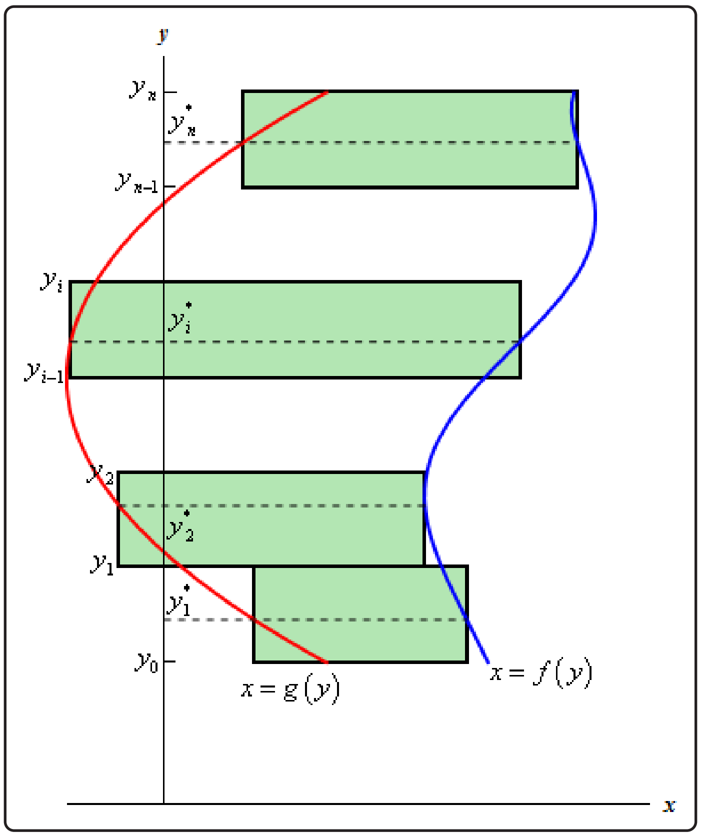
$$A = \lim_{n\to\infty} \sum_{i=1}^{n} \big(f(x_i^*) - g(x_i^*)\big)\Delta x$$

Now, recalling the definition of the definite integral this is nothing more than,

$$A = \int_a^b f(x) - g(x)\, dx$$

The formula above will work provided the two functions are in the form $y = f(x)$ and $y = g(x)$. However, not all functions are in that form.

Sometimes we will be forced to work with functions in the form between $x = f(y)$ and $x = g(y)$ on the interval $[c, d]$ (an interval of y values...). When this happens, the derivation is identical. First we will start by assuming that $f(y) \ge g(y)$ on $[c, d]$. We can then divide up the interval into equal subintervals and build rectangles on each of these intervals. Here is a sketch of this situation.



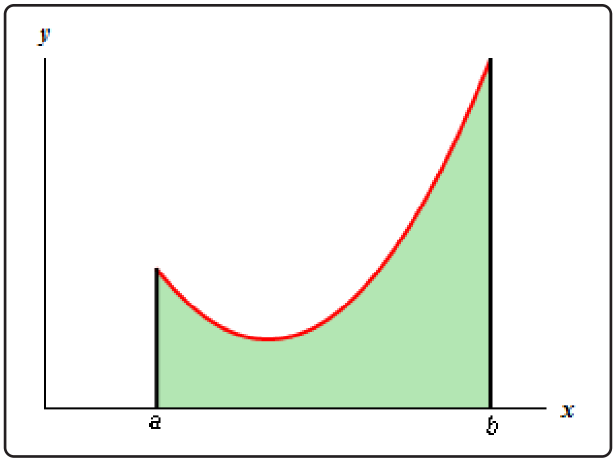
Following the work from above, we will arrive at the following for the area,

$$A = \int_c^d f(y) - g(y)\, dy$$

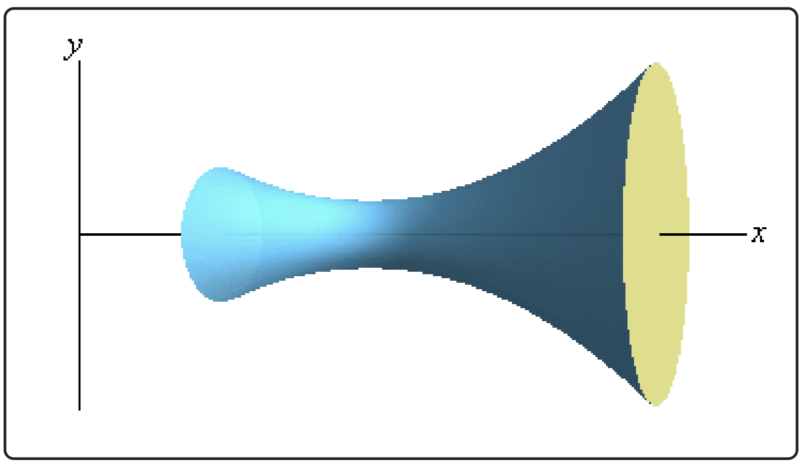
So, regardless of the form that the functions are in we use basically the same formula.

Volumes for Solid of Revolution

Before deriving the formula for this we should probably first define just what a solid of revolution is. To get a solid of revolution we start out with a function, $y = f(x)$, on an interval $[a, b]$.



We then rotate this curve about a given axis to get the surface of the solid of revolution. For purposes of this derivation let's rotate the curve about the x-axis. Doing this gives the following three dimensional region.



We want to determine the volume of the interior of this object. To do this we will proceed much as we did for the area between two curves case. We will first divide up the interval into n subintervals of width,

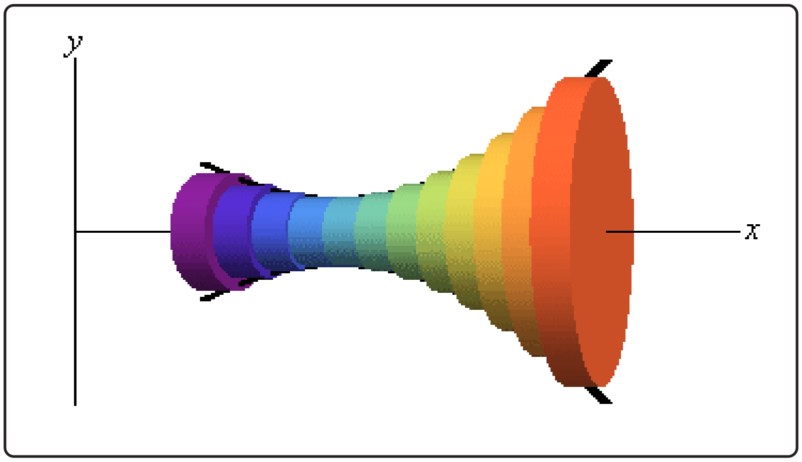
$$\Delta x = \frac{b-a}{n}$$

We will then choose a point from each subinterval, x_i^*.

Now, in the area between two curves case we approximated the area using rectangles on each subinterval. For volumes we will use disks on each subinterval to approximate the area. The area of the face of each disk is given by $A\left(x_i^*\right)$ and the volume of each disk is

$$V_i = A\left(x_i^*\right)\Delta x$$

Here is a sketch of this,



The volume of the region can then be approximated by,

$$V \approx \sum_{i=1}^{n} A\left(x_i^*\right) \Delta x$$

The exact volume is then,

$$\begin{aligned} V &= \lim_{n \to \infty} \sum_{i=1}^{n} A\left(x_i^*\right) \Delta x \\ &= \int_a^b A\left(x\right)\, dx \end{aligned}$$

So, in this case the volume will be the integral of the cross-sectional area at any x, $A\left(x\right)$. Note as well that, in this case, the cross-sectional area is a circle and we could go farther and get a formula for that as well. However, the formula above is more general and will work for any way of getting a cross section so we will leave it like it is.

In the sections where we actually use this formula we will also see that there are ways of generating the cross section that will actually give a cross-sectional area that is a function of y instead of x. In these cases the formula will be,

$$V = \int_c^d A\left(y\right)\, dy, \qquad c \le y \le d$$

In this case we looked at rotating a curve about the x-axis, however, we could have just as easily rotated the curve about the y-axis. In fact, we could rotate the curve about any vertical or horizontal axis and in all of these, case we can use one or both of the following formulas.

$$V = \int_a^b A\left(x\right)\, dx \qquad\qquad V = \int_c^d A\left(y\right)\, dy$$

A.7 Types of Infinity

Most students have run across infinity at some point in time prior to a calculus class. However, when they have dealt with it, it was just a symbol used to represent a really, really large positive or really, really large negative number and that was the extent of it. Once they get into a calculus class students are asked to do some basic algebra with infinity and this is where they get into trouble. Infinity is NOT a number and for the most part doesn't behave like a number. However, despite that we'll think of infinity in this section as a really, really, really large number that is so large there isn't another number larger than it. This is not correct of course but may help with the discussion in this section. Note as well that everything that we'll be discussing in this section applies only to real numbers. If you move into complex numbers for instance things can and do change.

So, let's start thinking about addition with infinity. When you add two non-zero numbers you get a new number. For example, $4 + 7 = 11$. With infinity this is not true. With infinity you have the following.

$$\begin{aligned} \infty + a &= \infty \qquad \text{where } a \neq -\infty \\ \infty + \infty &= \infty \end{aligned}$$

In other words, a really, really large positive number (∞) plus any positive number, regardless of the size, is still a really, really large positive number. Likewise, you can add a negative number (*i.e.* $a < 0$) to a really, really large positive number and stay really, really large and positive. So, addition involving infinity can be dealt with in an intuitive way if you're careful. Note as well that the a must NOT be negative infinity. If it is, there are some serious issues that we need to deal with as we'll see in a bit.

Subtraction with negative infinity can also be dealt with in an intuitive way in most cases as well. A really, really large negative number minus any positive number, regardless of its size, is still a really, really large negative number. Subtracting a negative number (*i.e.* $a < 0$) from a really, really large negative number will still be a really, really large negative number. Or,

$$\begin{aligned} -\infty - a &= -\infty \qquad \text{where } a \neq -\infty \\ -\infty - \infty &= -\infty \end{aligned}$$

Again, a must not be negative infinity to avoid some potentially serious difficulties.

Multiplication can be dealt with fairly intuitively as well. A really, really large number (positive, or negative) times any number, regardless of size, is still a really, really large number we'll just need to be careful with signs. In the case of multiplication we have

$$(a)(\infty) = \infty \quad \text{if } a > 0 \qquad\qquad (a)(\infty) = -\infty \quad \text{if } a < 0$$

$$(\infty)(\infty) = \infty \qquad\qquad (-\infty)(-\infty) = \infty \qquad\qquad (-\infty)(\infty) = -\infty$$

What you know about products of positive and negative numbers is still true here.

Some forms of division can be dealt with intuitively as well. A really, really large number divided by a number that isn't too large is still a really, really large number.

$$\frac{\infty}{a} = \infty \qquad \text{if } a > 0, a \neq \infty \qquad\qquad \frac{\infty}{a} = -\infty \qquad \text{if } a < 0, a \neq -\infty$$
$$\frac{-\infty}{a} = -\infty \qquad \text{if } a > 0, a \neq \infty \qquad\qquad \frac{-\infty}{a} = \infty \qquad \text{if } a < 0, a \neq -\infty$$

Division of a number by infinity is somewhat intuitive, but there are a couple of subtleties that you need to be aware of. When we talk about division by infinity we are really talking about a limiting process in which the denominator is going towards infinity. So, a number that isn't too large divided an increasingly large number is an increasingly small number. In other words, in the limit we have,

$$\frac{a}{\infty} = 0 \qquad\qquad \frac{a}{-\infty} = 0$$

So, we've dealt with almost every basic algebraic operation involving infinity. There are two cases that that we haven't dealt with yet. These are

$$\infty - \infty = ? \qquad\qquad \frac{\pm\infty}{\pm\infty} = ?$$

The problem with these two cases is that intuition doesn't really help here. A really, really large number minus a really, really large number can be anything ($-\infty$, a constant, or ∞). Likewise, a really, really large number divided by a really, really large number can also be anything ($\pm\infty$ - this depends on sign issues, 0, or a non-zero constant).

What we've got to remember here is that there are really, really large numbers and then there are really, really, really large numbers. In other words, some infinities are larger than other infinities. With addition, multiplication and the first sets of division we worked this wasn't an issue. The general size of the infinity just doesn't affect the answer in those cases. However, with the subtraction and division cases listed above, it does matter as we will see.

Here is one way to think of this idea that some infinities are larger than others. This is a fairly dry and technical way to think of this and your calculus problems will probably never use this stuff, but it is a nice way of looking at this. Also, please note that I'm not trying to give a precise proof of anything here. I'm just trying to give you a little insight into the problems with infinity and how some infinities can be thought of as larger than others. For a much better (and definitely more precise) discussion see,

http://www.math.vanderbilt.edu/ schectex/courses/infinity.pdf

Let's start by looking at how many integers there are. Clearly, I hope, there are an infinite number of them, but let's try to get a better grasp on the "size" of this infinity. So, pick any two integers completely at random. Start at the smaller of the two and list, in increasing order, all the integers that come after that. Eventually we will reach the larger of the two integers that you picked.

Depending on the relative size of the two integers it might take a very, very long time to list all the integers between them and there isn't really a purpose to doing it. But, it could be done if we wanted to and that's the important part.

Because we could list all these integers between two randomly chosen integers we say that the integers are *countably infinite*. Again, there is no real reason to actually do this, it is simply something that can be done if we should choose to do so.

In general, a set of numbers is called countably infinite if we can find a way to list them all out. In a more precise mathematical setting this is generally done with a special kind of function called a bijection that associates each number in the set with exactly one of the positive integers. To see some more details of this see the pdf given above.

It can also be shown that the set of all fractions are also countably infinite, although this is a little harder to show and is not really the purpose of this discussion. To see a proof of this see the pdf given above. It has a very nice proof of this fact.

Let's contrast this by trying to figure out how many numbers there are in the interval $(0, 1)$. By numbers, I mean all possible fractions that lie between zero and one as well as all possible decimals (that aren't fractions) that lie between zero and one. The following is similar to the proof given in the pdf above but was nice enough and easy enough (I hope) that I wanted to include it here.

To start let's assume that all the numbers in the interval $(0, 1)$ are countably infinite. This means that there should be a way to list all of them out. We could have something like the following,

$$\begin{aligned} x_1 &= 0.692096\cdots \\ x_2 &= 0.171034\cdots \\ x_3 &= 0.993671\cdots \\ x_4 &= 0.045908\cdots \\ \vdots & \qquad \vdots \end{aligned}$$

Now, select the i^{th} decimal out of x_i as shown below

$$\begin{aligned} x_1 &= 0.\underline{6}92096\cdots \\ x_2 &= 0.1\underline{7}1034\cdots \\ x_3 &= 0.99\underline{3}671\cdots \\ x_4 &= 0.045\underline{9}08\cdots \\ \vdots & \qquad \vdots \end{aligned}$$

and form a new number with these digits. So, for our example we would have the number

$$x = 0.6739\cdots$$

In this new decimal replace all the 3's with a 1 and replace every other numbers with a 3. In the case of our example this would yield the new number

$$\overline{x} = 0.3313\cdots$$

Notice that this number is in the interval $(0, 1)$ and also notice that given how we choose the digits of the number this number will not be equal to the first number in our list, x_1, because the first digit of each is guaranteed to not be the same. Likewise, this new number will not get the same number as the second in our list, x_2, because the second digit of each is guaranteed to not be the same. Continuing in this manner we can see that this new number we constructed, $\overline{x}$, is guaranteed to not be in our listing. But this contradicts the initial assumption that we could list out all the numbers in the interval $(0, 1)$. Hence, it must not be possible to list out all the numbers in the interval $(0, 1)$.

Sets of numbers, such as all the numbers in $(0, 1)$, that we can't write down in a list are called *uncountably* infinite.

The reason for going over this is the following. An infinity that is uncountably infinite is significantly larger than an infinity that is only countably infinite. So, if we take the difference of two infinities we have a couple of possibilities.

$$\infty\,(\text{uncountable}) - \infty\,(\text{countable}) = \infty$$

$$\infty\,(\text{countable}) - \infty\,(\text{uncountable}) = -\infty$$

Notice that we didn't put down a difference of two infinities of the same type. Depending upon the context there might still have some ambiguity about just what the answer would be in this case, but that is a whole different topic.

We could also do something similar for quotients of infinities.

$$\frac{\infty\,(\text{countable})}{\infty\,(\text{uncountable})} = 0$$

$$\frac{\infty\,(\text{uncountable})}{\infty\,(\text{countable})} = \infty$$

Again, we avoided a quotient of two infinities of the same type since, again depending upon the context, there might still be ambiguities about its value.

So, that's it and hopefully you've learned something from this discussion. Infinity simply isn't a number and because there are different kinds of infinity it generally doesn't behave as a number does. Be careful when dealing with infinity.

A.8 Summation Notation

In this section we need to do a brief review of summation notation or sigma notation. We'll start out with two integers, n and m, with $n < m$ and a list of numbers denoted as follows,

$$a_n,\ a_{n+1},\ a_{n+2},\ \ldots,\ a_{m-2},\ a_{m-1},\ a_m$$

We want to add them up, in other words we want,

$$a_n + a_{n+1} + a_{n+2} + \ldots + a_{m-2} + a_{m-1} + a_m$$

For large lists this can be a fairly cumbersome notation so we introduce summation notation to denote these kinds of sums. The case above is denoted as follows.

$$\sum_{i=n}^{m} a_i = a_n + a_{n+1} + a_{n+2} + \ldots + a_{m-2} + a_{m-1} + a_m$$

The i is called the index of summation. This notation tells us to add all the a_i's up for all integers starting at n and ending at m.

For instance,

$$\begin{aligned}
&\sum_{i=0}^{4} \frac{i}{i+1} = \frac{0}{0+1} + \frac{1}{1+1} + \frac{2}{2+1} + \frac{3}{3+1} + \frac{4}{4+1} = \frac{163}{60} = 2.71\overline{6}\overline{6}\overline{6} \\
&\sum_{i=4}^{6} 2^i x^{2i+1} = 2^4 x^9 + 2^5 x^{11} + 2^6 x^{13} = 16x^9 + 32x^{11} + 64x^{13} \\
&\sum_{i=1}^{4} f\left(x_i^*\right) = f\left(x_1^*\right) + f\left(x_2^*\right) + f\left(x_3^*\right) + f\left(x_4^*\right)
\end{aligned}$$

Properties

Here are a couple of formulas for summation notation.

Fact

1. $\sum\limits_{i=i_0}^{n} ca_i = c\sum\limits_{i=i_0}^{n} a_i$ where c is any number. So, we can factor constants out of a summation.
2. $\sum\limits_{i=i_0}^{n} (a_i \pm b_i) = \sum\limits_{i=i_0}^{n} a_i \pm \sum\limits_{i=i_0}^{n} b_i$ So, we can break up a summation across a sum or difference.

Note that we started the series at i_0 to denote the fact that they can start at any value of i that we need them to. Also note that while we can break up sums and differences as we did in **2** above we can't do the same thing for products and quotients. In other words,

$$\sum_{i=i_0}^{n} (a_i b_i) \neq \left(\sum_{i=i_0}^{n} a_i\right)\left(\sum_{i=i_0}^{n} b_i\right) \qquad \sum_{i=i_0}^{n} \frac{a_i}{b_i} \neq \frac{\sum_{i=i_0}^{n} a_i}{\sum_{i=i_0}^{n} b_i}$$

Formulas

Here are a couple of nice formulas that we will find useful in a couple of sections. Note that these formulas are only true if starting at $i = 1$. You can, of course, derive other formulas from these for different starting points if you need to.

Fact

1. $\displaystyle\sum_{i=1}^{n} c = cn$
2. $\displaystyle\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$
3. $\displaystyle\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$
4. $\displaystyle\sum_{i=1}^{n} i^3 = \left[\frac{n(n+1)}{2}\right]^2$

Here is a quick example on how to use these properties to quickly evaluate a sum that would not be easy to do by hand.

Example 1

Using the formulas and properties from above determine the value of the following summation.

$$\sum_{i=1}^{100} (3-2i)^2$$

Solution

The first thing that we need to do is square out the stuff being summed and then break up the summation using the properties as follows,

$$\begin{aligned}\sum_{i=1}^{100}(3-2i)^2 &= \sum_{i=1}^{100} 9-12i+4i^2\\ &= \sum_{i=1}^{100} 9-\sum_{i=1}^{100} 12i+\sum_{i=1}^{100} 4i^2\\ &= \sum_{i=1}^{100} 9-12\sum_{i=1}^{100} i+4\sum_{i=1}^{100} i^2\end{aligned}$$

Now, using the formulas, this is easy to compute,

$$\begin{aligned}\sum_{i=1}^{100}(3-2i)^2 &= 9\left(100\right)-12\left(\frac{100\left(101\right)}{2}\right)+4\left(\frac{100\left(101\right)\left(201\right)}{6}\right)\\ &= 1293700\end{aligned}$$

Doing this by hand would definitely taken some time and there's a good chance that we might have made a minor mistake somewhere along the line.

A.9 Constant of Integration

In this section we need to address a couple of topics about the constant of integration. Throughout most calculus classes we play pretty fast and loose with it and because of that many students don't really understand it or how it can be important.

First, let's address how we play fast and loose with it. Recall that technically when we integrate a sum or difference we are actually doing multiple integrals. For instance,

$$\int 15x^4 - 9x^{-2}\,dx = \int 15x^4\,dx - \int 9x^{-2}\,dx$$

Upon evaluating each of these integrals we should get a constant of integration for each integral since we really are doing two integrals.

$$\begin{aligned}\int 15x^4 - 9x^{-2}\,dx &= \int 15x^4\,dx - \int 9x^{-2}\,dx\\ &= 3x^5 + c + 9x^{-1} + k\\ &= 3x^5 + 9x^{-1} + c + k\end{aligned}$$

Since there is no reason to think that the constants of integration will be the same from each integral we use different constants for each integral.

Now, both c and k are unknown constants and so the sum of two unknown constants is just an unknown constant and we acknowledge that by simply writing the sum as a c.

So, the integral is then,

$$\int 15x^4 - 9x^{-2}\,dx = 3x^5 + 9x^{-1} + c$$

We also tend to play fast and loose with constants of integration in some substitution rule problems. Consider the following problem,

$$\int \cos\left(1+2x\right) + \sin\left(1+2x\right)\,dx = \frac{1}{2}\int \cos(u) + \sin(u)\,du \qquad\qquad u = 1+2x$$

Technically when we integrate we should get,

$$\int \cos\left(1+2x\right) + \sin\left(1+2x\right)\,dx = \frac{1}{2}\Big(\sin(u) - \cos(u) + c\Big)$$

Since the whole integral is multiplied by $\frac{1}{2}$, the whole answer, including the constant of integration, should be multiplied by $\frac{1}{2}$. Upon multiplying the $\frac{1}{2}$ through the answer we get,

$$\int \cos\left(1+2x\right) + \sin\left(1+2x\right)\,dx = \frac{1}{2}\sin(u) - \frac{1}{2}\cos(u) + \frac{c}{2}$$

However, since the constant of integration is an unknown constant dividing it by 2 isn't going to change that fact so we tend to just write the fraction as a c.

$$\int \cos\left(1+2x\right) + \sin\left(1+2x\right)\,dx = \frac{1}{2}\sin(u) - \frac{1}{2}\cos(u) + c$$

In general, we don't really need to worry about how we've played fast and loose with the constant of integration in either of the two examples above.

The real problem however is that because we play fast and loose with these constants of integration most students don't really have a good grasp of them and don't understand that there are times where the constants of integration are important and that we need to be careful with them.

To see how a lack of understanding about the constant of integration can cause problems consider the following integral.

$$\int \frac{1}{2x}\,dx$$

This is a really simple integral. However, there are two ways (both simple) to integrate it and that is where the problem arises.

The first integration method is to just break up the fraction and do the integral.

$$\int \frac{1}{2x}\,dx = \int \frac{1}{2}\frac{1}{x}\,dx = \frac{1}{2}\ln|x| + c$$

The second way is to use the following substitution.

$$u = 2x \qquad du = 2dx \qquad \Rightarrow \qquad dx = \frac{1}{2}du$$

$$\int \frac{1}{2x}\,dx = \frac{1}{2}\int \frac{1}{u}\,du = \frac{1}{2}\ln|u| + c = \frac{1}{2}\ln|2x| + c$$

Can you see the problem? We integrated the same function and got very different answers. This doesn't make any sense. Integrating the same function should give us the same answer. We only used different methods to do the integral and both are perfectly legitimate integration methods. So, how can using different methods produce different answer?

The first thing that we should notice is that because we used a different method for each there is no reason to think that the constant of integration will in fact be the same number and so we really should use different letters for each.

More appropriate answers would be,

$$\int \frac{1}{2x}\,dx = \frac{1}{2}\ln|x| + c \qquad\qquad \int \frac{1}{2x}\,dx = \frac{1}{2}\ln|2x| + k$$

Now, let's take another look at the second answer. Using a property of logarithms we can write the answer to the second integral as follows,

$$\begin{aligned}\int \frac{1}{2x}\,dx &= \frac{1}{2}\ln|2x| + k \\ &= \frac{1}{2}\left(\ln 2 + \ln|x|\right) + k \\ &= \frac{1}{2}\ln|x| + \frac{1}{2}\ln 2 + k\end{aligned}$$

Upon doing this we can see that the answers really aren't that different after all. In fact they only differ by a constant and we can even find a relationship between c and k. It looks like,

$$c = \frac{1}{2}\ln 2 + k$$

So, without a proper understanding of the constant of integration, in particular using different integration techniques on the same integral will likely produce a different constant of integration, we might never figure out why we got "different" answers for the integral.

Note as well that getting answers that differ by a constant doesn't violate any principles of calculus. In fact, we've actually seen a fact that suggested that this might happen. We saw a fact in the Mean Value Theorem section that said that if $f'(x) = g'(x)$ then $f(x) = g(x) + c$. In other words, if two functions have the same derivative then they can differ by no more than a constant.

This is exactly what we've got here. The two functions,

$$f(x) = \frac{1}{2}\ln|x| \qquad\qquad g(x) = \frac{1}{2}\ln|2x|$$

have exactly the same derivative,

$$\frac{1}{2x}$$

and as we've shown they really only differ by a constant.

There is another integral that also exhibits this behavior. Consider,

$$\int \sin(x)\cos(x)\,dx$$

There are actually three different methods for doing this integral.

Method 1 :

This method uses a trig formula,

$$\sin(2x) = 2\sin(x)\cos(x)$$

Using this formula (and a quick substitution) the integral becomes,

$$\int \sin(x)\cos(x)\,dx = \frac{1}{2}\int \sin(2x)\,dx = -\frac{1}{4}\cos(2x) + c_1$$

Method 2 :

This method uses the substitution,

$$u = \cos(x) \qquad\qquad du = -\sin(x)\,dx$$

$$\int \sin(x)\cos(x)\,dx = -\int u\,du = -\frac{1}{2}u^2 + c_2 = -\frac{1}{2}\cos^2(x) + c_2$$

Method 3 :

Here is another substitution that could be done here as well.

$$u = \sin(x) \qquad du = \cos(x)\,dx$$

$$\int \sin(x)\cos(x)\ dx = \int u\,du = \frac{1}{2}u^2 + c_3 = \frac{1}{2}\sin^2(x) + c_3$$

So, we've got three different answers each with a different constant of integration. However, according to the fact above these three answers should only differ by a constant since they all have the same derivative.

In fact, they do only differ by a constant. We'll need the following trig formulas to prove this.

$$\cos(2x) = \cos^2(x) - \sin^2(x) \qquad \cos^2(x) + \sin^2(x) = 1$$

Start with the answer from the first method and use the double angle formula above.

$$-\frac{1}{4}\left(\cos^2(x) - \sin^2(x)\right) + c_1$$

Now, from the second identity above we have,

$$\sin^2(x) = 1 - \cos^2(x)$$

so, plug this in,

$$\begin{aligned} -\frac{1}{4}\Big(\cos^2(x) - \left(1 - \cos^2(x)\right)\Big) + c_1 &= -\frac{1}{4}\Big(2\cos^2(x) - 1\Big) + c_1 \\ &= -\frac{1}{2}\cos^2(x) + \frac{1}{4} + c_1 \end{aligned}$$

This is then answer we got from the second method with a slightly different constant. In other words,

$$c_2 = \frac{1}{4} + c_1$$

We can do a similar manipulation to get the answer from the third method. Again, starting with the answer from the first method use the double angle formula and then substitute in for the cosine instead of the sine using,

$$\cos^2(x) = 1 - \sin^2(x)$$

Doing this gives,

$$\begin{aligned} -\frac{1}{4}\Big(\left(1 - \sin^2(x)\right) - \sin^2(x)\Big) + c_1 &= -\frac{1}{4}\Big(1 - 2\sin^2(x)\Big) + c_1 \\ &= \frac{1}{2}\sin^2(x) - \frac{1}{4} + c_1 \end{aligned}$$

which is the answer from the third method with a different constant and again we can relate the two constants by,

$$c_3 = -\frac{1}{4} + c_1$$

So, what have we learned here? Hopefully we've seen that constants of integration are important and we can't forget about them. We often don't work with them in a Calculus I course, yet without a good understanding of them we would be hard pressed to understand how different integration methods can apparently produce different answers.

Index

Quite a few entries in this index are duplicated in several places throughout the index. For example **Work** is listed both separately and as a sub entry of **Integral Applications**. The reason for this is that I tried to anticipate (probably badly at times....) just where in the index a reader would be looking for a particular topic. Again to use Work as an example. A reader might look for Work as a separate entry or, maybe, upon realizing it was an application of integration look for it there.

So, assuming that I've anticipated properly, you should be able to find most entries in the table no matter where you look (provided it's actually in the table of course).