Discover Calculus I

Single-Variable Differential Calculus Topics with Motivating Activities

Discover Calculus I

Single-Variable Differential Calculus Topics with Motivating Activities

Peter Keep
Moraine Valley Community College

Last revised: August 17, 2025

[1] www.discovercalculus.com
[2] creativecommons.org/licenses/by/4.0

Acknowledgements

Disclosure about the Use of AI

This book has been lovingly written by a human.

Me.

Peter Keep.

I have used a lot of different tools, both for inspiration and for actually creating resources for this book. *None* of those tools has involved any form of generative AI.

I could list all of the ways that I think using generative AI in education is, at minimum, problematic. More pointedly, I believe that it is unethical. More broadly, I believe that the use of generative AI for any use-case that I have encountered to be unethical.

In my classes, I try to help students realize the joy and value of working at something and creating something and struggling with something and knowing something. Giving worth to something, even an imperfect thing. Celebrating our accomplishments, even when (especially when?) there is room to grow in those accomplishments. And so I have taken that advice in the creation of this book. I have created a book that is definitely not perfect. I have struggled to write it. There are parts of it that could be (need to be) improved.

But I was the one that created it. I struggled with it. I know it.

I hope that this book can also be a useful tool for others to use, and I have left the copyright to be about as open as possible. Others can take this, use it, can change it, add to it, subtract from it, etc.

In leaving this copyright open for others to change this book, I cannot guarantee that every version of this book is free from the mindless and joyless output from some Large Language Model. But I want to leave this note up in hopes that anyone who *does* inject some output from some generative AI product into this book will take it down. If this note, or some statement similar to it, is not present in the version of the book you are accessing, please be cautious. Find a different calculus textbook to read!

Find something written by a human. Find the words of some other mathematician who tries, maybe imperfectly, to share the ideas of calculus.

Teaching and learning is about humans communicating with each other, and only humans can do that.

Notes for Instructors

Notes for Students

Contents

Chapter 1

Limits

1.1 The Definition of the Limit

We're going to start this textbook by stating a definition. This is a common practice in math classes: we need to agree upon a common definition of the mathematical objects and adjectives we are thinking about. We will state a lot of definitions in this textbook.

What I hope we will do, though, is motivate these definitions. We want to arrive at a point where it makes sense to give a name to this phenomena or object that we're thinking of. Or maybe we arrive at a point where the specifics of the definition don't just come down to us out of nowhere, but feel like reasonable and obvious things to consider.

So for now, we're going to work on defining a very important and very key mathematical object that is used in calculus: the limit.

A limit is all about closeness, so let's first interact with the idea of closeness, and then work on a definition of a limit.

Defining a Limit

Activity 1.1.1 Close or Not?

We're going to try to think how we might define "close"-ness as a property, but, more importantly, we're going to try to realize the struggle of creating definitions in a mathematical context. We want our definition to be meaningful, precise, and useful, and those are hard goals to reach! Coming to some agreement on this is a particularly tricky task.

(a) For each of the following pairs of things, decide on which pairs you would classify as "close" to each other.

- You, right now, and the nearest city with a population of 1 million or higher
- Your two nostrils
- You and the door of the room you are in
- You and the person nearest you
- The floor of the room you are in and the ceiling of the room you are in

(b) For your classification of "close," what does "close" mean? Finish the sentence: A pair of objects are *close* to each other if...

(c) Let's think about how close two things would have to be in order to satisfy everyone's definition of "close." Pick two objects that you think everyone would agree are "close," if by "everyone" we meant:

- All of the people in the building you are in right now.
- All of the people in the city that you are in right now.
- All of the people in the country that you are in right now.
- Everyone, everywhere, all at once.

(d) Let's put ourselves into the context of functions and numbers. Consider the linear function $y = 4x - 1$. Our goal is to find some x-values that, when we put them into our function, give us y-value outputs that are "close" to the number 2. You get to define what close means.

First, evaluate $f(0)$ and $f(1)$. Are these y-values "close" to 2, in your definition of "close?"

(e) Pick five more, different, numbers that are "close" to 2 in your definition of "close." For each one, find the x-values that give you those y-values.

(f) How far away from $x = \frac{3}{4}$ can you go and still have y-value outputs that are "close" to 2?

To wrap this up, think about your points that you have: you have a list of x-coordinates that are clustered around $x = \frac{3}{4}$ where, when you evaluate $y = 4x - 1$ at those x-values, you get y-values that are "close" to 2. Great!
Do you think others will agree? Or do you think that other people might look at your list of y-values and decide that some of them *aren't* close to 2?
Do you think you would agree with other peoples' lists? Or you do think that you might look at other peoples' lists of y-values and decide that some of them *aren't* close to 2?

The balance that we need to find, as we discovered in Activity 1.1.1, is about being able to leave room for those with a very strict idea of what "close" might be. We will want to think of an idea kind of like "infinite closeness," but we're not going to frame it this way: we're going to think about a function's output being so close to some specific number that literally everyone can agree. It is so close that it is within every possible definition of closeness.

The general idea is that we want to think about the behavior of a function at inputs that are near some specific input. Is there a trend with the outputs? Are they all centered around a specific value or do they differ wildly?

Definition 1.1.1 Limit of a Function.

For the function $f(x)$ defined at all x-xalues around a (except maybe at $x = a$ itself), we say that the **limit of** $f(x)$ as x approaches a is L if $f(x)$ is arbitrarily close to the single, real number L whenever x is sufficiently close to, but not equal to, a. We write this as:

$$\lim_{x \to a} f(x) = L$$

or sometimes we write $f(x) \to L$ when $x \to a$.

We can clarify a couple of things here:

- There are two types of "close" in this definition: "arbitrarily close" and "sufficiently close." One of these is in references to x-values being close to a number and the other is in reference to function outputs being close to a specific number.
- We are concerned with the behavior of a function around, but not at, a specific x-value: $x = a$. We don't really care about what the function is doing at that input (if anything at all), and we already have words to describe that kind of behavior!
- When we talk about x-values that are near a, that might reference x-values that are a bit bigger than a or x-values that are a bit smaller than a. We can be more specific by simply changing this definition to focus on only one "side" individually.

We can go back to Activity 1.1.1 and think about how we chose x-values that were larger than $\frac{3}{4}$ and smaller than $\frac{3}{4}$. Let's define these ideas a bit more formally!

Definition 1.1.2 Left-Sided Limit.

For the function $f(x)$ defined at all x-xalues around and less than a, we say that the **left-sided limit of** $f(x)$ as x approaches a is L if $f(x)$ is arbitrarily close to the single, real number L whenever x is sufficiently close to, but less than, a. We write this as:

$$\lim_{x \to a^-} f(x) = L$$

or sometimes we write $f(x) \to L$ when $x \to a^-$.

Definition 1.1.3 Right-Sided Limit.

For the function $f(x)$ defined at all x-xalues around and greater than a, we say that the **right-sided limit of** $f(x)$ as x approaches a is L if $f(x)$ is arbitrarily close to the single, real number L whenever x is sufficiently close to, but greater than, a. We write this as:

$$\lim_{x \to a^+} f(x) = L$$

or sometimes we write $f(x) \to L$ when $x \to a^+$.

This should lead us to our first result in this textbook. This first result will do two things:

1. Introduce some language that we can use when we talk about limits as well as a classification that we can apply to them.
2. Introduce how we will build our results throughout the course of this text. We want to discover these results as things that are required for us

to talk about (and do) calculus together, and hopefully we can motivate each one beforehand.

In lieu of a formal activity, let's just review Definition 1.1.1 Limit of a Function pose the following questions to think about:

- Why do we put emphasis on L being a number? What could happen if it wasn't?
- Why do we put the emphasis on the number L being a *real* number? What other type(s) of number could it be?
- Why do we put emphasis on L being a *single* number? How could we have the function be close to multiple real numbers?

We can look at one of the ways that we break the definition: by having two different values that the function gets close to.

Theorem 1.1.4 Mismatched Limits.

For a function $f(x)$, if both $\lim_{x\to a^-} f(x) \neq \lim_{x\to a^+} f(x)$, then we say that $\lim_{x\to a} f(x)$ ***does not exist****.*

Approximating Limits Using Our New Definition

We have defined a new term, and now we do the typical mathematical task: define a new thing and then investigate it.

> A common joke in mathematics is that we "make up a guy to get mad at." It's really only kind of a joke, because it really is a pretty good description of what we do! Here, we defined a new object and now we'll think about it and find ways that it frustrates us or some other weird behavior about it. That's mathematics!

We will eventually get really good at thinking about limits and using them, but for now we just want to get familiar with them. Let's approximate these values that our function is near by looking at some pictures of graphs and some tables of function outputs.

Later on, we'll formalize this more. For now, we just want to use these pictures and tables to get familiar with *what* a limit even is.

Activity 1.1.2 Approximating Limits.

For each of the following graphs of functions, approximate the limit in question. When you do so, approximate the values of the relevant one-sided limits as well.

(a) Approximate $\lim_{x\to 1} f(x)$ using the graph of the function $f(x)$ below.

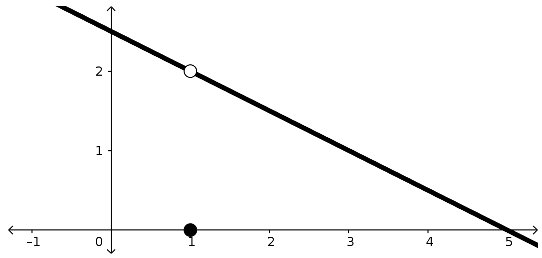


Figure 1.1.5

(b) Approximate $\lim_{x\to 2} g(x)$ using the graph of the function $g(x)$ below.

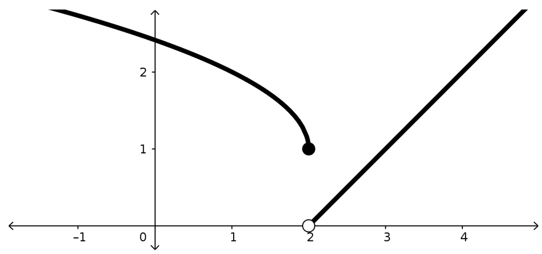


Figure 1.1.6

(c) Approximate the following three limits using the graph of the function $h(x)$ below.

- $\lim_{x\to -1} h(x)$
- $\lim_{x\to 0} h(x)$
- $\lim_{x\to 2} h(x)$

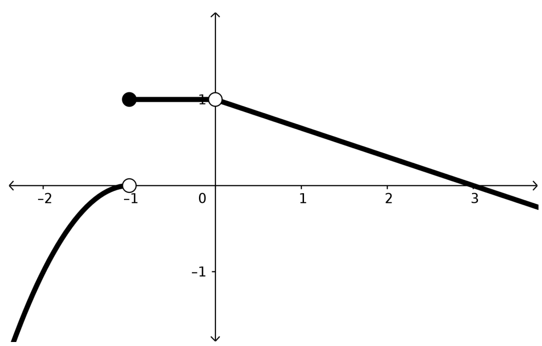


Figure 1.1.7

(d) Why do we say these are ”approximations” or ”estimations” of the limits we’re interested in?

(e) Are there any limit statements that you made that you are 100% confident in? Which ones?

(f) Which limit statements are you least confident in? What about them makes them ones you aren’t confident in?

(g) What extra details would you like to see to increase the confidence in your estimations? Are there changes we could make to the way these functions are represented that would make these approximations better or easier to make?

It can be hard to focus on the aspects of a graph that we really care about for the purpose of a limit. Let's build a small strategy to help us think about what we're looking at. We'll start by just considering some function, $f(x)$. Using our definition of the Limit of a Function as a guide, we'll make sure that it's defined around some x-value, $x = a$.

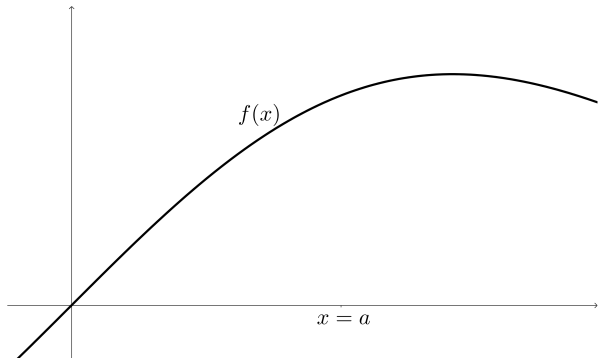


Figure 1.1.8 The function, $f(x)$.

Now we want to investigate more of our definition. We want to look at the x-values that are around, but not equal to, $x = a$.

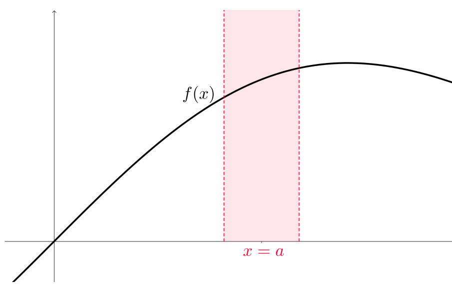


Figure 1.1.9 The x-values around $x = a$.

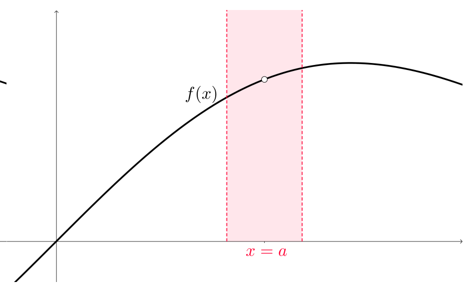


Figure 1.1.10 The x-values around, but not equal to, $x = a$.

We can see that we might as well remove any point at $x = a$ from our graph: we are only concerned with the behavior around that x-value instead of the function's behavior at it.

And now our focus can turn to the function outputs. For the x-values in this interval of inputs that we've constructed, is there some common real number that the corresponding function outputs are close to? We can visualize some interval of y-values. We'll think of this as a target: we want to build an interval of y-values that all of the function outputs from this interval of x-values land in.

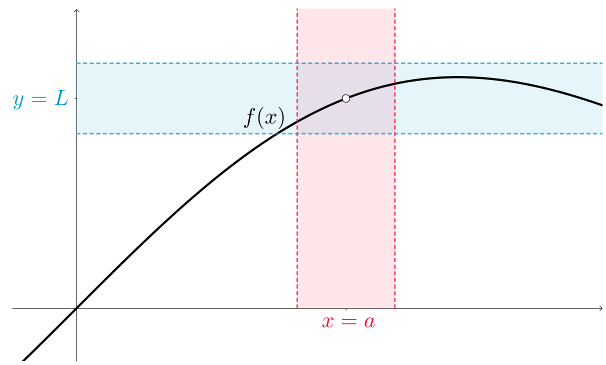


Figure 1.1.11 The corresponding function outputs $f(x)$ are all in the target interval of y-values.

This is a pretty wide range of y-values, but we can see that the graph of the function (when we limit to just the interval of x-values selected) produces function outputs that exist only in that interval. We don't *fill* the interval, but that's fine!

What we *really* care about, though, is if these function outputs are all close to the same, single, real number. What we can do is look at a more strict idea of "closeness" in the y-interval by shrinking it. In order for us to produce function outputs that are in this new, smaller, interval, we'll need to correspondingly shrink our interval of inputs to more closely surround $x = a$.

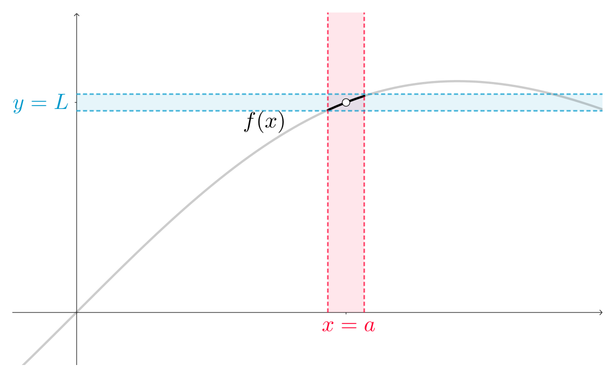


Figure 1.1.12 The x-values around, but not equal to, $x = a$.

In this visualization, we've also tried to focus on just the portion of our function that exists in this little intersection of intervals: we want to know what these functions values are close to, or more specifically if they are all close to the same thing. So we can de-emphasize the rest of our function!

All we're doing is working on a strategy to focus on the parts of this graph that matter: only the parts of the curve that are surrounding $x = a$ (but not that actual point specifically). From there, we just want to know what the function outputs are clustered around, if anything.

Let's look at this same kind of visualization for a limit that does not exist: we're going to think about the case where the one-sided limits don't match. We'll start a little further on in this visualization process: we have a function, and we can visualize an interval of x-values whose function outputs land inside a target interval of y-values.

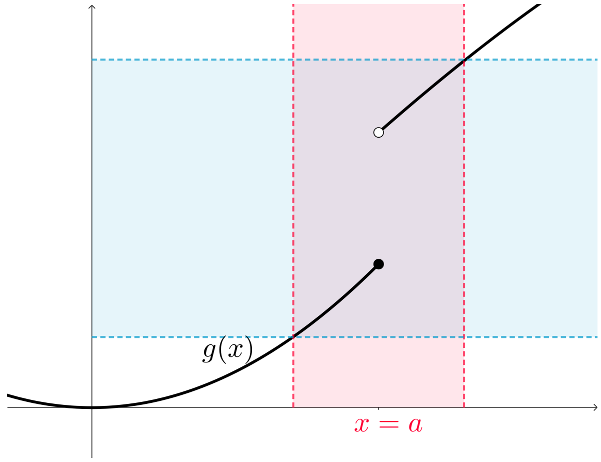


Figure 1.1.13

We can see the problem: that vertical space between the function on the left of $x = a$ and the where the function values are on the right of $x = a$ will make it so that horizontal bar cannot get much smaller. We can disregard the point at $x = a$ as well as the function outside of the interval, but once try to shrink the target interval of y-values, but we'll see the problem.

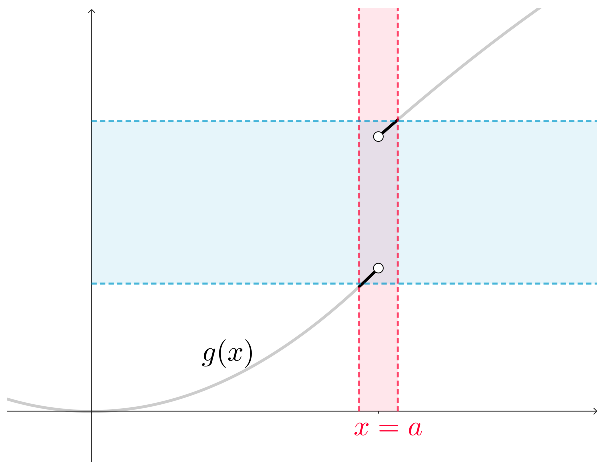


Figure 1.1.14

These function outputs are spread apart! They are not close to a single value. Instead, they're close to two! The function is close to a value on the left side, and then the function is close to a larger value on the right side.

$$\lim_{x\to a^-} g(x) \neq \lim_{x\to a^+} g(x) \text{ and so } \lim_{x\to a} g(x) \text{ does not exist.}$$

Now let's think about how we can approximate (and learn more about) limits using when we just think about the actual values of a function's inputs and corresponding outputs.

Activity 1.1.3 Approximating Limits Numerically.

For each of the following tables of function values, approximate the limit in question. When you do so, approximate the values of the relevant one-sided limits as well.

(a) Approximate $\lim\limits_{x\to 1} f(x)$ using the table of values of $f(x)$ below.

Table 1.1.15

x	0.5	0.9	0.99	1	1.01	1.1	1.5
$f(x)$	8.672	9.2	9.0001	-7	8.9998	9.5	7.59

(b) Approximate $\lim_{x \to -3} g(x)$ using the table of values of $g(x)$ below.

Table 1.1.16

x	-3.5	-3.1	-3.01	-3	-2.99	-2.9	-2.5
$g(x)$	-4.41	-3.89	-4.003	-4	7.035	2.06	-4.65

(c) Approximate $\lim_{x \to \pi} h(x)$ using the table of values of $h(x)$ below.

Table 1.1.17

x	3.1	3.14	3.141	π	3.142	3.15	3.2
$h(x)$	6	6	6	undefined	5.915	6.75	8.12

(d) Are you 100% confident about the existence (or lack of existence) of any of these limits?

(e) What extra details would you like to see to increase the confidence in your estimations? Are there changes we could make to the way these functions are represented that would make these approximations better or easier to make?

Overall, there's a common theme here: in either representation (graphically or numerically), we're making a best guess at the behavior of the function values around a point. We have limited information in these estimations, and so we're doing the best we can: in graphs, we're trying our best to make sense of the lack of precision in in the scales of our visual, and in the numerical tables we are only given a limited number of points to think about. In both cases, we are hoping to see more information to add more confidence to these estimations.

We want to make the jump from estimating these limits to evaluating them, and for that to happen, we'll need to add more information and more precision about the behavior of our function.

1.2 Evaluating Limits

Adding Precision to Our Estimations

Activity 1.2.1 From Estimating to Evaluating Limits (Part 1).

Let's consider the following graphs of functions $f(x)$ and $g(x)$.

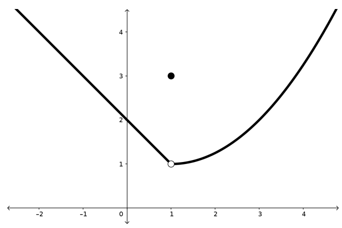


Figure 1.2.1 Graph of the function $f(x)$.

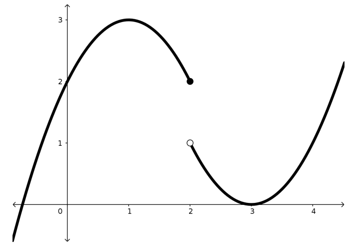


Figure 1.2.2 Graph of the function $g(x)$.

(a) Estimate the values of the following limits. If you believe that the limit does not exist, say so and explain why.

- $\lim\limits_{x\to 1^-} f(x)$
- $\lim\limits_{x\to 1^+} f(x)$
- $\lim\limits_{x\to 1} f(x)$

(b) Estimate the values of the following limits. If you believe that the limit does not exist, say so and explain why.

- $\lim\limits_{x\to 2^-} g(x)$
- $\lim\limits_{x\to 2^+} g(x)$
- $\lim\limits_{x\to 2} g(x)$

(c) Find the values of $f(1)$ and $g(2)$.

(d) For the limits and function values above, which of these are you most confident in? What about the limit, function value, or graph of the function makes you confident about your answer?

Similarly, which of these are you the least confident in? What about the limit, function value, or graph of the function makes you not have confidence in your answer?

We're going to repeat this process, but with a slight change to the representation of each function. Hopefully this will be illuminating in our attempt to add more precision to our estimations.

Activity 1.2.2 From Estimating to Evaluating Limits (Part 2).

Let's consider the following graphs of functions $f(x)$ and $g(x)$, now with the added labels of the equations defining each part of these functions.

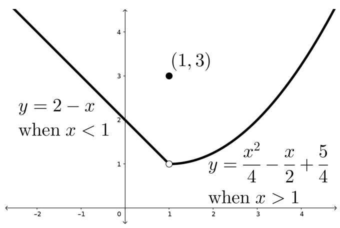


Figure 1.2.3 Graph of the function $f(x)$.

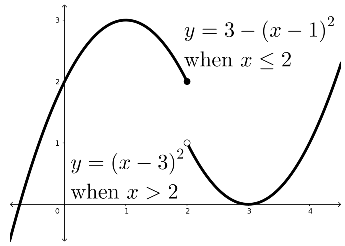


Figure 1.2.4 Graph of the function $g(x)$.

(a) Estimate the values of the following limits. If you believe that the limit does not exist, say so and explain why.

- $\lim\limits_{x \to 1^-} f(x)$
- $\lim\limits_{x \to 1^+} f(x)$
- $\lim\limits_{x \to 1} f(x)$

(b) Estimate the values of the following limits. If you believe that the limit does not exist, say so and explain why.

- $\lim\limits_{x \to 2^-} g(x)$
- $\lim\limits_{x \to 2^+} g(x)$
- $\lim\limits_{x \to 2} g(x)$

(c) Does the addition of the function rules change the level of confidence you have in these answers? What limits are you more confident in with this added information?

(d) Consider these functions without their graphs:

$$f(x) = \begin{cases} 2 - x & \text{when } x < 1 \\ 3 & \text{when } x = 1 \\ \frac{x^2}{4} - \frac{x}{2} + \frac{5}{4} & \text{when } x > 1 \end{cases}$$

$$g(x) = \begin{cases} 3 - (x-1)^2 & \text{when } x \le 2 \\ (x-3)^2 & \text{when } x > 2 \end{cases}$$

Find the limits $\lim\limits_{x \to 1} f(x)$ and $\lim\limits_{x \to 2} g(x)$. Compare these values of $f(1)$ and $g(2)$: are they related at all?

These two examples are hopefully helpful for us to see that when we are given the actual rule for a function f that connects x to the corresponding output $f(x)$, we are able to move past estimation. We suddenly have whatever level of precision we'd like, since we can immediately see what is happening with every x input to produce the corresponding $f(x)$ output.

In order for us to formalize this evaluation of limits, we're going to think about some properties of this limit object.

Limit Properties

Activity 1.2.3 Combinations of Functions.

We want to remind ourselves how we can combine functions using different operations, and how we might find outputs based on the different combinations. Our goal is to then think about how this might work with limits: how can we summarize the behavior of combinations of functions around some point?

Let's consider some functions $f(x) = x^2 + 3$ and $g(x) = x - \dfrac{1}{x}$. We'll say that the domain of both functions is $(0, \infty)$ for our own convenience.

(a) Let's consider the function $h(x) = f(x) + g(x)$. Describe at least two different ways of finding the value of $h(2)$.

(b) If we instead define the function $h(x) = f(x) - g(x)$, how would you describe at least two different ways of finding the value of $h(2)$?

(c) What about a scaled version of one of these functions? If we let $h(x) = 4f(x)$ and $j(x) = \dfrac{g(x)}{3}$, can you describe more than one way to find the value of $h(3)$ and $j(3)$?

(d) You can probably guess where we're going: we're going to define a function that is the product of f and g: $h(x) = f(x) \cdot g(x)$. Describe more than one way of evaluating $h(4)$.

(e) And finally, let's define $h(x) = \dfrac{f(x)}{g(x)}$. Now describe more than one way of finding $h(4)$.

(f) If $h(x) = \dfrac{f(x)}{g(x)}$, then are there any x-values that are in the domain of f and g (the domain is $x > 0$) that $h(x)$ cannot be defined for? Why?

Ok, we can confront this big idea: when we combine functions, we can either evaluate the combination of the functions at some x-value or evaluate each function separately and just combine the answers! Of course, there are some limitations (like when the combination isn't nicely defined because of division by 0 or something else), but this is a good framework to move forward with!

Maybe this activity was obvious for you, but it might not have been! This isn't something that we always think about with functions, even if (deep down) we know it to be true.

A nice extension that we can make is that moving past functions evaluated at a specific x-value towards descriptions of the behavior of functions *around* that specific x-value.

We'll apply this same kind of thinking (combining things by looking at each piece individually first, and then combining the answers together) to limits of combinations of functions.

Theorem 1.2.5 Combinations of Limits.

If $f(x)$ and $g(x)$ are two functions defined at x-values around, but maybe not at, $x = a$ and $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both exist, then we can evaluate limits of combinations of these functions.

1. Sums: *The limit of the sum of $f(x)$ and $g(x)$ is the sum of the limits of $f(x)$ and $g(x)$:*

$$\lim_{x\to a} (f(x) + g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$$

2. Differences: *The limit of a difference of $f(x)$ and $g(x)$ is the difference of the limits of $f(x)$ and $g(x)$:*

$$\lim_{x\to a} (f(x) - g(x)) = \lim_{x\to a} f(x) - \lim_{x\to a} g(x)$$

3. Coefficients: *If k is some real number coefficient, then:*

$$\lim_{x\to a} kf(x) = k \lim_{x\to a} f(x)$$

4. Products: *The limit of a product of $f(x)$ and $g(x)$ is the product of the limits of $f(x)$ and $g(x)$:*

$$\lim_{x\to a} (f(x) \cdot g(x)) = \left(\lim_{x\to a} f(x)\right) \cdot \left(\lim_{x\to a} g(x)\right)$$

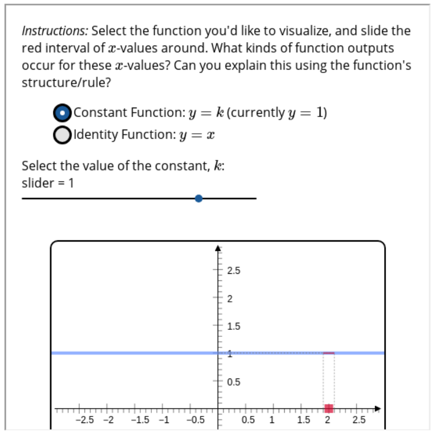
5. Quotients: *The limit of a quotient of $f(x)$ and $g(x)$ is the quotient of the limits of $f(x)$ and $g(x)$ (provided that you do not divide by 0):*

$$\lim_{x\to a} \left(\frac{f(x)}{g(x)}\right) = \frac{\left(\lim_{x\to a} f(x)\right)}{\left(\lim_{x\to a} g(x)\right)} \qquad (\textit{if } \lim_{x\to a} g(x) \neq 0)$$

We can summarize these properties: when we are thinking about our basic operations on functions, we can evaluate limits by just looking at the limits of each component function individually and then piecing those individual limit values back together.

This kind of structural “building-block” behavior is a really important one in mathematics. Whenever we define some new mathematical object, properties like this are typically good ideas for us to check in order to learn more about the object we’ve defined.

Ok, let’s move on. We’re going to turn our attention to something more concrete. We’re going to think of two function types: constant functions and the identity function.



Standalone
Embed

Theorem 1.2.6 Limits of Two Basic Functions.

Let a be some real number.

1. Limit of a Constant Function: *If k is some real number constant, then:*

$$\lim_{x\to a} k = k$$

2. Limit of the Identity Function:

$$\lim_{x\to a} x = a$$

These two functions might seem pretty simplistic (most functions that we think of are more complicated than these), but we can use these to build more functions!

Activity 1.2.4 Limits of Polynomial Functions.

We're going to use a combination of properties from Theorem 1.2.5 and Theorem 1.2.6 to think a bit more deeply about polynomial functions. Let's consider a polynomial function:

$$f(x) = 2x^4 - 4x^3 + \frac{x}{2} - 5$$

(a) We're going to evaluate the limit $\lim_{x\to 1} f(x)$. First, use the properties from Theorem 1.2.5 to re-write this limit as 4 different limits added or subtracted together.

(b) Now, for each of these limits, re-write them as products of things until you have only limits of constants and identity functions, as in Theorem 1.2.6. Evaluate your limits.

(c) Based on the definition of a limit (Definition 1.1.1), we normally say that $\lim_{x\to 1} f(x)$ is not dependent on the value of $f(1)$. Why do we say this?

(d) Compare the values of $\lim_{x\to 1} f(x)$ and $f(1)$. Why do these values feel connected?

(e) Come up with a new polynomial function: some combination of coefficients with x's raised to natural number exponents. Call your new polynomial function $g(x)$. Evaluate $\lim_{x\to -1} g(x)$ and compare the value to $g(-1)$. Explain why these values are the same.

(f) Explain why, for any polynomial function $p(x)$, the limit $\lim_{x\to a} p(x)$ is the same value as $p(a)$.

This leads us to an important result about a whole class of functions: polynomials! We can (finally) evaluate the limit of a polynomial without having to think too carefully about the distinction between the behavior of the function *around* $x = a$ and the behavior of the function *at* $x = a$.

Theorem 1.2.7 Limits of Polynomials.

If $p(x)$ *is a polynomial function and* a *is some real number, then:*

$$\lim_{x\to a} p(x) = p(a)$$

This result really just says that polynomials are friendly functions for limits: sure, a limit is really about the behavior of the function outputs around but not at $x = a$, but for polynomial functions, specifically, we can wave our hands and say "Ah, who cares, it's all the same anyways!"

Some questions that we might ask:

1. Are there other functions that have the same nice result about them that Theorem 1.2.7 says for polynomials?
2. Are there some typical functions that we'll work with where this result *doesn't* work (and we actually have to be aware of the behavior around a point instead of at it)?
3. What are we even going to use these limits for, anyways? Why do we care about these?

The answers to these questions will come slowly but surely, and we'll hopefully be able to start using these limits as a tool to think about more interesting and important topics soon: we just need to make sure we're familiar with them first.

1.3 First Indeterminate Forms

We're going to really focus on one of the main aspects of a limit in this next activity. The activity should serve two purposes:

1. We'll review a really important property or aspect of what a limit is!
2. We'll look at this thing that we already know from a slightly different perspective (or maybe just a specific perspective), and we'll discover a really important and helpful result from it!

Activity 1.3.1 Limits of (Slightly) Different Functions.

(a) Using the graph of $f(x)$ below, approximate $\lim_{x\to 1} f(x)$.

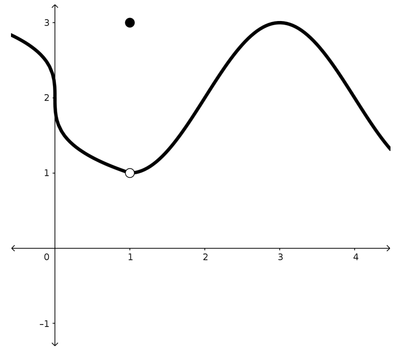


Figure 1.3.1

(b) Using the graph of the slightly different function $g(x)$ below, approximate $\lim_{x\to 1} g(x)$.

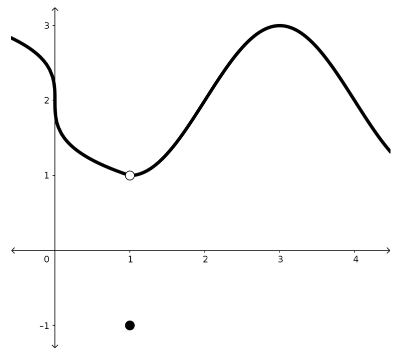


Figure 1.3.2

(c) Compare the values of $f(1)$ and $g(1)$ and discuss the impact that this difference had on the values of the limits.

(d) For the function $r(t)$ defined below, evaluate the limit $\lim_{x\to 4} r(t)$.

$$r(t) = \begin{cases} 2t - \frac{4}{t} & \text{when } t < 4 \\ 8 & \text{when } t = 4 \\ t^2 - t - 5 & \text{when } t > 4 \end{cases}$$

(e) For the slightly different function $s(t)$ defined below, evaluate the limit $\lim_{x\to 4} s(t)$.

$$s(t) = \begin{cases} 2t - \frac{4}{t} & \text{when } t \leq 4 \\ t^2 - t - 5 & \text{when } t > 4 \end{cases}$$

(f) Do the changes in the way that the function was defined impact the evaluation of the limit at all? Why not?

This is an important thing to notice: we can change our function by changing the value of the function at $x = a$ without changing the value of the limit as $x \to a$.

Theorem 1.3.3 Limits of (Slightly) Different Functions.

If $f(x)$ and $g(x)$ are two functions defined at x-values around a (but maybe not at $x = a$ itself) with $f(x) = g(x)$ for the x-values around a but with $f(a) \neq g(a)$ then $\lim_{x \to a} f(x) = \lim_{x \to a} g(x)$, if the limits exist.

Why will this be helpful? At the end of Section 1.2 we found that some functions (polynomials) are great: the limit of these functions is the same as the function value at the point (Theorem 1.2.7). This is a special case, and many functions won't be so nice to work with. But maybe we could use Theorem 1.3.3 to swap out an annoying-to-work-with function for a nice-to-work-with function!

A First Introduction to Indeterminate Forms

So before we begin applying this result, we will focus on a situation where we need it. We're going to do something strange: define a situation before we experience it.

Definition 1.3.4 Indeterminate Form.

We say that a limit has an **indeterminate form** if the general structure of the limit could take on any different value, or not exist, depending on the specific circumstances.
For instance, if $\lim_{x \to a} f(x) = 0$ and $\lim_{x \to a} g(x) = 0$, then we say that the limit $\lim_{x \to a} \left(\frac{f(x)}{g(x)} \right)$ has an indeterminate form. We typically denote this using the informal symbol $\frac{0}{0}$, as in:

$$\lim_{x \to a} \left(\frac{f(x)}{g(x)} \right) \stackrel{?}{\to} \frac{0}{0}.$$

Ok, so why do we need this definition? What does the word "indeterminate" even mean, here?

We're going to see in this next activity that this kind of $\dfrac{0}{0}$ form of a limit can actually lead to very different behavior: we call it indeterminate because we cannot determine, based solely on the form $\dfrac{0}{0}$, what the limit is or even if it will exist.

Activity 1.3.2

(a) Were going to evaluate $\lim_{x \to 3} \left(\dfrac{x^2 - 7x + 12}{x - 3} \right)$.

- First, check that we get the indeterminate form $\frac{0}{0}$ when $x \to 3$.
- Now we want to find a new function that is equivalent to $f(x) = \dfrac{x^2 - 7x + 12}{x - 3}$ for all x-values other than $x = 3$. Try factoring the numerator, $x^2 - 7x + 12$. What do you notice?
- "Cancel" out any factors that show up in the numerator and denominator. Make a special note about what that factor is.

- This function is equivalent to $f(x) = \dfrac{x^2 - 7x + 12}{x - 3}$ except at $x = 3$. The difference is that this function has an actual function output at $x = 3$, while $f(x)$ doesn't. Evaluate the limit as $x \to 3$ for your new function.

(b) Now we'll evaluate a new limit: $\lim_{x \to 1} \left(\dfrac{\sqrt{x^2+3} - 2}{x^2 - 5x + 4} \right)$.

- First, check that we get the indeterminate form $\frac{0}{0}$ when $x \to 1$.
- Now we want a new function that is equivalent to $g(x) = \dfrac{\sqrt{x^2+3} - 2}{x^2 - 5x + 4}$ for all x-values other than $x = 1$. Try multiplying the numerator and the denominator by $(\sqrt{x^2+3} + 2)$. We'll call this the "conjugate" of the numerator.
- In your multiplication, confirm that $(\sqrt{x^2+3} - 2)(\sqrt{x^2+3} + 2) = (x^2 + 3) - 4$.
- Try to factor the new numerator and denominator. Do you notice anything? Can you "cancel" anything? Make another note of what factor(s) you cancel.
- This function is equivalent to $g(x) = \dfrac{\sqrt{x^2+3} - 2}{x^2 - 5x + 4}$ except at $x = 1$. The difference is that this function has an actual function output at $x = 1$, while $g(x)$ doesn't. Evaluate the limit as $x \to 1$ for your new function.

(c) Our last limit in this activity is going to be $\lim_{x \to -2} \left(\dfrac{3 - \frac{3}{x+3}}{x^2 + 2x} \right)$.

- Again, check to see that we get the indeterminate form $\frac{0}{0}$ when $x \to -2$.
- Again, we want a new function that is equivalent to $h(x) = \dfrac{3 - \frac{3}{x+3}}{x^2 + 2x}$ for all x-values other than $x = -2$. Try completing the subtraction in the numerator, $3 - \frac{3}{x+3}$, using "common denominators."
- Try to factor the new numerator and denominator(s). Do you notice anything? Can you "cancel" anything? Make another note of what factor(s) you cancel.
- For the final time, we've found a function that is equivalent to $h(x) = \dfrac{3 - \frac{3}{x+3}}{x^2 + 2x}$ except at $x = -2$. The difference is that this function has an actual function output at $x = -2$, while $h(x)$ doesn't. Evaluate the limit as $x \to -2$ for your new function.

(d) In each of the previous limits, we ended up finding a factor that was shared in the numerator and denominator to cancel. Think back to each example and the factor you found. Why is it clear that these *must* have been the factors we found to cancel?

(e) Let's say we have some new function $f(x)$ where $\lim_{x\to 5} f(x) \overset{?}{\to} \frac{0}{0}$. You know, based on these examples, that you're going to apply *some* algebra trick to re-write your function, factor, and cancel. Can you predict what you will end up looking for to cancel in the numerator and denominator? Why?

This is great: we're applying Theorem 1.3.3 because in each algebraic manipulation, we change the domain of the function by removing some factor from the denominator!

These three algebra tricks are all we'll look at for now. In reality, there are plenty of little tricky manipulations we can use to slightly change functions, but if we focused on trying to build one for every situation we could run into, we'd spend the rest of this text just outlining different algebra tricks for different situations.

Algebra Tricks for Indeterminate Forms.

For limits with the $\frac{0}{0}$ indeterminate form, we can apply the following algebraic tricks:

1. *Factor and cancel:* This works well when we have polynomials divided by polynomials.
2. *Conjugates:* This works well when we have some difference of square roots in the numerator or denominator.
3. *Combine fractions with common denominators:* This works well when we have some subtraction with fractions inside of a numerator or denominator of another fraction.

What if There Is No Algebra Trick?

We've seen some nice examples above where we were able to use some algebra to manipulate functions in such was as to force some shared factor in the numerator and denominator into revealing itself. From there, we were able to apply Theorem 1.3.3 and swap out our problematic function with a new one, knowing that the limit would be the same.

But what if we can't do that? What if the specific structure of the function seems *resistent* somehow to our attempts at wielding algebra?

This happens a lot, and we'll investigate some more of those types of limits in Section 4.7. For now, though, let's look at a very famous limit and reason our way through the indeterminate form.

Activity 1.3.3

Let's consider a new limit:

$$\lim_{\theta\to 0} \frac{\sin(\theta)}{\theta}.$$

This one is strange!

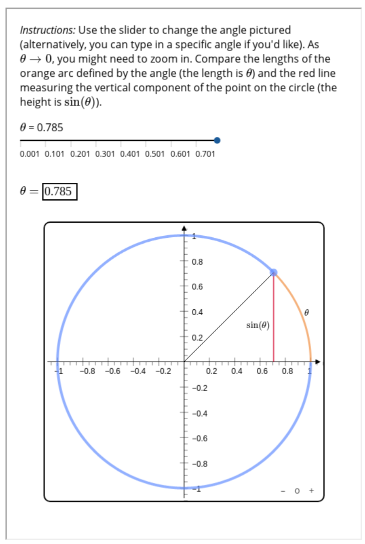
(a) Notice that this function, $f(\theta) = \dfrac{\sin(\theta)}{\theta}$, is resistent to our algebra tricks:

- There's nothing to "factor" here, since our trigonometric function is not a polynomial.
- We can't use a trick like the "conjugate" to multiply and rewrite, since there's no square roots and also only one term in the numerator.
- There aren't any fractions that we can combine by addition or subtraction.

(b) Be frustrated at this new limit for resisting our algebra tricks.

(c) Now let's think about the meaning of $\sin(\theta)$ and even θ in general. In this text, we will often use Greek letters, like θ, to represent angles. In general, these angles will be measured in radians (unless otherwise specified). So what does the sine function *do* or *tell us*? What is a radian?

(d) Let's visualize our limit, then, by comparing the length of the arc and the height of the point as $\theta \to 0$.



Standalone
Embed

(e) Explain to yourself, until you are absolutely certain, why the two lengths *must* be the same in the limit as $\theta \to 0$. What does this mean about $\lim_{\theta \to 0} \dfrac{\sin(\theta)}{\theta}$?

1.4 Limits Involving Infinity

Two types of limits involving infinity. In both cases, we'll mostly just consider what happens when we divide by small things and what happens when we divide by big things. We can summarize this here, though:

Fractions with small denominators are big, and fractions with big denominators are small.

Infinite Limits

Activity 1.4.1 What Happens When We Divide by 0?

First, let's make sure we're clear on one thing: there is no real number than is represented as some other number divided by 0.
When we talk about "dividing by 0" here (and in Section 1.3), we're talking about the behavior of some function in a limit. We want to consider what it might look like to have a function that involves division where the denominator *gets arbitrarily close to 0* (or, the limit of the denominator is 0).

(a) Remember when, once upon a time, you learned that dividing one a number by a fraction is the same as multiplying the first number by the reciprocal of the fraction? Why is this true?

(b) What is the relationship between a number and its reciprocal? How does the size of a number impact the size of the reciprocal? Why?

(c) Consider $12 \div N$. What is the value of this division problem when:

- $N = 6$?
- $N = 4$?
- $N = 3$?
- $N = 2$?
- $N = 1$?

(d) Let's again consider $12 \div N$. What is the value of this division problem when:

- $N = \frac{1}{2}$?
- $N = \frac{1}{3}$?
- $N = \frac{1}{4}$?
- $N = \frac{1}{6}$?
- $N = \frac{1}{1000}$?

(e) Consider a function $f(x) = \frac{12}{x}$. What happens to the value of this function when $x \to 0^+$? Note that this means that the x-values we're considering most are very small and positive.

(f) Consider a function $f(x) = \frac{12}{x}$. What happens to the value of this function when $x \to 0^-$? Note that this means that the x-values we're considering most are very small and negative.

Definition 1.4.1 Infinite Limit.

We say that a function $f(x)$ has an **infinite limit** at a if $f(x)$ is arbitrarily large (positive or negative) when x is sufficiently close to, but not equal to, $x = a$.
We would then say, depending on the sign of the values of $f(x)$, that:

$$\lim_{x \to a^-} f(x) = \pm\infty \qquad \lim_{x \to a^+} f(x) = \pm\infty.$$

If the sign of both one-sided limits are the same, we can say that $\lim_{x\to a} f(x) = \pm\infty$ (depending on the sign), but it is helpful to note that, by the definition of the Limit of a Function, this limit does not exist, since $f(x)$ is not arbitrarily close to a single real number.

Theorem 1.4.2 Dividing by 0 in a Limit.

If $f(x) = \dfrac{g(x)}{h(x)}$ *with* $\lim_{x\to a} g(x) \neq 0$ *and* $\lim_{x\to a} h(x) = 0$, *then* $f(x)$ *has an Infinite Limit at* a. *We will often denote this behavior as:*

$$\lim_{x\to a} f(x) \overset{?}{\to} \frac{\#}{0}$$

where # is meant to be some shorthand representation of a non-zero limit in the numerator (often, but not necessarily, some real number).

Evaluating Infinite Limits.

Once we know that $\lim_{x\to a} f(x) \overset{?}{\to} \frac{\#}{0}$, we know a bunch of information right away!

- This limit doesn't exist.
- The function $f(x)$ has a vertical asymptote at $x = a$, causing these unbounded y-values near $x = a$.
- The one sided limits *must* be either ∞ or $-\infty$.
- We only need to focus on the sign of the one sided limits! And signs of products and quotients are easy to follow.

So a pretty typical process is to factor as much as we can, and check the sign of each factor (in a numerator or denominator) as $x \to a^-$ and $x \to a^+$. From there, we can find the sign of $f(x)$ in both of those cases, which will tell us the one-sided limit.

Example 1.4.3

For each function, find the relevant one-sided limits at the input-value mentioned. If you can use a two-sided limit statement to discuss the behavior of the function around this input-value, then do so.

(a) $\left(\dfrac{2x^2 - 5x + 1}{x^2 + 8x + 16}\right)$ and $x = -4$

(b) $\left(\dfrac{4x^2 - x^5}{x^2 - 4x + 3}\right)$ and $x = 1$

(c) $\sec(\theta)$ and $\theta = \dfrac{\pi}{2}$

End Behavior Limits

Activity 1.4.2 What Happens When We Divide by Infinity?

Again, we need to start by making something clear: if we were really going to try divide some real number by infinity, then we would need to re-build our definition of what it means to divide. In the context we're in right now, we only have division defined as an operation for real (and maybe complex) numbers. Since infinity is neither, then we will not literally divide by infinity.
When we talk about "dividing by infinity" here, we're again talking about the behavior of some function in a limit. We want to consider what it might look like to have a function that involves division where the denominator *gets arbitrarily large (positive or negative)* (or, the limit of the denominator is infinite).

(a) Let's again consider $12 \div N$. What is the value of this division problem when:

- $N = 1$?
- $N = 6$?
- $N = 12$?
- $N = 24$?
- $N = 1000$?

(b) Let's again consider $12 \div N$. What is the value of this division problem when:

- $N = -1$?
- $N = -6$?
- $N = -12$?
- $N = -24$?
- $N = -1000$?

(c) Consider a function $f(x) = \frac{12}{x}$. What happens to the value of this function when $x \to \infty$? Note that this means that the x-values we're considering most are very large and positive.

(d) Consider a function $f(x) = \frac{12}{x}$. What happens to the value of this function when $x \to -\infty$? Note that this means that the x-values we're considering most are very large and negative.

(e) Why is there no difference in the behavior of $f(x)$ as $x \to \infty$ compared to $x \to -\infty$ when the sign of the function outputs are opposite ($f(x) > 0$ when $x \to \infty$ and $f(x) < 0$ when $x \to -\infty$)?

Definition 1.4.4 Limit at Infinity.

If $f(x)$ is defined for all large and positive x-values and $f(x)$ gets arbitrarily close to the single real number L when x gets sufficiently large, then we say:

$$\lim_{x\to\infty} f(x) = L.$$

Similarly, if $f(x)$ is defined for all large and negative x-values and $f(x)$ gets arbitrarily close to the single real number L when x gets sufficiently negative, then we say:

$$\lim_{x\to-\infty} f(x) = L.$$

In the case that $f(x)$ has a **limit at infinity** that exists, then we say $f(x)$ has a horizontal asymptote at $y = L$.

Lastly, if $f(x)$ is defined for all large and positive (or negative) x-values and $f(x)$ gets arbitrarily large and positive (or negative) when x gets sufficiently large (or negative), then we could say:

$$\lim_{x\to-\infty} f(x) = \pm\infty \text{ or } \lim_{x\to\infty} f(x) = \pm\infty.$$

Because the primary focus for limits at infinity is the end behavior of a function, we will often refer to these limits as **end behavior limits**.

Theorem 1.4.5 End Behavior of Reciprocal Power Functions.

If p is a positive real number, then:

$$\lim_{x\to\infty}\left(\frac{1}{x^p}\right) = 0 \textit{ and } \lim_{x\to-\infty}\left(\frac{1}{x^p}\right) = 0.$$

Theorem 1.4.6 Polynomial End Behavior Limits.

For some polynomial function:

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0$$

with n a positive integer (the degree) and all of the coefficients $a_0, a_1, \ldots, a_n$ real numbers (with $a_n \neq 0$), then

$$\lim_{x\to\infty} p(x) = \lim_{x\to\infty} a_n x^n$$

That is, the leading term (the term with the highest exponent) defines the end behavior for the whole polynomial function.

Proof.

Consider the polynomial function:

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0$$

where n is some integer and a_k is a real number for $k = 0, 1, 2, \ldots, n$. For simplicity, we will consider only the limit as $x \to \infty$, but we could easily repeat this exact proof for the case where $x \to -\infty$.

Before we consider this limit, we can factor out x^n, the variable with the highest exponent:

$$\begin{aligned} p(x) &= a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0 \\ &= x^n\left(\frac{a^n x^n}{x^n} + \frac{a_{n-1}x^{n-1}}{x^n} + \ldots + \frac{a_2x^2}{x^n} + \frac{a_1x}{x^n} + \frac{a_0}{x^n}\right) \\ &= x^n\left(a_n + \frac{a_{n-1}}{x} + \ldots + \frac{a_2}{x^{n-2}} + \frac{a_1}{x^{n-1}} + \frac{a_0}{x^n}\right) \end{aligned}$$

Now consider the limit of this product:

$$\begin{aligned} \lim_{x\to\infty} p(x) &= \lim_{x\to\infty} x^n\left(a_n + \frac{a_{n-1}}{x} + \ldots + \frac{a_2}{x^{n-2}} + \frac{a_1}{x^{n-1}} + \frac{a_0}{x^n}\right) \\ &= \left(\lim_{x\to\infty} x^n\right)\left(\lim_{x\to\infty} a_n + \frac{a_{n-1}}{x} + \ldots + \frac{a_2}{x^{n-2}} + \frac{a_1}{x^{n-1}} + \frac{a_0}{x^n}\right) \end{aligned}$$

We can see that in the second limit, we have a single constant term, a_n, followed by reciprocal power functions. Then, due to Theorem 1.4.5, we know that the second limit will by a_n, since the reciprocal power functions will all approach 0.

$$\begin{aligned}\lim_{x\to\infty} p(x) &= \left(\lim_{x\to\infty} x^n\right)\left(\lim_{x\to\infty} a_n + \frac{a_{n-1}}{x} + \ldots + \frac{a_2}{x^{n-2}} + \frac{a_1}{x^{n-1}} + \frac{a_0}{x^n}\right)\\ &= \left(\lim_{x\to\infty} x^n\right)(a_n + 0 + \ldots + 0 + 0 + 0)\\ &= \left(\lim_{x\to\infty} x^n\right)(a_n)\\ &= \lim_{x\to\infty} a_n x^n\end{aligned}$$

And so $\lim\limits_{x\to\infty} p(x) = \lim\limits_{x\to\infty} a_n x^n$ as we claimed.

Example 1.4.7

For each function, find the limits as $x \to \infty$ and $x \to -\infty$.

(a) $\left(\dfrac{2x^2 - 5x + 1}{x^2 + 8x + 16}\right)$

(b) $\left(\dfrac{4x^2 - x^5}{x^2 - 4x + 3}\right)$

(c) $\dfrac{|x|}{3x}$

Activity 1.4.3 Matching the Limits.

(a) We're going to look at four graphs of functions, as well as a list of limit statements. Match the limit statements with the graphs that match that behavior. Note that is is possible for a limit to be relevant on more than one graph.

(b) Now consider these four function definitions. Using your knowledge of limits, as well as the matching you've already done, match the definitions of these four functions with the graphs that go with them, and then also the limits that are relevant. (These limits will already be matched with the graphs, so you don't need to do further work here).

1.5 The Squeeze Theorem

Activity 1.5.1 A Weird End Behavior Limit.

In this activity, we're going to find the following limit:

$$\lim_{x\to\infty}\left(\frac{\sin^2(x)}{x^2+1}\right).$$

This limit is a bit weird, in that we really haven't looked at trigonometric functions that much. We're going to start by looking at a different limit in the hopes that we can eventually build towards this one.

(a) Consider, instead, the following limit:

$$\lim_{x\to\infty}\left(\frac{1}{x^2+1}\right).$$

Find the limit and connect the process or intuition behind it to at least one of the results from this text.

(b) Let's put this limit aside and briefly talk about the sine function. What are some things to remember about this function? What should we know? How does it behave?

(c) What kinds of values doe we expect $\sin(x)$ to take on for different values of x?

$$\underline{\qquad\qquad} \leq \sin(x) \leq \underline{\qquad\qquad}$$

(d) What happens when we square the sine function? What kinds of values can that take on?

$$\underline{\qquad\qquad} \leq \sin^2(x) \leq \underline{\qquad\qquad}$$

(e) Think back to our original goal: we wanted to know the end behavior of $\dfrac{\sin^2(x)}{x^2+1}$. Right now we have two bits of information:

- We know $\lim\limits_{x\to\infty}\left(\dfrac{1}{x^2+1}\right)$.
- We know some information about the behavior of $\sin^2(x)$. Specifically, we have some bounds on its values.

Can we combine this information?

In your inequality above, multiply $\left(\frac{1}{x^2+1}\right)$ onto all three pieces of the inequality. Make sure you're convinced about the direction or order of the inequality and whether or not it changes with this multiplication.

$$\underbrace{\frac{\qquad\qquad}{x^2+1}}_{\text{call this } f(x)} \leq \frac{\sin^2(x)}{x^2+1} \leq \underbrace{\frac{\qquad\qquad}{x^2+1}}_{\text{call this } h(x)}$$

(f) For your functions $f(x)$ and $h(x)$, evaluate $\lim\limits_{x\to\infty} f(x)$ and $\lim\limits_{x\to\infty} h(x)$.

(g) What do you think this means about the limit we're interested in, $\lim\limits_{x\to\infty}\left(\dfrac{\sin^2(x)}{x^2+1}\right)$?

Theorem 1.5.1 The Squeeze Theorem.

For some functions $f(x)$, $g(x)$, and $h(x)$ which are all defined and ordered $f(x) \leq g(x) \leq h(x)$ for x-values near $x = a$ (but not necessarily *at $x = a$ itself), and for some real number L, if we know that*

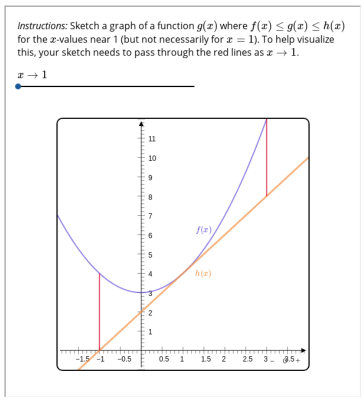
$$\lim_{x \to a} f(x) = L = \lim_{x \to a} h(x)$$

then we also know that $\lim_{x \to a} g(x) = L$.

Activity 1.5.2 Sketch This Function Around This Point.

(a) Sketch or visualize the functions $f(x) = x^2 + 3$ and $h(x) = 2x + 2$, especially around $x = 1$.

(b) Now we want to add in a sketch of some function $g(x)$, all the while satisfying the requirements of the Squeeze Theorem.



Standalone
Embed

(c) Use the Squeeze Theorem to evaluate and explain $\lim_{x \to 1} g(x)$ for your function $g(x)$.

(d) Is this limit dependent on the specific version of $g(x)$ that you sketched? Would this limit be different for someone else's choice of $g(x)$ given the same parameters?

(e) What information must be true (if anything) about $\lim_{x \to 3} g(x)$ and $\lim_{x \to 0} g(x)$?

Do we know that these limits exist? If they do, do we have information about their values?

1.6 Continuity and the Intermediate Value Theorem

Continuity as Connectedness

Continuity as Classification

Definition 1.6.1 Continuous at a Point.

The function $f(x)$ is **continuous** at an x-value in the domain of $f(x)$ if $x = a$ if $\lim_{x \to a} f(x) = f(a)$.
If $f(x)$ is not continuous at $x = a$, but one of the one-sided limits is equal to the function output, then we can define **directional continuity** at that point:

- We say $f(x)$ is **continuous on the left** at $x = a$ when $\lim_{x \to a^-} f(x) = f(a)$.
- We say $f(x)$ is **continuous on the right** at $x = a$ when $\lim_{x \to a^+} f(x) = f(a)$.

Definition 1.6.2 Continuous on an Interval.

We say that $f(x)$ is **continuous on the interval** (a, b) if $f(x)$ is continuous at every x-value with $a < x < b$.
If $f(x)$ is continuous on the right at $x = a$ and/or continuous on the left at $x = b$, then we will say that $f(x)$ is continuous on the interval $[a, b)$, $(a, b]$, or $[a, b]$, whichever is relevant.

Discontinuities

Where is a Function not Continuous?

Most of the functions that we consider in this text will be continuous everywhere that it makes sense: on their domain. That is, if there is a point defined at some x-value, it is likely that the function's limit matches the y-value of the point. More specifically, though:

- A function is discontinuous at any location that results in an infinite limit. These are locations where $f(x)$ is undefined and the limit is infinite (and so doesn't exist).
- A function is, in general, discontinuous wherever it is undefined. This seems silly to say! We probably could have left this unsaid.
- A function that is defined as a piecewise function could be discontinuous at locations where the pieces meet: maybe the limit doesn't exist, or maybe the function value is not defined, or maybe the limit exists and the function value is defined but they do not match.

Intermediate Value Theorem

Theorem 1.6.3 Intermediate Value Theorem.

If $f(x)$ is a function that is continuous on $[a, b]$ with $f(a) \neq f(b)$ and L is any real number between $f(a)$ and $f(b)$ (either $f(a) < L < f(b)$ or $f(b) < L < f(a)$), then there exists some c between a and b ($a < c < b$) such that $f(c) = L$.

This theorem was stated as early as the 5th century BCE by Bryson of Heraclea. Back then, a really interesting problem was related to "squaring the circle." That is, given a circle with some measurable radius, can we construct a square with equal area? This is obviously true, in that we can just use a square with the side length $r\sqrt{\pi}$. What we typically mean by "construct," though, is to create this square using only a compass and straightedge (a ruler without length markings) and only a finite number of steps. This was finally proven to be impossible in 1882, approximately 2300 years later.

Bryson of Heraclea knew that the square itself existed (even if he couldn't construct it) because he was able to find a circle with area less than the square (by inscribing a circle inside of the square) and a circle with area greater than the square (where the square is inscribed in the circle). Since he posited that he could increase the size of the circle in a continuous manner (without using those words), he claimed that a square with area equal to that of the circle must exist, since the area of the circle passes through all values from the smaller area to the larger area.

Chapter 2

Derivatives

2.1 Introduction to Derivatives

We'll start this off by thinking about slopes. Before we begin, you should be able to answer the following questions:

- What *is* a slope? How could you describe it?
- How do you calculate the slope of a line between two points?
- If we have a function $f(x)$ and we pick two points on the curve of the function, what does the slope of a straight line connecting the two points tell us? What kind of behavior about $f(x)$ does this slope describe?

Defining the Derivative

Activity 2.1.1 Thinking about Slopes.

We're going to calculate and make some conjectures about slopes of lines between points, where the points are on the graph of a function. Let's define the following function:

$$f(x) = \frac{1}{x+2}.$$

(a) We're going to calculate a lot of slopes! Calculate the slope of the line connecting each pair of points on the curve of $f(x)$:

- $(-1, f(-1))$ and $(0, f(0))$
- $(-0.5, f(-0.5))$ and $(0, f(0))$
- $(-0.1, f(-0.1))$ and $(0, f(0))$
- $(-0.001, f(-.001))$ and $(0, f(0))$

(b) Let's calculate another group of slopes. Find the slope of the lines connecting these pairs of points:

- $(0, f(0))$ and $(1, f(1))$
- $(0, f(0))$ and $(0.5, f(0.5))$
- $(0, f(0))$ and $(0.1, f(0.1))$

- $(0, f(0))$ and $(0.001, f(0.001))$

(c) Just to make it clear what we've done, lay out your slopes in this table:

Between $(0, f(0))$ **and...**	**Slope**
$(1, f(1))$	
$(0.5, f(0.5))$	
$(0.1, f(0.1))$	
$(0.01, f(0.01))$	
$(-0.01, f(-0.01))$	
$(-0.1, f(-0.1))$	
$(-0.5, f(-0.5))$	
$(-1, f(-1))$	

(d) Now imagine a line that is tangent to the graph of $f(x)$ at $x = 0$. We are thinking of a line that touches the graph at $x = 0$, but runs along side of the curve there instead of through it.

Make a conjecture about the slope of this line, using what we've seen above.

(e) Can you represent the slope you're thinking of above with a limit? What limit are we approximating in the slope calculations above? Set up the limit and evaluate it, confirming your conjecture.

Activity 2.1.2 Finding a Tangent Line.

Let's think about a new function, $g(x) = \sqrt{2-x}$. We're going to think about this function around the point at $x = 1$.

(a) Ok, we are going to think about this function at this point, so let's find the coordinates of the point first. What's the y-value on our curve at $x = 1$?

(b) Use a limit similar to the one you constructed in Activity 2.1.1 to find the slope of the line tangent to the graph of $g(x)$ at $x = 1$.

(c) Now that you have a slope of this line, and the coordinates of a point that the line passes through, can you find the equation of the line?

Definition 2.1.1 Derivative at a Point.

For a function $f(x)$, we say that the **derivative** of $f(x)$ at $x = a$ is:

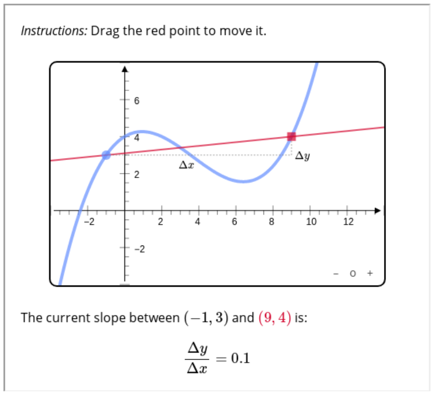
$$f'(a) = \lim_{x \to a} \left(\frac{f(x) - f(a)}{x - a} \right)$$

provided that the limit exists.
If $f'(a)$ exists, then we say $f(x)$ is **differentiable** at a.

We can investigate this definition visually. Consider the function $f(x)$ plotted below, where we will look at the point $(-1, f(-1))$. In the definition of the limit, we'll let $a = -1$, and so consider:

$$\lim_{x \to -1} \left(\frac{f(x) - f(-1)}{x - (-1)} \right).$$

Can you estimate the limit of the slope of the tangent line as $x \to -1$?



Standalone
Embed

Does it look like the limit of the slope between $(-1, f(-1))$ and $(x, f(x))$ exists as $x \to -1$? What do you think it is?

Calculating a Bunch of Slopes at Once

Activity 2.1.3 Calculating a Bunch of Slopes.

Let's do this all again, but this time we'll calculate the slope at a bunch of different points on the same function.
Let's use $j(x) = x^2 - 4$.

(a) Start calculating the following derivatives, using the definition of the Derivative at a Point:

- $j'(-2)$
- $j'(0)$
- $j'(1/3)$
- $j'(-1)$

(b) Stop calculating the above derivatives when you get tired/bored of it. How many did you get through?

(c) Notice how repetitive this is: on one hand, we have to set up a completely different limit each time (since we're looking at a different point on the function each time). On the other hand, you might have noticed that the work is all the same: you factor and cancel over and over. These limits are all ones that we covered in Section 1.3 First Indeterminate Forms, and so it's no surprise that we keep using the same algebra manipulations over and over again to evaluate these limits.

Do you notice any patterns, any connections between the x-value you used for each point and the slope you calculated at that point? You might need to go back and do some more.

(d) Try to evaluate this limit in general:

$$\begin{aligned} j'(a) &= \lim_{x\to a}\left(\frac{j(x)-j(a)}{x-a}\right)\\ &= \lim_{x\to a}\left(\frac{(x^2-4)-(a^2-4)}{x-a}\right). \end{aligned}$$

Remember, you know how this goes! You're going to do the same sorts of algebra that you did earlier!

What is the formula, the pattern, the way of finding the slope on the $j(x)$ function at any x-value, $x=a$?

(e) Confirm this by using your new formula to re-calculate the following derivatives:

- $j'(-2)$
- $j'(0)$
- $j'(1/3)$
- $j'(-1)$

We're going to try to think about the derivative as something that can be calculated in general, as well as something that can be calculated at a point. We'll define a new way of calculating it, still a limit of slopes, that will be a bit more general.

Definition 2.1.2 The Derivative Function.

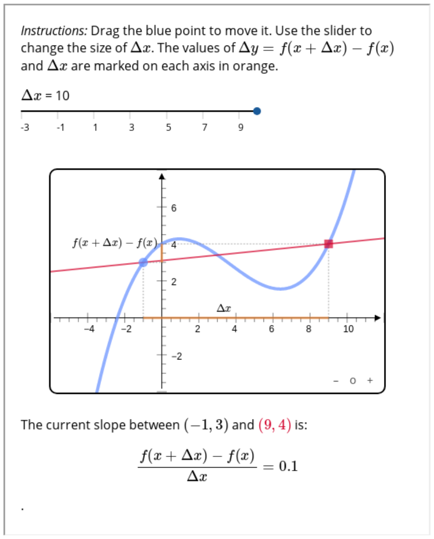
For a function $f(x)$, the derivative of $f(x)$, denoted $f'(x)$, is:

$$f'(x) = \lim_{\Delta x\to 0}\left(\frac{f(x+\Delta x)-f(x)}{\Delta x}\right)$$

for x-values in the domain of $f(x)$ where this limit exists.

This definition feels pretty different, but we can hopefully notice that this is really just calculating a slope. Notice, in the following plot, that there is a significant difference. In the visualization of the Derivative at a Point, the first point was fixed into place and the second point was the one that we moved and changed. It was the one with the variable x-value.

Notice in the following visualization that the *first* point is the one that is moveable while the *second* point is defined based on the first one (and the horizontal difference between the points, Δx). This means that we don't need to define one specific point, and can find the slope of the line tangent to $f(x)$ at some changing x-value.



Standalone
Embed

2.2 Interpreting Derivatives

What is a derivative?

This can feel like a silly question, since we're calculating it and getting used to finding them. But what is it?

In this section, we just want to remind ourselves of what this object is, how we should hold it in our minds as we move through the course, and then practice being flexible with this interpretation.

The Derivative is a Slope

Activity 2.2.1 Interpreting the Derivative as a Slope.

In Activity 2.1.1 Thinking about Slopes and Activity 2.1.2 Finding a Tangent Line, we built the idea of a derivative by calculating slopes and using them. Let's continue this by considering the function $f(x) = \frac{1}{x^2}$.

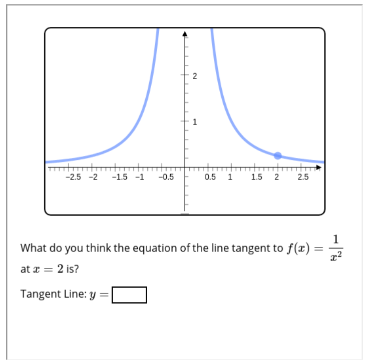
(a) Use Definition 2.1.1 Derivative at a Point to find $f'(2)$. What does this value represent?

(b) We want to plot the line that would be tangent to the graph of $f(x)$ at $x = 2$.

Remember that we can write the equation of a line in two ways:

- The equation of a line with slope m that passes through the point $(a, f(a))$ is:
$$y = m(x-a) + f(a).$$
- The equation of a line with slope m that passes the point $(0, b)$ (this is another way of saying that the y-intercept of the line is b) is:
$$y = mx + b.$$

Find the equation of the line tangent to $f(x)$ at $x = 2$. Add it to the graph of $f(x) = \frac{1}{x^2}$ below to check your equation.



Standalone
Embed

(c) This tangent line is very similar to the actual curve of the function $f(x)$ near $x = 2$. Another way of saying this is that while the slope of $f(x)$ is not always the value you found for $f'(2)$, it is close to that for x-values near 2.

Use this idea of slope to predict where the y-value of our function will be at 2.01.

(d) Compare this value with $f(2.01) = \dfrac{1}{2.01^2}$. How close was it?

The Derivative is a Rate of Change

Activity 2.2.2 Interpreting the Derivative as a Rate of Change.

This is going to somewhat feel redundant, since maybe we know that a slope is really just a rate of change. But hopefully we'll be able to explore this a bit more and see how we can use a derivative to tell us information about some specific context.
Let's say that we want to model the speed of a car as it races along a strip of the road. By the time we start measuring it (we'll call this time 0), the position the car (along the straight strip of road) is:

$$s(t) = 73t + t^2,$$

where t is time measured in seconds and $s(t)$ is the position measured in feet. Let's say that this function is only relevant on the domain $0 \leq t \leq 15$. That is, we only model the position of the car for a 15-second window as it speeds past us.

(a) How far does the car travel in the 15 seconds that we model it? What was the car's average velocity on those 15 seconds?

(b) Calculate $s'(t)$, the derivative of $s(t)$, using Definition 2.1.2 The Derivative Function. What information does this tell us about our vehicle?

(c) Calculate $s'(0)$. Why is this smaller than the average velocity you found? What does that mean about the velocity of the car?

(d) If we call $v(t) = s'(t)$, then calculate $v'(t)$. Note that this is a derivative of a derivative.

(e) Find $v'(0)$. Why does this make sense when we think about the difference between the average velocity on the time interval and the value of $v(0)$ that we calculated?

(f) What does it mean when we notice that $v'(t)$ is constant? Explain this by interpreting it in terms of both the velocity of the vehicle as well as the position.

The Derivative is a Limit

Look back at the definition of Derivative at a Point. The end of it is interesting: "provided that the limit exists." We need to keep in mind that this is a limit, and so a derivative exists or fails to exist whenever that limit exists or fails to exist.

What are some ways that a limit fails to exist?

- A limit doesn't exist if the left-side limit and the right-side limit do not match: Theorem 1.1.4 Mismatched Limits.
- A limit doesn't exist if it is an Infinite Limit.

What do each of these situations look like when we're considering the limit of slopes?

When Does a Derivative Not Exist?

1. A derivative doesn't exist at points where the slopes on either side of the point don't match.
2. A derivative doesn't exist at points with vertical tangent lines.
3. A derivative doesn't exist at points where the function is not continuous.

The Derivative is a Function

Activity 2.2.3 Interpreting the Derivative as a Function.

In Activity 2.1.3 Calculating a Bunch of Slopes, we calculated the derivative function for $j(x) = x^2 - 4$. Using the definition of The Derivative Function, we can see that $j'(x) = 2x$. Let's explore that a bit more.

(a) Sketch the graphs of $j(x) = x^2 - 4$ and $j'(x) = 2x$. Describe the shapes of these graphs.

(b) Find the coordinates of the point at $x = \frac{1}{2}$ on both the graph of $j(x)$ and $j'(x)$. Plot the point on each graph.

(c) Think back to our previous interpretations of the derivative: how do we interpret the y-value output you found for the j' function?

(d) Find the coordinates of another point at some other x-value on both the graph of $j(x)$ and $j'(x)$. Plot the point on each graph, and explain what the output of j' tells us at this point.

(e) Use your graph of $j'(x)$ to find the x-intercept of $j'(x)$. Locate the point on $j(x)$ with this same x-value. How do we know, visually, that this point is the x-intercept of $j'(x)$?

(f) Use your graph of $j'(x)$ to find where $j'(x)$ is positive. Pick two x-values where $j'(x) > 0$. What do you expect this to look like on the graph of $j(x)$? Find the matching points (at the same x-values) on the graph of $j(x)$ to confirm.

(g) Use your graph of $j'(x)$ to find where $j'(x)$ is negative. Pick two x-values where $j'(x) < 0$. What do you expect this to look like on the graph of $j(x)$? Find the matching points (at the same x-values) on the graph of $j(x)$ to confirm.

(h) Let's wrap this up with one final pair of points. Let's think about the point $(-3, 5)$ on the graph of $j(x)$ and the point $(-3, -6)$ on the graph of $j'(x)$. First, explain what the value of -6 (the output of j' at $x = -3$) means about the point $(-3, 5)$ on $j(x)$. Finally, why can we not use the value 5 (the output of j at $x = -3$) means about the point $(-3, -6)$ on $j'(x)$?

Notation for Derivatives

So far we've been using the "prime" notation to represent derivatives: the derivative of $f(x)$ is $f'(x)$. We will continue to use this notation, but we'll introduce a bunch of other ways of writing notation to represent the derivative. Each new notation will emphasize some aspect of the derivative that will serve to be useful, even though they all represent essentially the same thing.

Function	**Derivative**	**Derivative at $x = a$**	**Emphasis**
$f(x)$	$f'(x)$	$f'(a)$	The derivative is a function. The function takes in x-value inputs and returns the slope of f at that x-value.
y	y'	$y'\Big\rvert_{x=a}$	We can find slopes on any curve, not just functions. This is sometimes also used as a way to simplify the notation, especially when we want to manipulate equations involving y'.
y	$\frac{dy}{dx}$	$\frac{dy}{dx}\Big\rvert_{x=a}$	The derivative is a slope. It is $\frac{\Delta y}{\Delta x}$ as $\Delta x \to 0$, and we use dx and dy (called **differentials**) to represent Δx and Δy as the limits as $\Delta x \to 0$. This notation is also useful to tell us what the rate of change is: what is changing (in this case y) and what is it changing based on (in this case x).
$f(x)$	$\frac{d}{dx}\left(f(x)\right)$	$\frac{d}{dx}\left(f(x)\right)\Big\rvert_{x=a}$	The derivative is an action that we do to some function. We can call it an **operator**, although we won't formally define that term in this text. We'll look at this idea more in Section 3.1. We can specify what we are expecting the input variable to be, based on the differential dx in the denominator.

2.3 Some Early Derivative Rules

We are going to break this topic into two parts:

1. We will try to find some common patterns or connections between derivatives and specific functions. For instance, when we use Definition 2.1.2 The Derivative Function to build a derivative, are there patterns in the work of evaluating that limit that will allow us to get through the limit work quickly? Can we group some functions together based on how we might deal with the limit?

2. We will try to think about derivatives a bit more generally and show how we can build some basic properties to help us think about differentiating variations of the functions that we recognize.

Derivatives of Common Functions

Activity 2.3.1 Derivatives of Power Functions.

We're going to do a bit of pattern recognition here, which means that we will need to differentiate several different power functions. For our reference, a power function (in general) is a function in the form $f(x) = a(x^n)$ where n and a are real numbers, and $a \neq 0$.
Let's begin our focus on the power functions x^2, x^3, and x^4. We're going to use Definition 2.1.2 The Derivative Function a lot, so feel free to review it before we begin.

(a) Find $\frac{d}{dx}\left(x^2\right)$. As a brief follow up, compare this to the derivative $j'(x)$ that you found in Activity 2.1.3 Calculating a Bunch of Slopes. Why are they the same? What does the difference, the -4, in the $j(x)$ function do to the graph of it (compared to the graph of x^2) and why does this not impact the derivative?

(b) Find $\frac{d}{dx}\left(x^3\right)$.

(c) Find $\frac{d}{dx}\left(x^4\right)$.

(d) Notice that in these derivative calculations, the main work is in multiplying $(x + \Delta x)^n$. Look back at the work done in all three of these derivative calculations and find some unifying steps to describe how you evaluate the limit/calculate the derivative *after* this tedious multiplication was finished. What steps did you do? Is it always the same thing?

Another way of stating this is: if I told you that I knew what $(x + \Delta x)^5$ was, could you give me some details on how the derivative limit would be finished?

(e) Finish the following derivative calculation:

$$\begin{aligned}\frac{d}{dx}\left(x^5\right) &= \lim_{\Delta x\to 0}\left(\frac{(x+\Delta x)^5 - x^5}{\Delta x}\right)\\ &= \lim_{\Delta x\to 0}\left(\frac{(x^5 + 5x^4\Delta x + 10x^3\Delta x^2 + 10x^2\Delta x^3 + 5x\Delta x^4 + \Delta x^5) - x^5}{\Delta x}\right)\\ &=\rightsquigarrow \ldots\end{aligned}$$

(f) Make a conjecture about the derivative of a power function in general, $\frac{d}{dx}(x^n)$.

Something to notice here is that the calculation in this limit is really dependent on knowing what $(x + \Delta x)^n$ is. When n is an integer with $n \geq 2$, this really just translates to multiplication. If we can figure out how to multiply $(x + \Delta x)^n$ in general, then this limit calculation will be pretty easy to do. We noticed that:

1. The first term of that multiplication will combine with the subtraction of x^n in the numerator and subtract to 0.
2. The rest of the terms in the multiplication have at least one copy of Δx, and so we can factor out Δx and "cancel" it with the Δx in the denominator.

3. Once this has done, we've escaped the portion of the limit that was giving us the $\frac{0}{0}$ indeterminate form, and so we can evaluate the limit as $\Delta x \to 0$. The result is just that whatever terms still have at least one remaining copy of Δx in it "go to" 0, and we're left with just the terms that do not have any copies of Δx in them.

Triangle binomial theorem for coefficients.

Theorem 2.3.1 Power Rule for Derivatives.

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

We have shown that this is true for $n = 2, 3, 4, \ldots$, but this is also true for *any* value of n (including $n = 1$, non-integers, and non-positives). We will prove this more formally later (in Section 3.3), and until then we will be free to use this result.

Example 2.3.2

Let's confirm this Power Rule for two examples that we are familiar with.

(a) Find the derivative $\frac{d}{dx}(\sqrt{x})$ using the limit definition of the derivative function. Note that $\sqrt{x} = x^{1/2}$ and $\frac{1}{2\sqrt{x}} = \frac{1}{2}x^{-1/2}$.

(b) Find the derivative $\frac{d}{dx}\left(\frac{1}{x}\right)$ using the limit definition of the derivative function. Note that $\frac{1}{x} = x^{-1}$ and $-\frac{1}{x^2} = -x^{-2}$.

In this activity, we also found one other result.

Theorem 2.3.3 Derivative of a Constant Function.

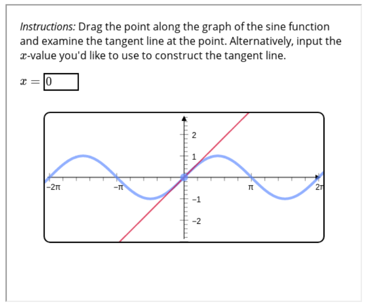
If $y = k$ where k is some real number constant, then $y' = 0$. Another way of saying this is:

$$\frac{d}{dx}(k) = 0.$$

Activity 2.3.2 Derivatives of Trigonometric Functions.

Let's try to think through the derivatives of $y = \sin(\theta)$ and $y = \cos(\theta)$. In this activity, we'll look at graphs and try to collect some information about the derivative functions. We'll be practicing out interpretations, so if you need to brush up on Section 2.2 before we start, that's fine!

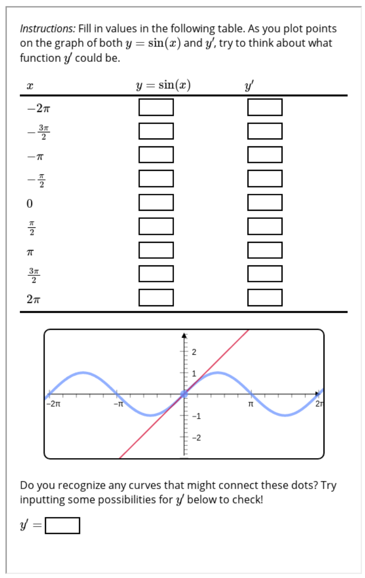
(a) The following plot includes both the graph of $y = \sin(x)$, and the line tangent to $y = \sin(x)$. Watch as the point where we build the tangent line moves along the graph, between $x = -2\pi$ and $x = 2\pi$.



Standalone
Embed

Collect as much information about the derivative, y', as you can. What do you know about it? What are some facts about the slopes of the tangent lines in this animation?

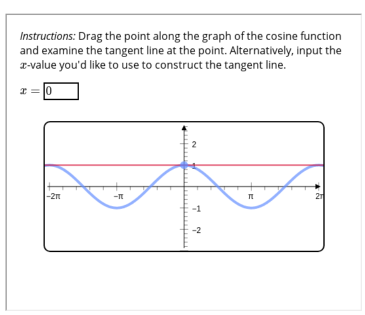
(b) We're going to get more specific here: let's find the coordinates of points that are on both the graph of $y = \sin(x)$ and it's derivative y'. Remember, to get the values for y', we're really looking at the slope of the tangent line at that point.



Standalone
Embed

(c) Let's repeat this process using the $y = \cos(x)$ function instead.

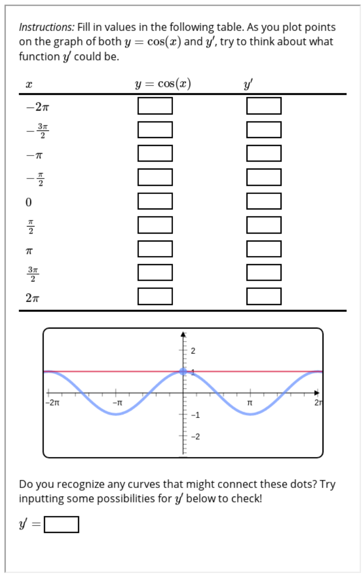
The following plot includes both the graph of $y = \cos(x)$, and the line tangent to $y = \cos(x)$. Watch as the point where we build the tangent line moves along the graph, between $x = -2\pi$ and $x = 2\pi$.



Standalone
Embed

Collect as much information about the derivative, y', as you can. What do you know about it? What are some facts about the slopes of the tangent lines in this animation?

(d) We're going to get more specific here: let's find the coordinates of points that are on both the graph of $y = \cos(x)$ and it's derivative y'. Remember, to get the values for y', we're really looking at the slope of the tangent line at that point.



Standalone
Embed

Theorem 2.3.4 Derivatives of the Sine and Cosine Functions.

$$\frac{d}{d\theta}(\sin(\theta)) = \cos(\theta)$$

$$\frac{d}{d\theta}(\cos(\theta)) = -\sin(\theta)$$

Proof.

In order to show why $\frac{d}{d\theta}(\sin(\theta)) = \cos(\theta)$ and $\frac{d}{dx}(\cos(\theta)) = -\sin(\theta)$, we will work with the limit definitions of both. Consider both:

$$\frac{d}{d\theta}(\sin(\theta)) = \lim_{\Delta\theta \to 0}\left(\frac{\sin(\theta + \Delta\theta) - \sin(\theta)}{\Delta\theta}\right)$$
$$\frac{d}{d\theta}(\cos(\theta)) = \lim_{\Delta\theta \to 0}\left(\frac{\cos(\theta + \Delta\theta) - \cos(\theta)}{\Delta\theta}\right)$$

Our goal is to re-write the numerators in both of these limits as something more usable. So far, we have been evaluating these types of limits (First Indeterminate Forms) using some algebraic manipulations. Instead of using algebra, we will use geometry.

Consider the unit circle below. We have plotted the angle θ and are reminded that the point on the circle that corresponds with the terminal side of the angle θ has coordinates $(\cos(\theta), \sin(\theta))$. We can label the sides of the triangle pictured below.

Now we consider a second point on the circle, this one formed by the terminal side of the angle $(\theta + \Delta\theta)$. This point has coordinates $(\cos(\theta + \Delta\theta), \sin(\theta + \Delta\theta))$. Note, below, that we want to find expressions for $\sin(\theta + \Delta\theta) - \sin(\theta)$ and $\cos(\theta + \Delta\theta) - \cos(\theta)$. We can find these geometrically.

Note, then, that the two triangles look to be similar triangles. In fact, we will find that in the limit as $\Delta\theta \to 0$, the length h matches the arc length $\Delta\theta$ perfectly, and thus lays at a right angle to the terminal side of the angle $\theta + \Delta\theta$. Since as $\Delta\theta \to 0$ we have $h \to \Delta\theta$, we can find the other side lengths as well: $(\sin(\theta + \Delta\theta) - \sin(\theta)) \to \Delta\theta\cos\theta$ and $(\cos(\theta + \Delta\theta) - \cos(\theta)) \to \Delta\theta\sin\theta$. So then $(\cos(\theta + \Delta\theta) - \cos(\theta)) \to -\Delta\theta\sin\theta$.

Consider, then, the limits involved in the derivative calculations that we built earlier.

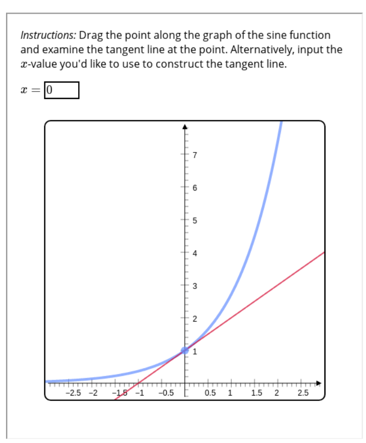
$$\begin{aligned}
\frac{d}{d\theta}(\sin(\theta)) &= \lim_{\Delta\theta\to 0}\left(\frac{\sin(\theta+\Delta\theta)-\sin(\theta)}{\Delta\theta}\right)\\
&= \lim_{\Delta\theta\to 0}\left(\frac{\Delta\theta\cos(\theta)}{\Delta\theta}\right)\\
&= \lim_{\Delta\theta\to 0}(\cos(\theta))\\
&= \cos(\theta)\\
\frac{d}{d\theta}(\cos(\theta)) &= \lim_{\Delta\theta\to 0}\left(\frac{\cos(\theta+\Delta\theta)-\cos(\theta)}{\Delta\theta}\right)\\
&= \lim_{\Delta\theta\to 0}\left(\frac{-(\cos(\theta)-\cos(\theta+\Delta\theta))}{\Delta\theta}\right)\\
&= \lim_{\Delta\theta\to 0}\left(\frac{-\Delta\sin(\theta)}{\Delta\theta}\right)\\
&= \lim_{\Delta\theta\to 0}(-\sin(\theta))\\
&= -\sin(\theta)
\end{aligned}$$

So we have shown that $\dfrac{d}{d\theta}(\sin(\theta)) = \cos(\theta)$ and $\dfrac{d}{d\theta}(\cos(\theta)) = -\sin(\theta)$ as we claimed.

Activity 2.3.3 Derivative of the Exponential Function.

We're going to consider a maybe-unfamiliar function, $f(x) = e^x$. We'll explore this function in a similar way to use thinking about the derivatives of sine and cosine in Activity 2.3.2: we'll look at a tangent line at different points, and think about the slope.

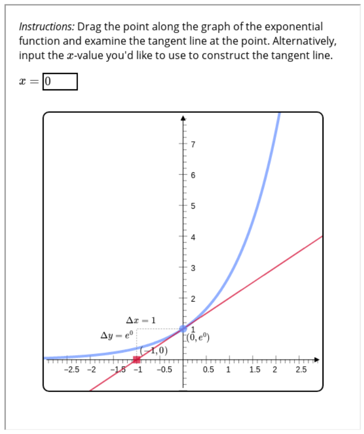
(a) The plot below includes both the graph of $y = e^x$ and the line tangent to $y = e^x$. Watch as the point moves along the curve.



Standalone
Embed

Collect as much information about the derivative, y', as you can. What do you know about it? What are some facts about the slopes of the tangent lines in this animation?

(b) There is a slightly hidden fact about slopes and tangent lines in this animation. In the following animation, we'll add (and label) one more point. Let's look at this again, this time noting the point at which this tangent line crosses the x-axis. This will make it easier to think about slopes!



Standalone
Embed

What information does this reveal about the slopes?

(c) Make a conjecture about the slope of the line tangent to the exponential function $y = e^x$ at any x-value. What do you believe the formula/equation for y' is then?

Theorem 2.3.5 Derivative of the Exponential Function.

$$\frac{d}{dx}\left(e^x\right) = e^x$$

Some Properties of Derivatives in General

Theorem 2.3.6 Combinations of Derivatives.

If $f(x)$ and $g(x)$ are differentiable functions, then the following statements about their derivatives are true.

1. Sums: *The derivative of the sum of $f(x)$ and $g(x)$ is the sum of the derivatives of $f(x)$ and $g(x)$:*

$$\begin{aligned}\frac{d}{dx}\left(f(x) + g(x)\right) &= \left(\frac{d}{dx}f(x)\right) + \left(\frac{d}{dx}g(x)\right)\\ &= f'(x) + g'(x)\end{aligned}$$

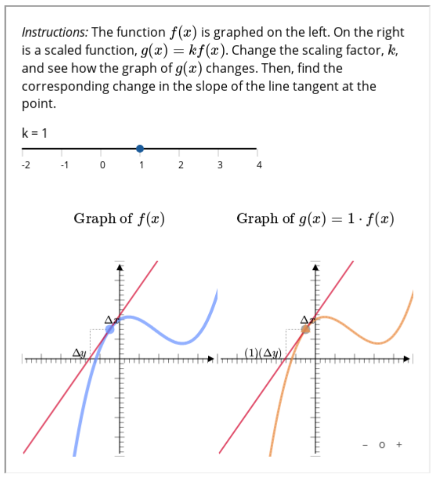
2. Differences: *The derivative of the difference of $f(x)$ and $g(x)$ is the difference of the derivatives of $f(x)$ and $g(x)$:*

$$\begin{aligned}\frac{d}{dx}\left(f(x) - g(x)\right) &= \left(\frac{d}{dx}f(x)\right) - \left(\frac{d}{dx}g(x)\right)\\ &= f'(x) - g'(x)\end{aligned}$$

3. Coefficients: *If k is some real number coefficient, then:*

$$\begin{aligned}\frac{d}{dx}\left(kf(x)\right) &= k\left(\frac{d}{dx}f(x)\right)\\ &= kf'(x)\end{aligned}$$

We can think about each of these properties through the lense of how combining these functions impacts the slopes. For instance, if we wanted to visualize the property about coefficients (that $\frac{d}{dx}\left(kf(x)\right) = k\frac{d}{dx}\left(f(x)\right)$), we can visualize this coefficient as a scaling factor.



Standalone
Embed

Example 2.3.7 Putting These Together.

Find the following derivatives:

(a) $\frac{d}{dx}\left(4x^5 - \frac{5x}{2} + 6\cos(x) - 1\right)$

Solution.

$$\begin{aligned}\frac{d}{dx}\left(4x^5 - \frac{5x}{2} + 6\cos(x) - 1\right) &= \frac{d}{dx}\left(4x^5\right) - \frac{d}{dx}\left(\frac{5x}{2}\right) + \frac{d}{dx}\left(6\cos(x)\right) - \frac{d}{dx}(1)\\ &= 4\frac{d}{dx}\left(x^5\right) - \frac{5}{2}\frac{d}{dx}(x) + 6\frac{d}{dx}\left(\cos(x)\right) - \frac{d}{dx}(1)\\ &= 4(5x^4) - \frac{5}{2}(1) + 6(-\sin(x)) - 0\\ &= 20x^4 - \frac{5}{2} - 6\sin(x)\end{aligned}$$

2.4 The Product and Quotient Rules

We saw in Theorem 2.3.6 Combinations of Derivatives that when we want to find the derivative of a sum or difference of functions, we can just find the derivatives of each function separately, and then combine the derivatives back together (by adding or subtracting). This, hopefully, is pretty intuitive: of course a slope of a sum of things is just the slopes of each thing added together!

In this section, we want to think about derivatives of product and quotients of functions. What happens when we differentiate a function that is really just two functions multiplied together?

Activity 2.4.1 Thinking About Derivatives of Products.

Let's start with two quick facts:

$$\frac{d}{dx}\left(x^3\right) = 3x^2 \text{ and } \frac{d}{dx}\left(\sin(x)\right) = \cos(x).$$

(a) What is $\frac{d}{dx}\left(x^3 + \sin(x)\right)$? What about $\frac{d}{dx}\left(x^3 - \sin(x)\right)$?

(b) Based on what you just explained, what is a reasonable assumption about what $\frac{d}{dx}\left(x^3 \sin(x)\right)$ might be?

(c) Let's just think about $\frac{d}{dx}\left(x^3\right)$ for a moment. What *is* x^3? Can you write this as a product? Call one of your functions $f(x)$ and the other $g(x)$. You should have that $x^3 = f(x)g(x)$ for whatever you used.

(d) Use your example to explain why, in general, $\frac{d}{dx}\left(f(x)g(x)\right) \neq \frac{d}{dx}\left(f(x)\right) \cdot \frac{d}{dx}\left(g(x)\right)$.

(e) Let's show another way that we know this. Think about $\sin(x)$. We know two things:

$$\sin(x) = (1)(\sin(x)) \text{ and } \frac{d}{dx}\left(\sin(x)\right) = \cos(x).$$

What, though, is $\frac{d}{dx}(1) \cdot \frac{d}{dx}\left(\sin(x)\right)$?

(f) Use all of this to reassure yourself that even though the derivative of a sum of functions is the sum of the derivatives of the functions, we will need to develop a better understanding of how the derivatives of products of functions work.

A thing that I like to think about is this: if $\frac{d}{dx}\left(f(x)g(x)\right) = f'(x)g'(x)$, then every function's derivative would be 0.

In the example with the $\sin(x)$ function, we noticed that we could write the function as $(1)(\sin(x))$. This is true of every function!

If $\frac{d}{dx}\left(f(x)g(x)\right) = f'(x)g'(x)$, then we could say that for any function $f(x)$ with a derivative $f'(x)$:

$$\begin{aligned}\frac{d}{dx}\left(f(x)\right) &= \frac{d}{dx}\left(1 \cdot f(x)\right)\\ &= \frac{d}{dx}(1)\,\frac{d}{dx}\left(f(x)\right)\\ &= 0 \cdot f'(x)\\ &= 0.\end{aligned}$$

This, obviously, can't be true! We know of *tons* of functions that have non-zero slopes...*most* of them do!

So, we hopefully have some motivation for building a rule to that helps us think about derivatives of products of functions.

The Product Rule

Activity 2.4.2 Building a Product Rule.

Let's investigate how we might differentiate the product of two functions:

$$\frac{d}{dx}\left(f(x)g(x)\right).$$

We'll use an area model for multiplication here: we can consider a rectangle where the side lengths are functions $f(x)$ and $g(x)$ that change for different values of x. Maybe x is representative of some type of time component, and the side lengths change size based on time, but it could be anything.

If we want to think about $\frac{d}{dx}\left(f(x)g(x)\right)$, then we're really considering the change in area of the rectangle.

(a) Find the area of the two rectangles. The second rectangle is just one where the input variable for the side length has changed by some amount, leading to a different side length.

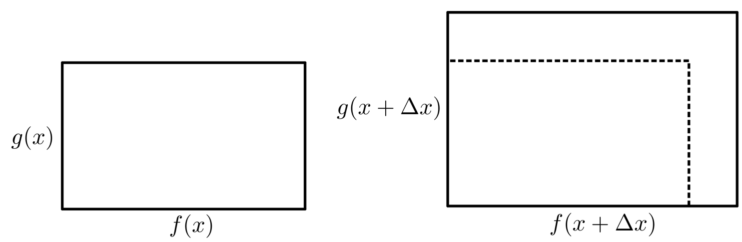


Figure 2.4.1

(b) Write out a way of calculating the difference in these areas.

(c) Let's try to calculate this difference in a second way: by finding the actual area of the region that is new in the second rectangle.

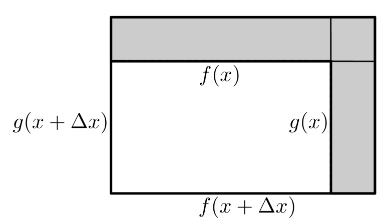


Figure 2.4.2

In order to do this, we've broken the region up into three pieces. Calculate the areas of the three pieces. Use this to fill in the following equation:

$f(x+\Delta x)g(x+\Delta x)-f(x)g(x) = \underline{\qquad\qquad\qquad\qquad}$.

(d) We do not want to calculate the total change in area: a derivative is a *rate of change*, so in order to think about the derivative we need to divide by the change in input, Δx.

We'll also want to think about this derivative as an *instantaneous* rate of change, meaning we will consider a limit as $\Delta x \to 0$. Fill in the following:

$$\begin{aligned}\frac{d}{dx}\left(f(x)g(x)\right) \lim_{\Delta x \to 0}&\left(\frac{f(x+\Delta x)g(x+\Delta x) - f(x)g(x)}{\Delta x}\right)\\ = \lim_{\Delta x \to 0}&\left(\frac{\boxed{\qquad\qquad\qquad\qquad}}{\Delta x}\right)\end{aligned}$$

We can split this fraction up into three fractions:

$$\begin{aligned}\frac{d}{dx}\left(f(x)g(x)\right) = &\lim_{\Delta x \to 0}\left(\frac{\boxed{\qquad\qquad}}{\Delta x}\right)\\ &+ \lim_{\Delta x \to 0}\left(\frac{\boxed{\qquad\qquad}}{\Delta x}\right)\\ &+ \lim_{\Delta x \to 0}\left(\frac{\boxed{\qquad\qquad}}{\Delta x}\right)\end{aligned}$$

(e) In any limit with $f(x)$ or $g(x)$ in it, notice that we can factor part out of the limit, since these functions do not rely on Δx, the part that changes in the limit. Factor these out.

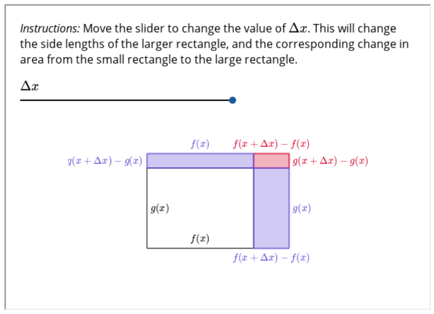
In the third limit, factor out either $\lim\limits_{\Delta x \to 0}\left(f(x+\Delta x) - f(x)\right)$ or $\lim\limits_{\Delta x \to 0}\left(g(x+\Delta x) - g(x)\right)$.

$$\begin{aligned}\frac{d}{dx}\left(f(x)g(x)\right) = f(x) &\lim_{\Delta x \to 0}\left(\frac{\boxed{\qquad\qquad}}{\Delta x}\right)\\ + g(x) &\lim_{\Delta x \to 0}\left(\frac{\boxed{\qquad\qquad}}{\Delta x}\right)\\ + &\lim_{\Delta x \to 0}\left(\boxed{\qquad\qquad}\right)\left(\lim_{\Delta x \to 0}\left(\frac{\boxed{\qquad\qquad}}{\Delta x}\right)\right)\end{aligned}$$

(f) Use Definition 2.1.2 The Derivative Function to re-write these limits. Show that when $\Delta x \to 0$, we get:

$$f(x)g'(x) + g(x)f'(x) + 0.$$

We can investigate this visual a bit more dynamically: see the differences in area as $\Delta x \to 0$. What parts are left, when $\Delta x \to 0$? What areas aren't visible?



Standalone Embed

Theorem 2.4.3 Product Rule.

If $u(x)$ and $v(x)$ are functions that are differentiable at x and $f(x) = u(x)\cdot v(x)$, then:

$$\frac{d}{dx}\left(f(x)\right) = u'(x)\cdot v(x) + u(x)\cdot v'(x).$$

For convenience, this is often written as:

$$\frac{d}{dx}\left(u\cdot v\right) = u'v + uv' \qquad \textit{or} \qquad \frac{d}{dx}\left(u\cdot v\right) = v\left(\frac{du}{dx}\right) + u\left(\frac{dv}{dx}\right).$$

Example 2.4.4

Use the Product Rule to find the following derivatives.

(a) $\dfrac{d}{dx}\left(x^3\sin(x)\right)$

Hint. Use $u = x^3$ and $v = \sin(x)$. Now find u' and v', and use:

$$\frac{d}{dx}\left(uv\right) = u'v + uv'.$$

(b) $\dfrac{d}{dx}\left((x^3+4x)e^x\right)$

(c) $\dfrac{d}{dx}\left(\sqrt{x}\cos(x)\right)$

What about Dividing?

So we can differentiate a product of functions, and the obvious next question should be about division: if we needed to build a reasonable way of differentiating a product, shouldn't we also need to build a new way of thinking about derivatives of quotients?

A nice thing that we can do is think about division as really just multiplication. For instance, if we want to differentiate $\dfrac{d}{dx}\left(\dfrac{\sin(x)}{x^2}\right)$, then we can just think about this quotient as really a product:

$$\frac{d}{dx}\left(\frac{\sin(x)}{x^2}\right) = \frac{d}{dx}\left(\frac{1}{x^2}\left(\sin(x)\right)\right).$$

Now we can just apply product rule!

$$\begin{aligned}\frac{d}{dx}\left(\frac{1}{x^2}(\sin(x))\right) &= \frac{d}{dx}\left(x^{-2}\sin(x)\right)\\ u = \sin(x) \quad & v = x^{-2}\\ u' = \cos(x) \quad & v' = -2x^{-3}\\ \frac{d}{dx}\left(\sin(x)x^{-2}\right) &= x^{-2}\cos(x) + (-2x^{-3}\sin(x))\\ &= \frac{\cos(x)}{x^2} - \frac{2\sin(x)}{x^3}\end{aligned}$$

This works great! We can *always* think about quotients as just products of reciprocals! But the problem is that we can't always differentiate these reciprocals (for now). We were able to differentiate $\frac{1}{x^2}$ by re-writing this as just a power function (with a negative exponent).

What about this flipped example:

$$\frac{d}{dx}\left(\frac{x^2}{\sin(x)}\right)?$$

In order for us to do the same thing, we need to re-write this as

$$\frac{d}{dx}\left(x^2(\sin(x))^{-1}\right)$$

but we don't know how to differentiate $(\sin(x))^{-1}$ (yet).

So let's try to build a general way of differentiating quotients.

Activity 2.4.3 Constructing a Quotient Rule.

We're going to start with a function that is a quotient of two other functions:

$$f(x) = \frac{u(x)}{v(x)}.$$

Our goal is that we want to find $f'(x)$, but we're going to shuffle this function around first. We won't calculate this derivative directly!

(a) Start with $f(x) = \frac{u(x)}{v(x)}$. Multiply $v(x)$ on both sides to write a definition for $u(x)$.

$u(x) =$

(b) Find $u'(x)$.

(c) Wait: we don't care about $u'(x)$. Right? We care about finding $f'(x)$!

Use what you found for $u'(x)$ and solve for $f'(x)$.

$f'(x) =$

(d) This is a strange formula: we have a formula for $f'(x)$ written in terms of $f(x)$! But we said earlier that $f(x) = \frac{u(x)}{v(x)}$.

In your formula for $f'(x)$, replace $f(x)$ with $\dfrac{u(x)}{v(x)}$.

$f'(x) =$

This formula is fine, but a little clunky. We'll try to re-write it in some nice ways, but it is a bit more complex than the Product Rule.

Theorem 2.4.5 Quotient Rule.

If $u(x)$ and $v(x)$ are differentiable at x and $f(x) = \dfrac{u(x)}{v(x)}$ then:

$$f'(x) = \frac{v(x)u'(x) - u(x)v'(x)}{(v(x))^2}.$$

For convenience, this is often written as:

$$\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{vu' - uv'}{v^2}.$$

Example 2.4.6

Use the Quotient Rule to find the following derivatives.

(a) $\dfrac{d}{dx}\left(\dfrac{\sin(x)}{x^2}\right)$

Once you have this derivative, confirm that it is the same as $\dfrac{\cos(x)}{x^2} - \dfrac{2\sin(x)}{x^3}$, the way that we found it using the Product Rule.

(b) $\dfrac{d}{dx}\left(\dfrac{x^2}{\sin(x)}\right)$

(c) $\dfrac{d}{dx}\left(\dfrac{x+4}{x^2+1}\right)$

Derivatives of (the Rest of the) Trigonometric Functions

You might remember that of the six main trigonometric functions, we can write four of them in terms of the other two: here are the different trigonometric functions written in terms of sine and cosine functions:

$$\tan(x) = \left(\frac{\sin(x)}{\cos(x)}\right)$$
$$\sec(x) = \left(\frac{1}{\cos(x)}\right)$$
$$\cot(x) = \left(\frac{\cos(x)}{\sin(x)}\right)$$
$$\csc(x) = \left(\frac{1}{\sin(x)}\right)$$

Example 2.4.7

Find the derivatives of the remaining trigonometric functions.

(a) $\frac{d}{dx}(\tan(x))$

Hint. Write $\frac{d}{dx}(\tan(x)) = \frac{d}{dx}\left(\frac{\sin(x)}{\cos(x)}\right)$ and use the Quotient Rule.

(b) $\frac{d}{dx}(\sec(x))$

Hint. Write $\frac{d}{dx}(\sec(x)) = \frac{d}{dx}\left(\frac{1}{\cos(x)}\right)$ and use the Quotient Rule.

(c) $\frac{d}{dx}(\cot(x))$

Hint. Write $\frac{d}{dx}(\cot(x)) = \frac{d}{dx}\left(\frac{\cos(x)}{\sin(x)}\right)$ and use the Quotient Rule.

(d) $\frac{d}{dx}(\csc(x))$

Hint. Write $\frac{d}{dx}(\csc(x)) = \frac{d}{dx}\left(\frac{1}{\cos(x)}\right)$ and use the Quotient Rule.

2.5 The Chain Rule

We've been building up some intuition and rules to help us think about differentiating different functions and combinations of functions. We can find derivatives of scaled functions, sums of functions, differences of functions, products of functions, and also quotients of functions.

In this section, we'll look at our last operation between functions: composition.

Composition and Decomposition

An important part of finding derivatives of products and quotients is identifying the component functions that are being multiplied/divided (often labeled $u(x)$ or just u and $v(x)$ or just v). From there, we find the derivatives of each of the component functions, and then use the formula from the Product Rule or Quotient Rule to put the pieces together.

Thinking about derivatives of composed functions will be the same: we'll need to identify what functions are being composed inside of other functions, and use those pieces in some formulaic way to represent the derivative. On that note, let's remind ourselves and practice working with composition (and decomposition) of functions.

Activity 2.5.1 Composition (and Decomposition) Pictionary.

This activity will involve a second group, or at least a partner. We'll go through the first part of this activity, and then connect with a second group/person to finish the second part.

(a) Build two functions, calling them $f(x)$ and $g(x)$. Pick whatever kinds of functions you'd like, but this activity will work best if these functions are in a kind of sweet-spot between "simple" and "complicated," but don't overthink this.

(b) Compose $g(x)$ inside of $f(x)$ to create $(f \circ g)(x)$, which we can also write as $f(g(x))$.

(c) Write your composed $f(g(x))$ function on a separate sheet of paper. Do not leave any indication of what your chosen $f(x)$ and $g(x)$ are. Just write your composed function by itself.

Now, pass this composed $f(g(x))$ to your partner/a second group.

(d) You should have received a new function from some other person/group. It is different than yours, but also labeled $f(g(x))$ (with different choices of $f(x)$ and $g(x)$).

Identify a possibility for $f(x)$, the outside function in this composition, as well as a possibility for $g(x)$, the inside function in this composition. You can check your answer by composing these!

(e) Write a different pair of possibilities for $f(x)$ and $g(x)$ that will still give you the same composed function, $f(g(x))$.

(f) Check with your partner/the second group: did you identify the pair of functions that they originally used?

Did whoever you passed your composed function to correctly identify your functions?

A big thing to notice here is that when we pick the pieces of functions that we think were composed inside of each other, there's not a single obvious answer. This is pretty different compared to, say, using the Quotient Rule. In these quotients, we have a natural division (literally) between the pieces. Here, it's much more subjective for us when we decide to label an "inside" function and an "outside" function.

We will build up our intuition to find a good balance for how we pick these.

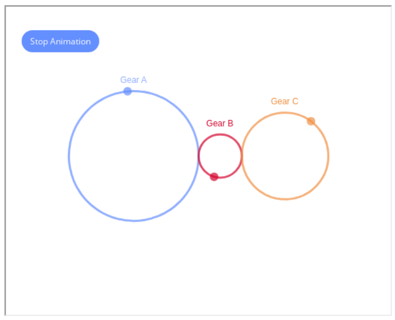
The Chain Rule, Intuitively

Before we build the Chain Rule for differentiating composed functions, we should talk about some notation. Earlier (in Subsection) we talked about the derivative notation, $\frac{dy}{dx}$. One of the things we mentioned is that while we know that the derivative is an instantaneous rate of change, this notation is helpful to tell us *what* is changing with regard to *what*.

In $\frac{dy}{dx}$, we are calculating how much the y-variable changes when x increases. If we talked about $\frac{df}{dt}$, then we are discussing how much f changes for an increase in t, whatever these variables represent.

Activity 2.5.2 Gears and Chains.

Let's think about some gears. We've got three gears, all different sizes. But the gears are linked together, and a small motor works to spin one of the gears. Since the gears are linked, when one gear spins, they all do. But since they are different sizes, they complete a different number of revolutions: the smaller ones spin more times than the larger ones, since they have a smaller circumference.



Standalone
Embed

For our purpose, let's say that Gear A is being driven by the motor.

(a) Let's try to quantify how much "faster" Gear B is spinning compared to Gear A. How many revolutions does Gear B complete in the time it takes Gear A to complete one revolution?

(b) Now quantify the speed of Gear C compared to its neighbor, Gear B. How many revolutions does Gear C complete in the time it takes Gear B to complete one revolution?

(c) Use the above relative "speeds" to compare Gear C and Gear A: how many revolutions does Gear C complete in the time it takes Gear A to complete one revolution?

More importantly, how do we find this?

(d) Now let's translate this into some derivative notation: we've really been finding rates at which one thing changes (the speed of the gear spinning) relative to another's.

Call the speed of Gear B compared to Gear A: $\frac{dB}{dA}$. Now call the speed of Gear C compared to Gear B: $\frac{dC}{dB}$. Come up with a formula to find $\frac{dC}{dA}$.

So what we need to do now is to somehow translate this intuitive idea of multiplying rates of change to build a strategy for thinking about derivatives of composed functions.

We can think of these linked gears as functions: Gear C changes based on what is happening with Gear B, which changes based on Gear A. We can translate Gear A to be an input variable, like x. Then Gear B is a function based on that: we can call it $g(x)$. Then Gear C is a function that takes in the position of Gear B (the function $g(x)$), and so we can think of it as $f(g(x))$.

To build the derivative rule for composite functions, we need to find how the "outside" function changes as the "inside" function changes ($\frac{dC}{dB}$ in this case) and multiply that by how the "inside" function changes as the input variable changes ($\frac{dB}{dA}$ here).

Theorem 2.5.1 The Chain Rule.

For the composite function $y = f(g(x))$, if we define $u = g(x)$ and $y = f(u)$, then, as long as both f and g are differentiable at u and x respectively:

$$\frac{d}{dx}\left(f(g(x))\right) = \frac{d}{du}\left(f(u)\right) \cdot \frac{d}{dx}\left(g(x)\right).$$

Alternatively, this can be written as:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} \qquad \textit{or} \qquad \frac{d}{dx}\left(f(g(x))\right) = f'(g(x)) \cdot g'(x).$$

Doing is Different than Knowing

It is lovely to know that the Chain Rule is really just linking the two rates of change together to connect a function with an input variable through a middle processing function. That's great!

But doing the Chain Rule is different than just knowing it, so let's walk through a first example. Let's find the following derivative:

$$\frac{d}{dx}\left(\sin(x^2)\right)$$

We'll call the "inside" function $u = x^2$, so we can really write the whole function (normally we're calling this y) as $y = \sin(u)$.

$$\begin{aligned}\frac{d}{dx}\left(\underbrace{\sin(\overbrace{x^2}^{u})}_{y}\right) &= \frac{dy}{du} \cdot \frac{du}{dx} \\ &= \frac{d}{du}\left(\sin(u)\right) \cdot \frac{d}{dx}\left(x^2\right)\end{aligned}$$

What we can notice, here, is that $\sin(u)$ is just a function of some variable u, and we want to find $\frac{dy}{du}$, the rate at which $y = \sin(u)$ changes with regard to its input variable. This might feel a bit strange, since u isn't just an input variable: it *means* something, since we have that $u = x^2$. This is fine! The extra $\frac{du}{dx}$ that we multiply will take care of linking this derivative to the input variable x.

$$\begin{aligned}\frac{d}{dx}\left(\underbrace{\sin(\overbrace{x^2}^{u})}_{y}\right) &= \frac{d}{du}\left(\sin(u)\right) \cdot \frac{d}{dx}\left(x^2\right) \\ &= \cos(u) \cdot 2x \\ &= \cos(x^2) \cdot 2x \\ &= 2x\cos(x^2)\end{aligned}$$

After we finished differentiating $\frac{d}{du}\left(\sin(u)\right)$, you'll notice that we used the fact that $u = x^2$ to write our combination of derivatives (the derivative of the "outside" function and the derivative of the "inside" function) in terms of the same input variable again.

The last line, rewriting $\cos(x^2) \cdot 2x$ as $2x\cos(x^2)$, is just for aesthetics.

Now you're ready to try some more examples! In each, focus on identifying a natural selection for the "inside" function, u.

Example 2.5.2

Use the Chain Rule to differentiate the following:

(a) $\frac{d}{dx}\left(\sqrt{x^2+4}\right)$

Hint. Notice that x^2+4 is composed under the square root. Use $u = x^2+4$.

(b) $\frac{d}{dx}\left(e^{\tan(x)}\right)$

Hint. Try letting $u = \tan(x)$, since it's composed inside the exponent of the exponential function.

(c) $\frac{d}{dx}\left(\sin^5(x)\right)$

Hint. You could think about this as $\frac{d}{dx}\left(\sin(x)\sin(x)\sin(x)\sin(x)\sin(x)\right)$ and try to use a very annoying product rule, but it might be easier to think about this as $\frac{d}{dx}\left((\sin(x))^5\right)$.

Generalizing the Derivative of the Exponential

Earlier, we looked at the specific derivative for $f(x) = e^x$ (Theorem 2.3.5), but we haven't talked about derivatives of other exponential functions. What about things like $y = 2^x$ or $y = \left(\frac{1}{2}\right)^x$? We can use a nice fact about exponentials and *logarithms*. We'll think more about log functions later (starting in Section 3.2), but we can think a bit about them now.

A big fact to recall is that a logarithm is a way of finding an exponent with a specific property. If we want to find the exponent that we would need to put on the number e to give us 9 as an answer, we could use $\ln(9)$.

$$e^{\ln(9)} = 9$$

This is just because logs are defined in this circular way: they are, by definition, the exponent you would need to output whatever number is inside the log.

This means that if we want to think about the number 2, but written in a different way, we can think of $e^{\ln(2)}$.

Ok, but why would we *ever* use this? This seems like a ridiculous way to write a number as basic as 2!

Consider the following:

$$2^x = \Bigg(\underbrace{e^{\ln(2)}}_{=2}\Bigg)^x$$

But we also might notice that we can re-write this using an exponent rule! We know that in general: $\left(a^b\right)^c = a^{b\cdot c}$. Let's re-write this exponential function:

$$\begin{aligned} 2^x &= \left(e^{\ln(2)}\right)^x \\ &= e^{\ln(2)\cdot x} \end{aligned}$$

Remember, $\ln(2)$ is just a number: it's specifically the number you have to put in the exponent on e to get 2. So this is just a coefficient on x. We can differentiate and use the Chain Rule!

$$\begin{aligned} \frac{d}{dx}\left(2^x\right) &= \frac{d}{dx}\left(e^{\ln(2)\cdot x}\right) \\ &= e^{\ln(2)\cdot x}\cdot \ln(2) \end{aligned}$$

Now we can remember that $e^{\ln(2)\cdot x}$ is really $\left(e^{\ln(2)}\right)^x$ which is just 2^x.

So we get $\frac{d}{dx}\left(2^x\right) = 2^x \ln(2)$. We can notice that we can re-create this with *any* (reasonable) value for the base of this exponential function.

Theorem 2.5.3 Derivative of the Generalized Exponential Function.

If $b > 0$ and $b \neq 1$, then:

$$\frac{d}{dx}\left(b^x\right) = b^x \ln(b).$$

Chapter 3

Implicit Differentiation

3.1 Implicit Differentiation

Let's quickly recap what we've done with this new calculus object, the derivative:

1. We defined the derivative at a point (Definition 2.1.1) to find the slope of a line touching a graph of a function $f(x)$ at a single point. We call this the "slope of the tangent line" at a point.

2. Once we calculated this slope, we quickly found a way to think about the derivative as a *function* (Definition 2.1.2) that connects x-values with the slope of the line tangent to $f(x)$ at that x-value.

3. We thought about how we could interpret the derivative as more than just a slope (Section 2.2). We can think about this as an instantaneous rate of change, and use it to add detail to how we think about mathematical models of different things.

4. We spent some time building up shortcuts, noticing patterns, and generalizing ways of finding these derivative functions for specific functions (Section 2.3) as well as combinations of those functions (Section 2.4 and Section 2.5).

Our goal, now, is to generalize this a bit. What happens when we push past the restriction of only dealing with *functions*? Can we think of some reasonable *non-functions* that might produce curves? Might we think about tangent lines and slopes in these contexts?

Explicit vs. Implicit Definitions

Definition 3.1.1 Explicitly and Implicitly Defined Curves.

A function or curve that is defined **explicitly** is one where the relationship between the variables is stated in with an equation like $y = f(x)$. Here, x is the input variable and we can find each corresponding value of the y-variable by applying some operations to x. As an example, we might consider the following function:

$$y = 3x + 1.$$

A function or curve that is defined **implicitly** us one where the relationship between the variables is stated with an equation connecting the variables, but not necessarily one which is "solved" for a single variable. Here, the relationship between variables is not stated with the typical "input" and "output" variables. As an example, we might consider the same function as above, but defined as:

$$y - 3x - 1 = 0.$$

Often, an implicitly defined curve is one where we *cannot* solve for a single variable by isolating it.

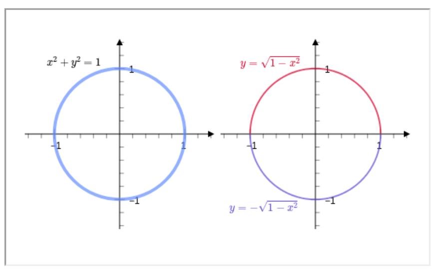
A classic and important implicitly defined curve is the unit circle:

$$x^2 + y^2 = 1.$$

We can try to isolate for y and write this as an explicitly defined curve, and end up with:

$$y = \sqrt{1 - x^2}.$$

Unfortunately, this only displays the upper half of the circle, since the square root will only output positive values. In this case, we can get around this by defining the circle with two functions.



Standalone
Embed

As we move forward, let's work with this circle using the implicitly defined version ($x^2 + y^2 = 1$). How might we find a slope of a line tangent to this circle at some point?

Using a Derivative as an Operator

Let's recall back to Subsection Notation for Derivatives. We talked about how we can use the notation $\frac{d}{dx}(f(x))$ as a kind of action: the notation says "find the derivative of $f(x)$ with respect to x." When we say "with respect to x," we mean that we are treating x as an input variable, and trying to find out how f changes based on changes to that input. The notation says, "find the rate at which $f(x)$ changes as x increases."

Because this notation is a call to action, we can use it when we're dealing with an equation. We can call back to our early algebra days, where we learn that whatever we do to one side of an equation needs to be done to the other side as well, in order to maintain the equality.

We can apply this to our derivative actions: we can differentiate both sides of an equation!

Activity 3.1.1 Thinking about the Chain Rule.

(a) Explain to someone how (and why) we use the The Chain Rule to find the following derivative:

$$\frac{d}{dx}\left(\sqrt{\sin(x)}\right).$$

(b) Let's say that $f(x) = \sin(x)$. Explain how we find the following derivative:

$$\frac{d}{dx}\left(\sqrt{f(x)}\right).$$

How is this different, or not different, than the previous derivative?

(c) Let's say that we have some other function, $g(x)$. Explain how we find the following derivative:

$$\frac{d}{dx}\left(\sqrt{g(x)}\right).$$

How is this different, or not different, than the previous derivatives?

(d) What is the difference between the following derivatives:

$$\frac{d}{dx}\left(\sqrt{x}\right) \qquad \frac{d}{dx}\left(\sqrt{y}\right) \qquad \frac{d}{dy}\left(\sqrt{y}\right)$$

Because we'll be applying our derivative notation to pieces of some equation, we'll need to be very aware of where we need to apply the Chain Rule.

Now we can look at some examples of implicitly defined curves and think about questions about the derivative. Let's start with our circle.

Activity 3.1.2 Slopes on a Circle.

Visualize the unit circle. Feel free to draw one, or find the picture above. We're going to think about slopes on this circle.

(a) Point out the locations on the unit circle where you would expect to see tangent lines that are perfectly horizontal. What do you think the value of the derivative, $\frac{dy}{dx}$, would be at these points?

(b) Point out the locations on the unit circle where you would expect to see tangent lines that are perfectly vertical. What do you think the value of the derivative, $\frac{dy}{dx}$, would be at these points?

(c) Find the point(s) where $x = \frac{1}{2}$. What do you think the value of the derivative, $\frac{dy}{dx}$, would be at these points?

(d) For the unit circle defined by the equation $x^2 + y^2 = 1$, apply the derivative to both sides of this equation to get the following:

$$\frac{d}{dx}\left(x^2 + y^2\right) = \frac{d}{dx}(1)$$
$$\frac{d}{dx}\left(x^2\right) + \frac{d}{dx}\left(y^2\right) = \frac{d}{dx}(1)$$

Carefully consider each of these derivatives (each of the terms). Which of these will you need to apply the Chain Rule for?

(e) Differentiate. Solve for $\frac{dy}{dx}$ or y', whichever notation you decide to use.

(f) Go back to the first few questions, and try to answer them again:

(a) Find the locations of any horizontal tangent lines (where $\frac{dy}{dx} = 0$).

(b) Find the locations of any vertical tangent lines (where $\frac{dy}{dx}$ doesn't exist, or where you would divide by 0).

(c) Find the values of $\frac{dy}{dx}$ for the points on the circle where $x = \frac{1}{2}$.

Example 3.1.2

Let's repeat some of this process, but using a new curve. Consider the curve defined by the equation:

$$y^2 = x^3 - x + 1.$$

This curve is a special curve with some interesting mathematical properties, and is actually a part of a family of curves called **elliptic curves**. For now, let's just consider it as a fun curve to look at, and use implicit differentiation to think about it.

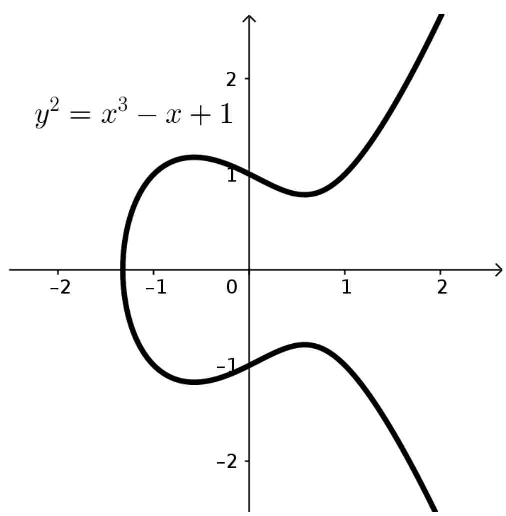


Figure 3.1.3

(a) Mark the locations on the curve where it looks like the curve will have horizontal tangent lines. How many did you find?

(b) Mark the locations on the curve where it looks like the curve will have vertical tangent lines. How many did you find?

(c) Find the point(s) where $x = 0$. What do you think the value of the derivative, $\frac{dy}{dx}$, would be at these points?

(d) For the elliptic curve defined by the equation $y^2 = x^3 - x + 1$, apply the derivative to both sides of this equation:

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

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

Carefully consider each of these derivatives (each of the terms). Which of these will you need to apply the Chain Rule for?

(e) Differentiate. Solve for $\frac{dy}{dx}$ or y', whichever notation you decide to use.

Hint 1. Make sure to use the Chain Rule when necessary!

Hint 2. $\frac{d}{dx}\left(y^2\right) = \frac{d}{dy}\left(y^2\right)\frac{dy}{dx} = 2y \cdot \frac{dy}{dx}$ or $2yy'$

(f) Go back to the first few questions, and try to answer them again:

(a) Find the locations of any horizontal tangent lines (where $\frac{dy}{dx} = 0$).

(b) Find the locations of any vertical tangent lines (where $\frac{dy}{dx}$ doesn't exist, or where you would divide by 0).

(c) Find the values of $\frac{dy}{dx}$ for the points on the curve where $x = 0$.

This example was pretty similar to the first activity: in both of these, we use the Chain Rule to differentiate $\frac{d}{dx}\left(y^2\right)$. Let's look at another example.

Activity 3.1.3 .

Let's consider a new curve:

$$\sin(x) + \sin(y) = x^2y^2.$$

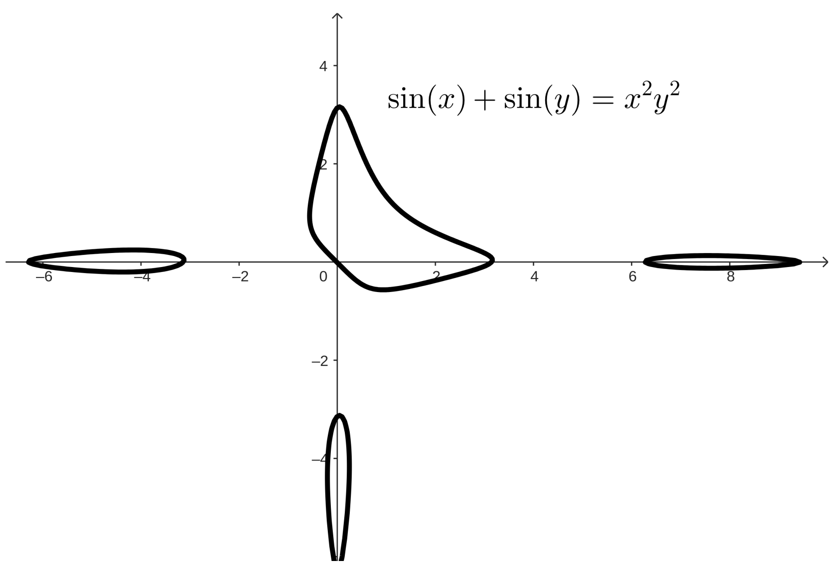


Figure 3.1.4

(a) We are going to find $\frac{dy}{dx}$ or y'. Let's dive into differentiation:

$$\frac{d}{dx}(\sin(x) + \sin(y)) = \frac{d}{dx}\left(x^2y^2\right)$$

$$\frac{d}{dx}(\sin(x)) + \frac{d}{dx}(\sin(y)) = \frac{d}{dx}\left(x^2y^2\right)$$

Think carefully about these derivatives. For each of the three, how will you approach it? What kinds of nuances or rules or strategies will you need to think about? Why?

(b) Implement your ideas or strategies from above to differentiate each term.

(c) Now we need to solve for $\frac{dy}{dx}$ or y', whichever you are using. While this equation can look complicated, we can notice something about the "location" of $\frac{dy}{dx}$ or y' in our equation.

Why do we always know that $\frac{dy}{dx}$ or y' will be *multiplied* on a term whenever it shows up?

(d) Now that we are confident that we will *always* know that we are multiplying this derivative, we can employ a consistent strategy:

(a) Rearrange our equation so that every term with a $\frac{dy}{dx}$ or y' is on one side, and everything without is on the other.

(b) Now we are guaranteed that $\frac{dy}{dx}$ or y' is a common factor: factor it out.

(c) Now we can solve for $\frac{dy}{dx}$ or y' by dividing!

Solve for $\frac{dy}{dx}$ or y' in your equation!

(e) Build the equation of a line that lays tangent to the curve at the origin. Does the value of $\frac{dy}{dx}$ at $(0,0)$ look reasonable to you?

Some Summary and Strategy

Let's wrap this up with some general strategy and summary of what we've seen.

The first thing we can notice is that we have talked through how to employ two of the three big derivative rules: we used Chain Rule throughout these examples, and in Activity 3.1.3 we needed to use the Product Rule in order to differentiate $\frac{d}{dx}\left(x^2y^2\right)$. We have a glaring omission from our examples so far, though. Where is the Quotient Rule?

In these implicitly defined curves, we can manipulate the equations to never have to think about division! Consider the curve:

$$\frac{\sin(x)}{x} + \frac{\sin(y)}{x} = xy^2.$$

Graph this curve in a graphing utility like desmos. Does it look any different than the curve in Activity 3.1.3?

The only difference, really, is that the curve with the division is not defined at $x = 0$. As long as we keep those domain issues in mind, we can multiply everything by x to get our familiar equation:

$$\sin(x) + \sin(y) = x^2y^2.$$

And there we go, we never have to think about the Quotient Rule in these kinds of contexts!

So we really only need a strategy for using the Chain Rule and the Product Rule.

Strategy for Implicit Differentiation.

- We use the *Chain Rule* whenever we differentiate something like $\frac{d}{dy}(f(y))$. We differentiate whatever the function is, and multiply by the derivative of y: $f'(y)y'$.

  This generalizes more: any time the variable in our derivative notation does not match the variable in the function/term, we need to use the Chain Rule:

  $$\frac{d}{dy}(e^x) \qquad \frac{d}{dt}(\sin(x)) \qquad \frac{d}{dx}(y^4)$$

- We use the *Product Rule* whenever we differentiate a term with some combination of x and y variables. More generally, we can use the Product Rule any time we have a combination of at least two variables. We have to treat these as different kinds of functions getting multiplied!

  $$\frac{d}{dy}(xe^y) \qquad \frac{d}{dt}(\cos(t)\sin(x)) \qquad \frac{d}{dx}\left(y^4\sqrt{x+1}\right)$$

From here on out, we will use the ideas of implicit differentiation to accomplish two things:

1. We have a bit more flexibility with how we think of derivatives! We do not need to be restricted to only thinking about functions in order to think about rates of change or slopes at a point. We can think about any curve, any relationship between variables, and think about the relationship between them based on how one changes with regard to the other.

2. Implicit differentiation will be a very useful tool. Even when we have functions that can be written explicitly, they might be hard to deal with -- overly complex or maybe involving functions that we aren't used to. It is absolutely possible, and a really useful strategy, to re-write the relationship between variables implicitly! We can maybe create a version of these equations that we can differentiate!

We're going to use this second idea first: in the next section we'll be thinking about inverse functions. We do not have any idea of how to think about derivatives of logarithmic functions, like $y = \ln(x)$.

We can re-write this:

$$y = \ln(x) \longleftrightarrow x = e^y.$$

This second representation is something we can differentiate!

Similarly, if we wanted to think about the derivative of $y = \tan^{-1}(x)$, we might instead think about re-writing this:

$$y = \tan^{-1}(x) \longleftrightarrow x = \tan(y).$$

There are some weird issues to think about with the domains and ranges of these functions, but this is how we'll approach these derivatives next.

3.2 Derivatives of Inverse Functions

We should start here by saying: we're going to be thinking about inverse functions, and so maybe it will be helpful to recap some facts about inverse functions.

- If $y = f(x)$ is some function, then we can use the inverse function to represent this relationship between variables: $x = f^{-1}(y)$.Some examples:
  - $y = e^x \longleftrightarrow x = \ln(y)$. That is, the logarithm function "solves" for the exponent (sometimes this is easier to just say that the logarithm *is* the exponent).
  - $y = \sin(\theta) \longleftrightarrow \theta = \sin^{-1}(y)$. That is, this inverse sine function (or sometimes $\arcsin(y)$) finds the angle at which sine of that angle is y. With these trigonometric functions, we need to make some restrictions: because there are an infinite number of angles that will produce the same output of the sine function (reflecting the angle across the y-axis will do it, as will adding any multiple of 2π), we restrict the angles that the inverse sine function can output to being in $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.
- Based on this re-representation above, we can always compose a function and its inverse to get the identity function, $y = x$. In general, if $y = f(x)$ has an inverse function f^{-1}, then $\left(f \circ f^{-1}\right)(x) = f\left(f^{-1}(x)\right) = x$. Similarly, we can compose in the opposite order: $\left(f^{-1} \circ f\right)(x) = f^{-1}\left(f(x)\right) = x$. This can be a bit trickier to think about for the inverse trigonometric functions, since this only works on intervals of x where that inverse is defined. So we end up with strange things like:

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

  This is because the inverse sine function finds only angles in the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, and the angles $\frac{3\pi}{2}$ and $-\frac{\pi}{2}$ are *coterminal* (and so have the same output from the sine function).

With these small facts in mind, we can think about derivatives of inverse functions.

Wielding Implicit Differentiation

We're going to do a very cool thing: in order to find derivatives of inverse functions, we can invert the relationship between x and y, and then use Implicit Differentiation to find $\frac{dy}{dx}$.

Activity 3.2.1 Building the Derivative of the Logarithm.

We're going to accomplish two things here:

1. By the end of this activity, we'll have a nice way of thinking about $\frac{d}{dx}(\ln(x))$, and we will now be able to differentiate functions involving logarithms!
2. Throughout this activity, we're going to develop a way of approaching derivatives of inverse functions more generally. Then we can apply this framework to other functions!

Let's think about this logarithmic function!

(a) We have stated (a couple of times now) how we define the log function:

$$y = e^x \longleftrightarrow x = \ln(y).$$

This arrow goes both directions: the log function is the inverse of the exponential, but the exponential is the inverse of the log function!

Can you re-write the relationship $y = \ln(x)$ using its inverse (the exponential)?

(b) For your inverted $y = \ln(x)$ from above (it should be $x =$ ____), apply a derivative operator to both sides, and use implicit differentiation to find $\frac{dy}{dx}$ or y'.

(c) A weird thing that we can notice is that when we use implicit differentiation, it is common to end up with our derivative written in terms of both x and y variables. This makes sense for our earlier examples: we needed specific coordinates of the point on the circle, for instance, to find the slope there.

But if $y = \ln(x)$, we want $\frac{dy}{dx}$ or y' to be a function of x:

$$f(x) = \ln(x) \longrightarrow f'(x) = \underline{\qquad}.$$

Your derivative is written with only y values.

Since $y = \ln(x)$, replace any instance of y with the log function. What do you have left?

(d) Remember that $y = \ln(x)$. Substitute this into your equation for $\frac{dy}{dx}$. Can you write this in a pretty simplistic way?

(e) Before we go much further, we should be a bit careful. What is the domain of this derivative?

What are the values of x where $\frac{d}{dx}(\ln(x))$ makes sense to think about?

Theorem 3.2.1 Derivative of the Logarithmic Function.

$$\frac{d}{dx}(\ln(x)) = \frac{1}{x}$$

for $x > 0$.
Further, since $\log_b(x) = \frac{\ln(x)}{\ln(b)}$ *(for* $b \geq 0$ *and* $b \neq 1$*), we can say that:*

$$\frac{d}{dx}(\log_b(x)) = \frac{1}{x\ln(b)}$$

for $x > 0$.

Derivatives of the Inverse Trigonometric Functions

Let's try a similar thing with inverse trigonometric functions. We'll start with the inverse sine function, $y = \sin^{-1}(x)$.

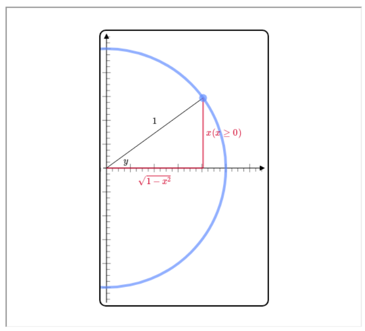
Activity 3.2.2 Finding the Derivative of the Inverse Sine Function.

We're going to do the same trick, except that there will be a couple of small differences due to thinking specifically about trigonometric functions.

Let's think about the function $y = \sin^{-1}(x)$. We know that this is equivalent to $x = \sin(y)$ (for y-values in $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$).

(a) Move the point around the portion of the unit circle in the graph below. Convince yourself that:

- $\sin(y) = x$
- $\sin(y) \geq 0$ when $0 \leq y \leq \frac{\pi}{2}$
- $\sin(y) < 0$ when $-\frac{\pi}{2} \leq y < 0$



Standalone
Embed

What is $\cos(y)$ in this figure? Does the sign change depending on the value of y?

(b) Use implicit differentiation and the equation $x = \sin(y)$ to find $\frac{dy}{dx}$ or y'

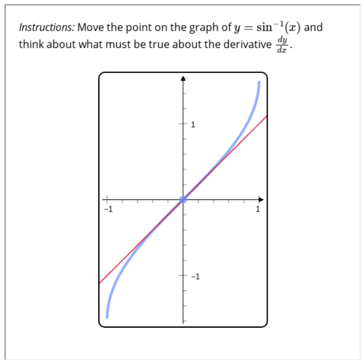
(c) If you still have your derivative written in terms of y, make sure to write $\cos(y)$ in terms of x!

(d) Let's think about the domain of this derivative: what x-values make sense to think about?

Think about this both in terms of what x-values reasonably fit into your formula of $\frac{d}{dx}\left(\sin^{-1}(x)\right)$ as well as the domain of the inverse sine function in general.

(e) Notice that in the denominator of $\frac{d}{dx}\left(\sin^{-1}(x)\right)$, you have a square root. Based on that information (and the visual above), what do you expect to be true about the sign of the derivative of the inverse sine function?

Confirm this by playing with the graph of $y = \sin^{-1}(x)$ below.



Standalone
Embed

(f) Investigate the behavior of $\frac{dy}{dx}$ at the end-points of the function: at $x = -1$ and $x = 1$. What do the slopes look like they're doing, graphically?

How does this work when you look at the function you built above? What happens when you try to find $\left.\dfrac{dy}{dx}\right|_{x=-1}$ or $\left.\dfrac{dy}{dx}\right|_{x=1}$?

Let's repeat the process to find the derivatives of $y = \tan^{-1}(x)$ and $y = \sec^{-1}(x)$.

Activity 3.2.3 Building the Derivatives for Inverse Tangent and Secant.

(a) Consider the triangle representing the case when $y = \tan^{-1}(x)$.

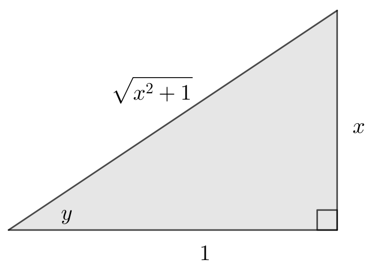


Figure 3.2.2

For $x = \tan(y)$, find $\frac{dy}{dx}$ using implicit differentiation. Find an appropriate expression for $\sec(y)$ based on the triangle above, but we will refer back to the version with the $\sec(y)$ in it later.

(b) Consider the triangle representing the case when $y = \sec^{-1}(x)$.

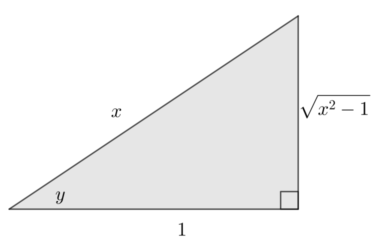
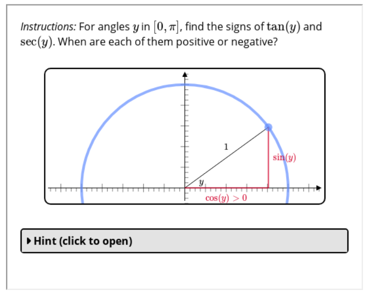


Figure 3.2.3

For $x = \sec(y)$, find $\frac{dy}{dx}$ using implicit differentiation. Find an appropriate expression for $\sec(y)$ and $\tan(y)$ based on the triangle above, but we will refer back to the version with the functions of y in it later.

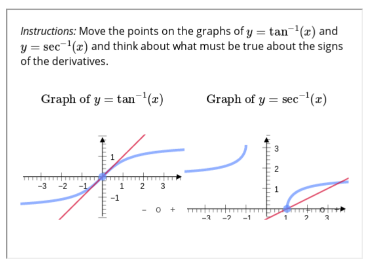
(c) Here's a graph of just the unit circle for angles $[0, \pi]$. We are choosing to focus on this region, since these are the angles that the inverse tangent and inverse secant functions will return to us. We want to investigate the signs of $\tan(y)$ and $\sec(y)$.



Standalone
Embed

(d) Go back to our derivative expressions for both the inverse tangent and inverse secant functions. What do we know about the signs of these derivatives?

(e) Confirm your idea about the sign of the derivatives by investigating the graphs of each function.



Standalone
Embed

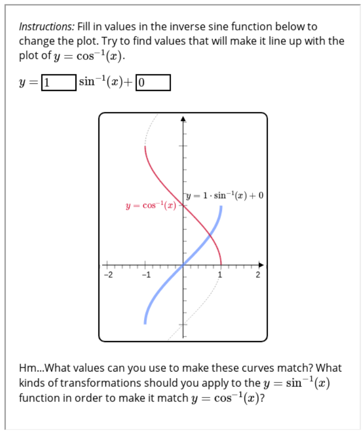
(f) How do we need to write these derivatives, when we write them in terms of x to account for the sign of the derivative?

It is important to think carefully about how things might change when we start thinking about other trigonometric functions. For instance, what happens when we think about $y = \cos^{-1}(x)$ instead? We *could* repeat the process from Activity 3.2.2 with $y = \cos^{-1}(x)$ instead (and we'll do that for $y = \tan^{-1}(x)$), but for now let's think about the graph of $y = \cos^{-1}(x)$.

Activity 3.2.4 Connecting These Inverse Functions.

We're going to look at a graph of $y = \cos^{-1}(x)$, but we're specifically going to try to compare it to the graph of $y = \sin^{-1}(x)$. We'll use some graphical transformations to make these functions match up, and then later we'll think about derivatives.

(a) Ok, consider the graph of $y = \cos^{-1}(x)$ and a transformed version of the inverse sine function. Apply some graphical transformations to make these match!



Standalone Embed

(b) It might be fun to think about another reason that this connection between $\sin^{-1}(x)$ and $\cos^{-1}(x)$ exists.

Consider this triangle:

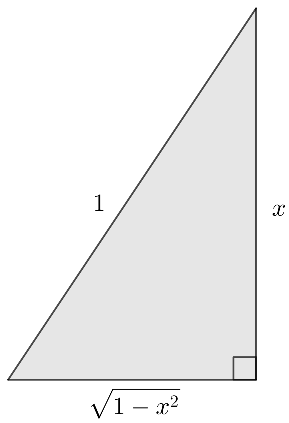


Figure 3.2.4

We're going to think about these inverse trigonometric functions as angles: let $\alpha = \cos^{-1}(x)$ and $\beta = \sin^{-1}(x)$. We can re-write these as:

$$\cos(\alpha) = x$$
$$\sin(\beta) = x.$$

Can you fill in your triangle using this information?

Why does $\alpha + \beta = \frac{\pi}{2}$? Convince yourself that this is what we did with the graphical transformations above, as well.

(c) Use this equation above, re-writing $\cos^{-1}(x)$ as some expression involving the inverse sine function, and then find

$$\frac{d}{dx}\left(\cos^{-1}(x)\right).$$

We could repeat this task to try to connect the graph of $y = \tan^{-1}(x)$ with $y = \cot^{-1}(x)$ as well as the graph of $y = \sec^{-1}(x)$ with $y = \csc^{-1}(x)$, but we can hopefully see what will happen. In each case, we

have the same kind of connection that we saw in the triangle, since these are cofunctions of each other!
We can summarize by believing that:

$$\frac{d}{dx}\left(\cos^{-1}(x)\right) = -\frac{d}{dx}\left(\sin^{-1}(x)\right) = -\frac{1}{\sqrt{1-x^2}}$$
$$\frac{d}{dx}\left(\cot^{-1}(x)\right) = -\frac{d}{dx}\left(\tan^{-1}(x)\right) = -\frac{1}{x^2+1}$$
$$\frac{d}{dx}\left(\csc^{-1}(x)\right) = -\frac{d}{dx}\left(\sec^{-1}(x)\right) = -\frac{1}{|x|\sqrt{x^2-1}}$$

Theorem 3.2.5 Derivatives of the Inverse Trigonometric Functions.

$$\frac{d}{dx}\left(\sin^{-1}(x)\right) = \frac{1}{\sqrt{1-x^2}} \qquad \textit{Domain: } -1 < x < 1$$
$$\frac{d}{dx}\left(\cos^{-1}(x)\right) = -\frac{1}{\sqrt{1-x^2}} \qquad \textit{Domain: } -1 < x < 1$$
$$\frac{d}{dx}\left(\tan^{-1}(x)\right) = \frac{1}{x^2+1} \qquad \textit{Domain: all Real numbers}$$
$$\frac{d}{dx}\left(\cot^{-1}(x)\right) = -\frac{1}{x^2+1} \qquad \textit{Domain: all Real numbers}$$
$$\frac{d}{dx}\left(\sec^{-1}(x)\right) = \frac{1}{|x|\sqrt{x^2-1}} \qquad \textit{Domain: } x < -1 \textit{ and } x > 1$$
$$\frac{d}{dx}\left(\csc^{-1}(x)\right) = -\frac{1}{|x|\sqrt{x^2-1}} \qquad \textit{Domain: } x < -1 \textit{ and } x > 1$$

3.3 Logarithmic Differentiation

We're going to start with a quick fact about logs and their derivatives. The derivative rule for $\frac{d}{dx}\left(\ln(x)\right)$ is still relatively new for us, so it is ok to still be getting comfortable with it, but we should notice this nice fact.

Fact 3.3.1 Derivative of the Log of a Function.

$$\frac{d}{dx}\left(\ln(f(x))\right) = \frac{f'(x)}{f(x)} \quad (\textit{when } f(x) > 0)$$

Note that there's nothing really special going on here: this is just an application of the The Chain Rule to the Derivative of the Logarithmic Function.

Throughout this section, the goal is to convince any open-minded readers of one thing:

Logs are friends.

Let us be informal and technically not quite correct but hopefully clear in this. Logs really are friendly mathematical objects. They were *created* to be friendly objects! In a time when doing arithmetic with large numbers was difficult due to a lack of computing technology, logs were introduced to make arithmetic easier.

The general idea is that, if there is some kind of hierarchy of operations, then logs transform operations between things into different operations that are lower on the hierarchy of operations. So logs turn things like products (repeated addition) and quotients (repeated subtraction) into addition and subtraction. Logs turn exponents (repeated multiplication) into coefficients.

Using math notation, we can write the following log properties.

Fact 3.3.2 Properties of Logarithms. *We will make use of the following properties of logarithms.*

- $\ln(xy) = \ln(x) + \ln(y)$
- $\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)$
- $\ln(x^y) = y\ln(x)$

Because of the domain of the log function, we need $x, y > 0$ for these properties to make sense. We will use them relatively loosely, with functions that take on negative and positive values, and not worry too much about the domain issues when we don't need to.

Logs Are Friends!

Ok, so how will we use these new-found friends? We're going to think about some functions (and combinations of functions) that are new to us and that aren't so clear for us to use things like the Product, Quotient, or Chain Rule. We'll try to use logs to re-write our functions into some easier to approach implicitly defined relationships in order for us to differentiate.

But first, let's build an explanation for the Power Rule for Derivatives.

Activity 3.3.1 Returning to the Power Rule.

Back in Section 2.3 we built an explanation for why $\frac{d}{dx}(x^n) = nx^{n-1}$ that relied on thinking about exponents as repeated multiplication: it relied on n being some positive integer. We said, at the time, that the Power Rule generalizes and works for *any* integer, but did so without explanation.

Let's consider $y = x^n$ where n is just some real number without any other restrictions.

(a) Apply a logarithm to both sides of this equation:

$$\ln(y) = \ln(x^n)$$

Now use one of the Properties of Logarithms to re-write this equation.

(b) Use implicit differentiation to find $\frac{dy}{dx}$ or y'.

(c) Explain to yourself why this is equivalent to the Power Rule that we built so long ago:

$$\frac{d}{dx}(x^n) = nx^{n-1}.$$

(d) Let's get weird. What if we have a not-quite-power function? Where the thing in the exponent isn't simply a number, but another variable?

Let's use the same technique to think about $y = x^x$ and its derivative. First, though, confirm that this is not a power function (and so we cannot use the Power Rule to find the derivative) and is also not an exponential function (and so the derivative isn't itself or itself scaled by a log).

(e) Now apply a log to both sides:

$$\ln(y) = \ln(x^x).$$

Re-write this using the same log property as before, and then use implicit differentiation to find $\frac{dy}{dx}$ or y'.

(f) Explain to yourself why logs are friends, especially when trying to differentiate functions in the form of $y = (f(x))^{g(x)}$.

This idea that we've just implemented (applying a logarithm to make some function more friendly and then using implicit differentiation to differentiate) is often called **logarithmic differentiation**. It works so well because *logs are friends*.

Wow, So Friendly!

This is incredible! We can now differentiate a whole new class of functions! Functions raised to functions, what could we think of next!

How about products and quotients of functions? I know, I know, we have The Product and Quotient Rules...what about *big* products and quotients? Annoying ones. Complicated ones.

Activity 3.3.2 Logarithmic Differentiation with Products and Quotients.

Let's say we've got some function that has products and quotients in it. Like, a lot. Consider the function:

$$y = \frac{(x-4)^2\sqrt{3x+1}}{(x+1)^7(x+5)^3}.$$

(a) Work out a general strategy for how you would find y' directly. Where would you have to use Quotient Rule? What are the pieces? Where would you have to use Product Rule? What are the pieces? Where would you have to use Chain Rule? What are the pieces?

To be clear: do not actually differentiate this. Just look at it in horror and try to outline a plan that some other fool would use.

Click on the "Solution" below to see what the fool did.

(b) Let's instead use logarithmic differentiation. First, apply the log to both sides to get:

$$\ln(y) = \ln\left(\frac{(x-4)^2\sqrt{3x+1}}{(x+1)^7(x+5)^3}\right).$$

Since this function is just a bunch of products of things with exponents all put into some big quotient, we can use our log properties to re-write this!

(c) We should have:

$$\ln(y) = 3\ln(x-4) + \frac{1}{2}\ln(3x+1) - 7\ln(x+1) - 3\ln(x+5).$$

Confirm this.

(d) Now differentiate both sides! You'll have to use some Chain Rule (but not a lot)! Refer back to Fact 3.3.1 for help here.

(e) Solve for $\frac{dy}{dx}$ or y'.

(f) While this is not a *nice* looking expression for the derivative, spend some time confirming that this was a nicer *process* then differentiating directly. This is because logs are friends.

So how do we wrap this up? I hope we can see that logs are a useful and powerful tool: we can use logarithmic differentiation to transform our functions to become "easier to work with" versions of themselves: we put everything on a log-scale and allow our properties of logarithms to make the operations become a bit more accessible.

This is a commonly used trick in many applications of calculus. Specifically, this is used often enough in statistics that there is a whole paradigm of estimation (called Maximum Likelihood Estimation) that uses a log-transformed version of a likelihood function and then applies some basic calculus ideas (that we'll cover in Section 4.5) to perform some powerful estimations.

While I hope that we end up leaving this section knowing that *logs are friends*, we can probably add a second (and related result) that we're using over and over.

Using the Chain Rule is probably easier than any other option.

We apply logs in order to re-write these functions in a friendly way *because* we would rather use the Chain Rule than any combination of other derivative strategies.

Chapter 4

Applications of Derivatives

4.1 Mean Value Theorem

Let's begin here with some weird questions. The questions aren't weird because of what they're asking. Instead, they're weird because of the logic of how we interpret them compared to how we *want* to interpret them.

1. Why is the derivative of a constant function 0?
2. Why do $y = x^2 + 7$ and $y = x^2 - 3$ have the same derivative?
3. If a function is only increasing on the interval $(0, 1)$, what do we know about the derivative at any of these x-values in $(0, 1)$?

These questions are ones we can think through and answer.

> Here are some answers for these first three questions:
>
> 1. A constant function has all of the same y-values for any x-value in the domain: of course the slope everywhere is 0! There isn't any change in the y-values!
> 2. We can differentiate these functions term by term: we know that the x^2 term has a derivative of $2x$, and then for each function's constant, the derivative has to be 0. So it doesn't matter what each constant was, the derivatives wouldn't rely on that value!
> 3. If a function is increasing on an interval, then it means that for any pair of x-values, $x_1 < x_2$, we get y-values in the same order: $f(x_1) < f(x_2)$. If we find the limit of slopes as $x_1 \to x_2$, we'll always get a positive slope, so the derivative has to be positive.

What's tricky is that these don't always say what we *want* to say. For instance, I might sometimes want to say the following:

1. If you know a function's derivative is 0, then you know the function is constant. Another way of saying this is that the *only* functions whose derivative is 0 are constant functions.
2. Similarly, we might want to say that if you know two functions that have the same derivative, then the only difference between the functions is a constant. There aren't any other ways for functions to be different with the same derivative.
3. We might want to say that if you know the derivative is positive on an interval, that means that the function has to be increasing.

Do you see the (slight) difference?

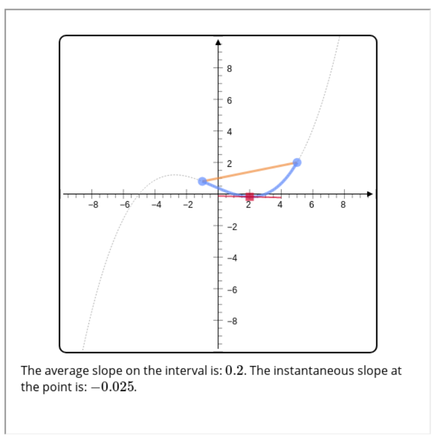
What we'll secretly see is that all of these statements are actually correct but require a result for us to say them. The Mean Value Theorem will be the result that we use to build important and helpful results throughout the rest of this course.

Slopes

We have two different kinds of slopes that we think of right now: a slope between two points, and a slope at a single point.

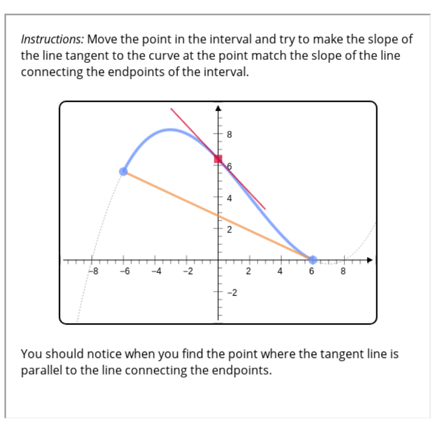
We can translate this into a rate of change! We think of two rates of change: an average rate of change on some interval and the instantaneous rate of change at some point.

We will try to connect these two different slopes/rates of change for "well-behaved" functions. Let's take a look at an example. In the graph below, we have a curve where we are considering some closed interval, as well as a point within that interval. Both slopes are visualized and calculated: the slope between the ending points of the interval is the average slope, while the slope of the line tangent to the curve at a point is the instantaneous slope. Move the point/interval around and get a feel for how these two different slopes relate (or don't relate!) to each other.



Standalone Embed

If you move the interval/point around enough, you'll see that sometimes these two slopes are really similar and sometimes they're very different. But what if the point in the middle of the interval wasn't so "set" at being stuck in the exact middle of the interval? What if we stuck the interval in place, and then had the freedom to move the point anywhere in the interval?



Standalone Embed

Do you think we will *always* be able to do this? What kinds of functions will have these points where the slope at the point matches the average slope on the interval?

The Mean Value Theorem

Theorem 4.1.1 Mean Value Theorem.

For a function $f(x)$ that is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), then there is some value $x = c$ with $a < c < b$ where:

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

This theorem is really just guaranteeing the existence of at least one of the points we found above: a point where the slope of the line tangent to the curve is the same as the slope between the endpoints of the interval. We can (and, very briefly, we will) use this theorem to find the point that is guaranteed to exist, but we will, more generally, use this theorem as a tool for connection.

We will try to use it as a way of connecting derivatives to the behavior of a function. The Mean Value Theorem gives us this equation where on one side we have a derivative, and on the other side we have function outputs. This is really similar to the definition of the Derivative at a Point, except that in this case we have no limit: we just get to use the behavior of the function on an interval around the point.

Secretly, the Mean Value Theorem is the driving force behind most of the results we will look at from here on out, at least when the requirements include continuity *and* differentiability on an interval. You can almost guarantee that if a theorem or result has these two requirements, then the Mean Value Theorem is likely at work.

Let's look at one example, at least, before we move on.

Example 4.1.2

Let's say that a person is planning on biking to their college campus, 8 miles away. At 2:39pm they send a text to their friend with a selfie of them on their bike about to start their trip, captioned "Beautiful day for a ride!" At 3:27pm, they post a picture on their social media of them in front of the bike rack on campus.

(a) What was the average velocity of the student between 2:39pm and 3:27pm?

Solution. The total trip took 48 minutes (or 0.8 hours), and the student traveled a total distance of 8 miles.

The student's average velocity is $\frac{8}{0.8} = 10$ miles per hour.

(b) The Mean Value Theorem connects average slopes with slopes of tangent lines. What does that mean, in general, in this context?

Solution. In this case, the average rate of change of the function on the interval is the average velocity of the biker. So then the instantaneous rate of change must correspond with an instantaneous velocity, of their velocity at some specific point in time.

(c) Make a claim about the instantaneous velocity of the student on their bike at some point in time.

Solution. We know that at some point between 2:39pm and 3:27pm, the cyclist had to be traveling at exactly 10 miles per hour.

Example 4.1.3

(a) For a function $f(x) = \sqrt{x} + 1$ on the interval $[1, 4]$, find the point that the Mean Value Theorem guarantees the existence of, and explain what it is.

Solution. Let's calculate the average slope on the interval:

$$\begin{aligned} \frac{f(4) - f(1)}{4 - 1} &= \frac{(\sqrt{4} + 1) - (\sqrt{1} + 1)}{3} \\ &= \frac{1}{3} \end{aligned}$$

So we know that there is some x-value between 1 and 4 where $f'(x) = \frac{1}{3}$.

The derivative is $f'(x) = \frac{1}{2\sqrt{x}}$, so we can solve for x:

$$\begin{aligned} f'(x) &= \frac{1}{3} \\ \frac{1}{2\sqrt{x}} &= \frac{1}{3} \\ 2\sqrt{x} &= 3 \\ \sqrt{x} &= \frac{3}{2} \\ x &= \frac{9}{4} \end{aligned}$$

So the point guaranteed to exist by the Mean Value Theorem is located at $\left(\frac{9}{4}, \frac{5}{2}\right)$, where the slope of the tangent line is $f'\left(\frac{9}{4}\right) = \frac{1}{3}$.

These examples are fine, but the real power of the Mean Value Theorem comes in how we can use it to general more interesting results. Let's check those out!

More Results due to the Mean Value Theorem

First, let's remind ourselves of what it means for a function to be increasing or decreasing on an interval.

Definition 4.1.4 Increasing and Decreasing on an Interval.

A function $f(x)$ is **increasing** on the interval (a, b) if, for every pair of x-values x_1 and x_2 (with $x_1 < x_2$), $f(x_1) < f(x_2)$.
If $f(x_1) > f(x_2)$, then we say that f is **decreasing** on the interval (a, b).

Note that we classify a function as increasing or decreasing based on comparing two function outputs. This is a perfect time to use the Mean Value Theorem, since it can connect these pairs of function outputs with a derivative!

Theorem 4.1.5 Sign of the Derivative and Direction of a Function.

If f is a function that is differentiable on the interval (a, b) and $f'(x) > 0$ for all x in the interval (a, b), then f must be increasing on (a, b). Similarly, if $f'(x) < 0$ for all x in the interval (a, b), then f must be decreasing on (a, b).

Proof.

Before we begin, let's just agree to think about only the case where $f'(x) > 0$ on the interval (a, b). The other case (where $f'(x) < 0$) ends up being the exact same argument, with some changes in signs and directions of inequalities. So we'll say $f'(x) > 0$ for all x-values in the interval (a, b).

Ok, let's begin!

Let's pick two arbitrary x-values from the interval (a, b). Call them x_1 and x_2, and we'll make sure that we name them in a way where $x_1 < x_2$. Now, f must be continuous on $[x_1, x_2]$ and differentiable on (x_1, x_2). We also know that $f'(x) > 0$ for every x in the interval $(x_1 < x < x_2)$.

The Mean Value Theorem says that there is some $x = c$ in (x_1, x_2) with

$$f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

Equivalently, this means that

$$f'(c)(x_2 - x_1) = f(x_2) - f(x_1).$$

Notice that $f'(c) > 0$ (since the derivative is positive everywhere in the interval) and $(x_2 - x_1) gt0$ (by the wat we named these x-values). This means that $f'(c)(x_2 - x_1) > 0$, and so $(f(x_2) - f(x_1)) > 0$ as well.

Ok so notice what just happened: we arbitrarily chose x-values x_1 and x_2 and noticed that for any of these pairs where $x_1 < x_2$, we found that $f(x_1) < f(x_2)$. This is exactly what it means for f to be increasing on the interval (a, b) (based on Definition 4.1.4).

We'll use this result pretty extensively in the next couple of topics. It is helpful for us to use this to connect the signs of a derivative with our intuition about what that must mean for the direction of a function.

Theorem 4.1.6 If a Function's Derivative is 0, it's Constant.

If $f(x)$ is a function defined on (a, b) with $f'(x) = 0$ for all x-values in the interval (a, b), then $f(x)$ is a constant function.

Proof.

We can almost exactly replicate the proof from Theorem 4.1.5 here!

Let's pick two arbitrary x-values from the interval (a, b). Call them x_1 and x_2, and we'll again make sure that we name them in a way where $x_1 < x_2$. Now, f must be continuous on $[x_1, x_2]$ and differentiable on (x_1, x_2). This time, we know that $f'(x) = 0$ for every x in the interval $(x_1 < x < x_2)$.

The Mean Value Theorem says that there is some $x = c$ in (x_1, x_2) with

$$f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

Equivalently, this means that

$$f'(c)(x_2 - x_1) = f(x_2) - f(x_1).$$

Notice that $f'(c) = 0$ (since the derivative is zero everywhere in the interval). This means that $f'(c)(x_2 - x_1) = 0$, and so $(f(x_2) - f(x_1)) = 0$ as well.

This means that every y-value from the function is the same as every other one, since we picked these points arbitrarily. So f must be a constant function!

Theorem 4.1.7 Equal Derivatives Correspond with Related Functions.

For two functions f and g, both differentiable on (a, b), if $f'(x) = g'(x)$ for all x-values on (a, b), then we know that $f(x) = g(x) + C$ for some real number constant C. That is, the only differences in f and g are due to a difference in the constant term.

Proof.

This one is pretty easy to prove: we're going to use a fun little trick where we connect this theorem to Theorem 4.1.6.

Let's think about these two functions f and g with $f'(x) = g'(x)$, and let's think about a function $h(x) = f(x) - g(x)$. Now we can notice that $h'(x) = f'(x) - g'(x)$ which has to be 0.

Oh wow, $h(x)$ is a function with $h'(x) = 0$ on the interval (a, b). It must be a constant function (based on Theorem 4.1.6)! Let's call it $h(x) = C$, where C is some real number constant.

This means that $f(x) - g(x) = C$, and we can see that the only difference between these two functions is a constant.

We'll use this result a bit later on, but it's a great example of how we can use the Mean Value Theorem to connect information about the derivative to information about how a function might work.

Let me interject my own opinion here: I think the Mean Value Theorem is extremely important. But I also don't think that it is that important for you to *use*.

I think you should know the general idea of connecting the slopes of the line tangent to the curve and the average slope on an interval. I think you should remember the picture we looked at. I think you should be content to see some results later on in the course and smile fondly about how the Mean Value Theorem is under the surface, churning away and getting us cool results to think about. I think we should be happy that the Mean Value Theorem is here and we can recognize it as, maybe, the most important result in this textbook.

But we don't need to pretend that we need to actually *use* it directly...we aren't going to need to compute a lot or anything like that. We'll just *know* it.

4.2 Increasing and Decreasing Functions

Activity 4.2.1 How Should We Think About Direction?

Our goal in this activity is to motivate some new terminology and results that will help us talk about the "direction" of a function and some interesting points on a function (related to the direction of a function). For us to do this, we'll look at some different examples of functions and try to think about some unifying ideas.

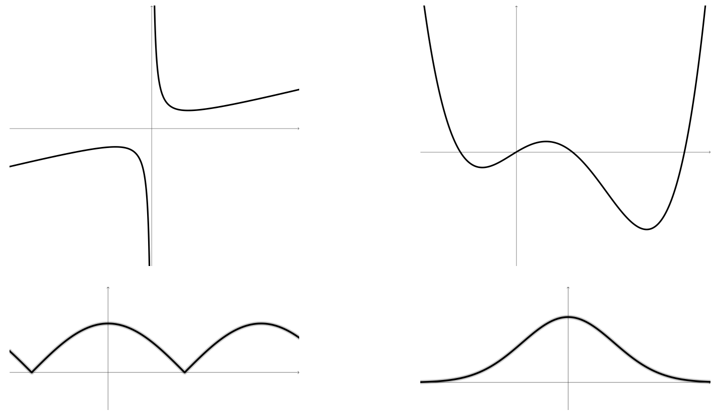


Figure 4.2.1

These examples do not cover all of the possibilities of how a function can act, but will hopefully provide us enough fertile ground to think about some different situations.

(a) In each graph, find and identify:

- The intervals where the function is increasing.
- The intervals where the function is decreasing.
- The points (or locations) around and between these intervals, the points where the function changes direction or the direction terminates.

(b) Make a conjecture about the behavior of a function at any point where the function changes direction.

(c) Look at the highest and lowest points on each function. You can even include the points that are highest and lowest just compared to the points around it. Make a conjecture about the behavior of the function at these points.

We want to turn this little bit of thinking and exploring into some useful definitions for us. To craft these definitions, we need to start with thinking about what we care about and why we might care about it.

Critical Points, Local Maximums, and Local Minimums

Let's start by saying what we're really looking for is the highest and lowest points on a function. These points are interesting, have useful applications, and are difficult to find in general without calculus. We hopefully noticed, though, that these points always end up showing up at the same kinds of locations! They're points where the direction of a function changes (or terminates).

We also noticed that there are some common characteristics of those points. They're points where the derivative was either 0 or didn't exist. So we'll start by defining these points, and then we'll define the idea of "highest" and "lowest" points. Then we'll put together a result that we hopefully noticed here: that the highest and lowest points show up at these points where the derivative is 0 or doesn't exist.

Definition 4.2.2 Critical Point of a Function.

We say that a point $(c, f(c))$ on the graph of $y = f(x)$ is a **critical point** of the function f if $f'(c) = 0$ or $f'(c)$ doesn't exist.
If $(c, f(c))$ is a critical point of f, then we sometimes will call $x = c$ a **critical number** and $y = f(c)$ a **critical value**.

So these are the points we will look for to find the ”highest” and ”lowest” points.

Now we need to define this idea so that we can build the result that tells us how to find these highest and lowest points.

Definition 4.2.3 Local Maximum/Minimum.

A point $(c, f(c))$ is a **local maximum** of $f(x)$ if there is some open interval of real numbers (a, b) around $x = c$ (that is, $a < c < b$) and $f(c) \geq f(x)$ for all x-values in the intersection of (a, b) and the domain of f.
Similarly, a point $(c, f(c))$ is a **local minimum** of $f(x)$ if there is some open interval of real numbers (a, b) around $x = c$ (that is, $a < c < b$) and $f(c) \leq f(x)$ for all x-values in the intersection of (a, b) and the domain of f.
These are really just slightly technical ways of saying that $f(c)$ is either the highest or lowest y-value produced by the function f for the x-values near $x = c$.

> If you look around online, or in other textbooks, you'll find different definitions of these same kinds of pints. Some of those definitions have silly stipulations, like saying that an ending point of a function cannot be a local maximum/local minimum.
>
> This seems to come from some sense that the derivative must be defined on each ”side” of a local max/min. In this book, we'll not worry about this restriction, and instead just look at the highest and lowest points relative to the other points near it.

Now we want to build the connection between these points. In Activity 4.2.1, we pointed out that the highest and lowest points on a function all had the common theme of showing up at places where the derivative was 0 or didn't exist.

Theorem 4.2.4 Every Local Maximum/Minimum Occurs at a Critical Point.

Every local maximum or local minimum of f occurs at a critical point of f.
Another way of saying this is that if $(c, f(c))$ is a local maximum or local minimum of f, then it must also be a critical point of f.

WAIT! STOP! Before you move on, let's make sure we understand what this theorem is saying. Or maybe what this theorem is *not* saying.

Notice that we are not saying that every critical point is a local maximum or local minimum! This is a classic ”every square is a rectangle but not every rectangle is a square” situation.

Every local maximum/minimum occurs at a critical point, but not every critical point is a local maximum/minimum.

Direction of a Function (and Where it Changes)

Let's build up a way of classifying critical points as local maximums, local minimums, or neither.

Activity 4.2.2 Comparing Critical Points.

Let's think about four different functions:

- $f(x) = 4 + 3x - x^2$
- $g(x) = \sqrt[3]{x+1} + 1 + x$
- $h(x) = (x-4)^{2/3}$
- $j(x) = 1 - x^3 - x^5$

Our goal is to find the critical points on the interval $(-\infty, \infty)$ and then to try to figure out if these critical points are local maximums or local minimums or just points that the function increases or decreases through.

(a) To start, we're going to be finding critical points. Without looking at a picture of the graph of the function, find the derivative.

Are there any x-values (in the domain of the function) where the derivative doesn't exist? We are normally looking for things like division by 0 here, but we could be finding more than that. Check out When Does a Derivative Not Exist? to remind yourself if needed.

Are there any x-values (in the domain of the function) where the derivative is 0?

(b) Now that we have the critical points for the function, let's think about where the derivative might be positive and negative. These will correspond to the direction of a function, based on Theorem 4.1.5 Sign of the Derivative and Direction of a Function.

Write out the intervals of x-values around and between your list of critical points. For each interval, what is the sign of the derivative? What do these signs mean about the direction of your function?

(c) Without looking at the graph of your function:

- What changes about how your function increases up to or decreases down to a critical point based on whether the derivative was 0 or the derivative didn't exist?
- Does your function change direction at a critical point? What will that look like, whether it does or does not change direction?

(d) Give a basic sketch of your graph. It might be helpful to find the y-values for any critical points you've got. Then you can sketch your function increasing/decreasing in the intervals between these points.

In your sketch, include enough detail to tell whether the derivative is 0 or doesn't exist at each critical point.

(e) Compare your sketch to the actual graph of the function (you can find all of the graphs in the hint).

This is great, we have a nice strategy for thinking about critical points!

Something we can notice: in finding these critical points (as well as thinking about the domain of the function), we found *all* of the locations where the derivative is both not positive and not negative. This is a weird way of saying that all of the intervals in between the critical points we found and any breaks in the domain of the function (like if there were vertical asymptotes or holes or something) are places where the derivative is positive or negative.

Even more exciting: if the derivative function we found is continuous, then the Intermediate Value Theorem says that it will *only* change signs at these critical points (or places like vertical asymptotes or holes). So this means that we can always construct a little chart or something, think about the x-values around and at critical points or other breaks in the domain, and then look at what the derivative does as we move through those intervals and x-values.

This will serve as a nice way of thinking about what's going on with our functions!

Theorem 4.2.5 First Derivative Test.

If $(c, f(c))$ is a critical point of f and we can evaluate the derivative f' on either side of this point, then we can use the signs of the first derivative to classify the critical point:

- *If the sign of f' changes from positive to negative as x passes through $x = c$, then $(c, f(c))$ is a local maximum.*
- *If the sign of f' changes from negative to positive as x passes through $x = c$, then $(c, f(c))$ is a local minimum.*
- *If the sign of f' does not change as x passes through $x = c$, then the function f increases or decreases (depending on whether $f' > 0$ or $f' < 0$) through $x = c$.*

We will often lay these results out in a chart or table, like the following:

x		c	
f'	$\oplus$		$\ominus$
f	$\nearrow$		$\searrow$
		local max	

x		c	
f'	$\ominus$		$\oplus$
f	$\searrow$		$\nearrow$
		local min	

x		c	
f'	$\oplus$		$\oplus$
f	$\nearrow$		$\nearrow$
		increasing through	

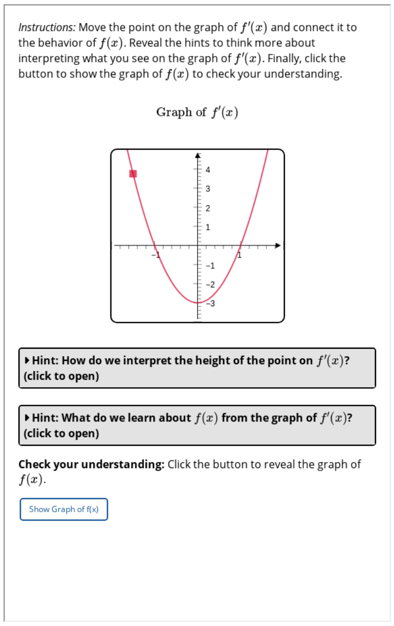
x		c	
f'	$\ominus$		$\ominus$
f	$\searrow$		$\searrow$
		decreasing through	

Using the Graph of the First Derivative

Activity 4.2.3 First Derivative Test Graphically.

Let's focus on looking at a picture of a derivative, $f'(x)$, and trying to collect information about the function $f(x)$. This is what we've done already, except that we've done it by thinking about the representation of $f'(x)$ as a function rule written out with algebraic symbols. Here we'll focus on connecting all of that to the picture of the graphs.

For all of the following questions, refer to the plot below. You can add information with the hints whenever you need to. Don't reveal the picture of $f(x)$ until you're really ready to check what you know.



Standalone Embed

(a) Based on the graph of $f'(x)$, estimate the interval(s) of x-values where $f(x)$ is increasing.

(b) Based on the graph of $f'(x)$, estimate the interval(s) of x-values where $f(x)$ is decreasing.

(c) Find the x-values of the critical points of $f(x)$. Once you've estimated these, classify them as local maximums, local minimums, or neither. Explain your reasoning.

(d) What do you think the graph of $f(x)$ looks like? Do your best to sketch it or explain it before revealing it!

(e) Why could we estimate the x-values of the critical numbers of $f(x)$, but not find the actual coordinates? How come we can't find the y-value based on looking at the graph of $f'(x)$?

Reading the graphs of functions is, in general, an important skill. But it's an especially important idea to be able to read and understand the graph of a function like a derivative and then interpret what we are seeing into some other context.

So for us to really excel here, we'll want to focus on the fact that a first derivative tells about the *slope* or *direction* of a function. Whatever y-values we find on the graph of a $f'(x)$ needs to be interpreted as a slope or rate of change of $f(x)$. Then we can string these slopes or rates of changes together to try to think about the behavior of $f(x)$ by knowing how the y-values are changing as we move along the curve of $y = f(x)$!

Strange Domains

We'll look at two more examples, both of them using functions whose domain is *not* $(-\infty, \infty)$.

Example 4.2.6

(a) Consider the function

$$f(x) = \frac{x^2}{(x-3)^2}.$$

Find the domain of f, the critical points of f, and then the intervals where f is increasing/decreasing. Then, classify any critical points local maximums/minimums if necessary.

Hint 1. $f'(x) = -\dfrac{6x}{(x-3)^3}$

Hint 2. The function $f(x)$ has one critical point at $(0, 0)$. Why isn't there a critical point at $x = 3$? What is happening there instead?

Hint 3.

x	$(-\infty, 0)$	0	$(0, 3)$	3	$(3, \infty)$
f'					
f					

Answer. The domain of $f(x)$ is $(-\infty, 3) \cup (3, \infty)$ due to the vertical asymptote at $x = 3$. The only critical point is at $(0, 0)$. The table below shows where f is increasing and decreasing, as well as any local maximums or minimums.

x	$(-\infty, 0)$	0	$(0, 3)$	3	$(3, \infty)$
f'	$\ominus$	0	$\oplus$		$\ominus$
f	$\searrow$ decreasing	$(0, 0)$ local min	$\nearrow$ increasing	asymptote	$\searrow$ decreasing

(b) Consider the function

$$g(x) = \sqrt{x} - x + 1.$$

Find the domain of g, the critical points of g, and then the intervals where g is increasing/decreasing. Then, classify any critical points local maximums/minimums if necessary.

Hint 1. $g'(x) = \dfrac{1}{2\sqrt{x}} - 1$

Hint 2. Notice that, by our definition of critical points, both $(0, 1)$ and $\left(\frac{1}{4}, \frac{3}{4}\right)$ are critical points.

Hint 3.

x	0	$\left(0, \frac{1}{4}\right)$	$\frac{1}{4}$	$\left(\frac{1}{4}, \infty\right)$
g'				
g				

Answer. The domain of $g(x)$ is $[0, \infty)$. There are two critical points: one at $(0, 1)$ and another at $\left(\frac{1}{4}, \frac{3}{4}\right)$. The table below shows where g is increasing and decreasing, as well as any local maximums or minimums.

x	0	$\left(0, \frac{1}{4}\right)$	$\frac{1}{4}$	$\left(\frac{1}{4}, \infty\right)$
g'	DNE	$\ominus$	0	$\oplus$
g	$(0, 1)$	$\searrow$	$\left(\frac{1}{4}, \frac{3}{4}\right)$	$\nearrow$
	local max	decreasing	local min	increasing

So we have two things to notice:

1. When we have some gap or missing spot in the domain of a function, that can still divide up the intervals where our function is increasing or decreasing! We should notice, though, that since this isn't actually a *point* on the curve of our function, it won't be a critical point and so we have to interpret it differently: we can't use the First Derivative Test!

2. An ending point of an interval is a location where the derivative cannot exist! We could define a *one-sided derivative* (similar to how we defined one-sided continuity in Definition 1.6.1), but for now we'll just say that the derivative doesn't exist, and call those ending points critical points. That means that depending on the direction that a function goes away (or leading up to) that ending point, we can classify it as a local maximum or minimum.

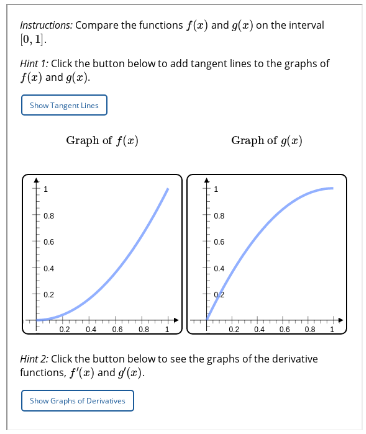
Lastly, just a couple of notes: in these little tables or charts (sometimes called **sign charts**, since they are showing the signs of the derivative), we'll use some shorthand notation. In Example 4.2.7, we used "DNE" to mean that a derivative "does not exist" at a point. Similarly, we used to represent the vertical asymptote at that x-value (so that we didn't accidentally think it was a local maximum or minimum based on the signs of the derivative around it).

Moving forward, we'll use this same kind of analysis to think about how the derivative might be changing on these intervals. This rate of change of the slopes, the **second derivative**, will be a useful tool for gathering more information about how a function might be acting.

4.3 Concavity

Activity 4.3.1 Compare These Curves.

(a) Consider the two curves pictured below. Compare and contrast them. What characteristics of these functions are the same? What characteristics of these functions are different?



Standalone Embed

(b) Explain the similarities you found by only talking about the slopes of each function (the values of $f'(x)$ and $g'(x)$).

(c) Explain the differences you found by only talking about the slopes of each function (the values of $f'(x)$ and $g'(x)$).

(d) Make a conjecture about the *rate of change* of both f' and g'. We'll call these things **second derivative** functions, $f''(x)$ and $g''(x)$.

Defining the Curvature of a Curve

Definition 4.3.1 Concavity and Inflection Points.

We say that a function if **concave up** on an interval (a,b) if $f'(x)$ is increasing on the interval. If $f'(x)$ is decreasing on the interval, then we say that $f(x)$ is **concave down**.
We say that a point $(c, f(c))$ is an **inflection point** if it is a point at which f changes concavity (from concave up to concave down or vice versa).

Note that when we think about a function being concave up or down on some interval, we can think about this in different ways. Curvature can sometimes be hard to recognize visually, but one of the things we can see from the visual above is the interaction between the tangent line and the curve:

- If the function is concave up on some interval, then the tangent line sits below the function at every point on that interval. The function curves *upward* away from the tangent line. Sometimes people will say that the curvature is concave "up, like a cup."

- If the function is concave down on some interval, then the tangent line sits above the function at every point on that interval. The function curves *downward* away from the tangent line. Sometimes people will say that the curvature is concave "down, like a frown."

So we have some visual ways of thinking about these different types of curvature. Annoyingly, though, it is still relatively difficult to pinpoint the exact (or even close) location of an inflection point. We might be able to pretty easily point at the locations of local maximums and local minimums on a graph of a function, but it can be hard to see the exact point at which a graph of a function changes from concave up to down or vice versa. We'll focus on finding them algebraically first, but then we'll think a bit more about the graphs of functions later.

Activity 4.3.2 Finding a Function's Concavity.

We're going to consider a pretty complicated function to work with, and employ a strategy similar to what we did with the first derivative:

- Our goal is to find the sign of $f''(x)$ on different intervals and where that sign *changes*.
- To do' this, we'll find the places that $f''(x) = 0$ or where $f''(x)$ doesn't exist. These are the critical points of $f'(x)$.
- From there, we can build a little sign chart, where we split up the x-values based on the domain of f and the critical numbers of f'. Then we can attach each interval of x-values with a sign of the second derivative, f'', on that interval.
- We'll interpret these signs. For any intervals where $f''(x) > 0$, we know that $f'(x)$ must be increasing and so $f(x)$ is concave up. Similarly, for any intervals where $f''(x) < 0$, we know that $f'(x)$ must be decreasing and so $f(x)$ is concave down.

Let's consider the function

$$f(x) = \ln(x^2+1) - \frac{x^2}{2}.$$

(a) Find the first derivative, $f'(x)$, and use some algebra to confirm that it is really:

$$f'(x) = -\frac{x(x+1)(x-1)}{x^2+1}.$$

While we have this first derivative, we *could* find the critical points of $f(x)$ and then classify those critical points using the First Derivative Test.

For our goal of finding the intervals where $f(x)$ is concave up and concave down, and then finding the inflection points of f, let's move on.

(b) Find the second derivative, $f''(x)$, and confirm that this is really:

$$f''(x) = -\frac{x^4+4x^2-1}{(x^2+1)^2}.$$

(c) Our goal is to find the x-values where $f''(x) = 0$ or where $f''(x)$ doesn't exist.

Note that in order to find where $f''(x) = 0$, we are really looking at the x-values that make the numerator 0:

$$x^4 + 4x^2 - 1 = 0.$$

Then, to find where $f''(x)$ doesn't exist, we are finding the x-values that make the denominator 0:

$$(x^2 + 1)^2 = 0.$$

Solve these equations.

(d) You have two critical points of $f'(x)$ (let's just call them x_1 and x_2). These are possible inflection points of $f(x)$, but we need to check to see if the concavity changes at these points.

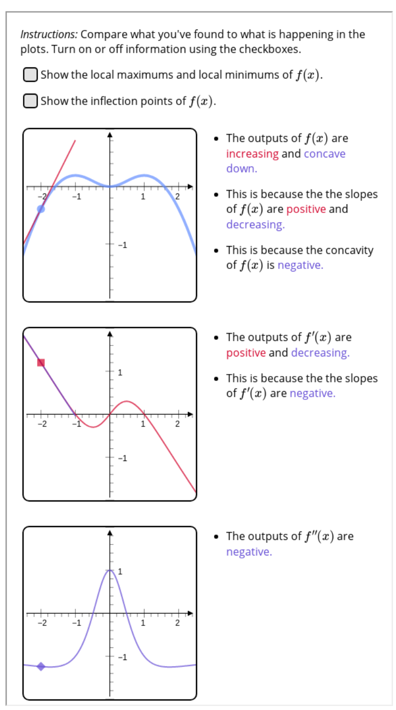
Fill in the following sign chart and interpret.

x	$(-\infty, x_1)$	x_1	(x_1, x_2)	x_2	(x_2, ∞)
f''					
f					

Let's confirm what we've just calculated graphically. Below, we have three graphs:

1. $f(x) = \ln(x^2 + 1) - \dfrac{x^2}{2}$
2. $f'(x) = -\dfrac{x(x+1)(x-1)}{x^2 + 1}$
3. $f''(x) = -\dfrac{x^4 + 4x^2 - 1}{(x^2 + 1)^2}$

Move the point on any graph and make sure the statements about signs, directions, and concavity match what you found! You should notice that signs of the first and second derivative change at the local maximums/ minimums and the inflection points that we found.



Standalone Embed

Let's circle back to the definition of Concavity and Inflection Points and think about these from the perspective of $f'(x)$.

We can notice that, by the definition, any inflection point is a point at which $f(x)$ changes concavity, and so is a point where $f'(x)$ changes direction. That means we are looking at local maximums or local minimums of $f'(x)$ (as long as they're not at the end points of some domain)! Similarly, these are points at which $f''(x)$ changes sign, and so we are thinking about the x-intercepts of these second derivatives (or other kinds of locations where the second derivative could change signs).

Example 4.3.2

Let's look at a few more examples, but this time they can be ones with derivatives that are a bit easier to work with.

(a) Consider the function $f(x) = \ln(x^2+1)$. Find the intervals where f is concave up, the intervals where it is concave down, and then find the locations of any inflection points.

(b) Consider the function $g(x) = e^{-x^2}$. Find the intervals where g is concave up, the intervals where it is concave down, and then find the locations of any inflection points.

At this point, we have three different functions that we are juggling: a function $f(x)$ (or whatever name we give it), the first derivative $f'(x)$, and the second derivative $f''(x)$. We'll want to keep in mind the role of each of these.

- $f(x)$ tells us the height, the y-value, of the function at some point.
- $f'(x)$ tells us the direction, the slope, of the function at some point.
- $f''(x)$ tells us the curvature, the concavity, of the function at some point.

We can try to summarize this in a small table:

$f(x)$	$\nearrow$	$\searrow$	$\smile$	$\frown$		
$f'(x)$	$\oplus$	$\ominus$	$\nearrow$	$\searrow$	$\smile$	$\frown$
$f''(x)$			$\oplus$	$\ominus$	$\nearrow$	$\searrow$
$f'''(x)$					$\oplus$	$\ominus$

Notice that we could extend this table and think about any triplet of functions/ derivatives in a row: we just need to think about what is positive/negative, what is increasing/decreasing, and what is concave up/down. If we wanted, we could try to define some adjective to describe what is happening to a function when $f'(x)$ is concave up/down, but let's not. It's hard enough to visualize concavity, and it will be difficult to visualize rates of change of the concavity.

Interpreting the Concavity at Critical Points

Theorem 4.3.3 Second Derivative Test for Local Maximums or Local Minimums.

If $(c, f(c))$ is a critical point of f with $f'(c) = 0$, then we can use the value of the second derivative at $x = c$ to classify the critical point:

1. *If $f''(c) > 0$, then f is concave up at and around $x = c$, and so the function has a local minimum at $(c, f(c))$.*

2. *If $f''(c) < 0$, then f is concave down at and around $x = c$, and so the function has a local maximum at $(c, f(c))$.*

3. *If $f''(c) = 0$, then the Second Derivative Test is inconclusive.*

Example 4.3.4

Find any critical points of the following functions. For each, use the Second Derivative Test to classify the critical point. If the Second Derivative Test fails (we get that the second derivative evaluated at the critical point is 0, and so is inconclusive), then classify the critical point in some other way.

(a) $f(x) = x\ln(x) - x$ where the domain of f is $(0, \infty)$.

(b) $g(x) = 5x + \frac{1}{x}$ where the domain of g is $(0, \infty)$.

Let's leave this here, with a few questions for you to think about:

- When, for you, do you think it would be reasonable to use the Second Derivative Test instead of the First Derivative Test? What goes into making this decision?

- When, for you, do you think it would be reasonable to use the First Derivative Test instead of the Second Derivative Test? What does into making this decision?

4.4 Global Maximums and Minimums

We need to start with a definition, and we can start with contrasting what we want the difference between a *local* maximum/minimum and a *global* maximum/minimum. Sometimes these are also called *absolute* maximum/minimums.

What do you want the difference to be?

If we focus on the terms or the names we're giving, then the difference should be based on the distinction between the words "local" and "global." In one, we're considering some confined and relatively arbitrary geographic area, just the things around or in the neighborhood. In the other, our context grows until we're considering the whole picture, the whole space that we're interested in!

Definition 4.4.1 Global Maximum and Global Minimum.

A function has a **global maximum** value of $f(c)$ if $f(c) \geq f(x)$ for all x-values in the domain of f.
A function has a **global minimum** value of $f(c)$ if $f(c) \leq f(x)$ for all x-values in the domain of f.

Note that the difference between this definition and Definition 4.2.3 Local Maximum/Minimum is the types of y-values we're comparing $f(c)$ to: in this new definition, we just use all of the x-values in the domain. In the definition for a local max/min, we had to construct some interval to intersect with the domain in order to just consider the "local" picture.

Activity 4.4.1 When Would We Not Have Maximums or Minimnums?

In this section, we're going to define these global maximums and then, most importantly, try to predict when these global maximums or global minimums might actually exist for a function.
To start, maybe we should come up with some examples of functions that do not have them!

(a) Come up with some situations where a function does not have some combination of global maximum/minimums. Sketch some graphs!

(b) Come up with some examples of graphs of functions that are bounded (do not ever have y-values that tend towards infinity in a limit) that do not have some combination of global maximum/ minimums.

(c) For the examples of graphs that you have built or collected, features of the functions that allow for the examples you picked? If you could impose some requirements that would "fix" the examples you found (so that they had both a global maximum and a global minimum), what requirements could you use?

When Do We Guarantee Both a Global Maximum and a Global Minimum?

Theorem 4.4.2 Extreme Value Theorem.

If $f(x)$ is a continuous function on a closed interval $[a, b]$ then f must have both a global maximum and a global minimum on the interval.

The Extreme Value Theorem guarantees the existence of both the global maximums and minimums, but we actually get more than just this out of the Extreme Value Theorem. Once we *know* that both of the global maximums and minimums exist, we can find them pretty easily.

The global maximum of a function, if it exists for the function on the domain/interval, is just the local maximum with the largest y-value. Similarly, the global minimum, if it exists, is the local minimum with the lowest y-value.

So once we know they both exist for a function on an interval, we can simply collect the critical points of the function (including the ending points of the domain) and compare the y-value function outputs!

Example 4.4.3

Check to see if each function (on the stated domain) satisfies the conditions of the Extreme Value Theorem, and then find any global maximums/minimums of the function on the interval.

(a) $f(x) = \ln(x^2 + 4x + 7)$ on $[-1, 3]$

(b) $g(x) = 3x^4 - 5x$ on $[-3, 4]$

(c) $j(x) = \sqrt[3]{x + 4}$ on $[-6, -1]$

What about Domains of Functions that Aren't Closed?

Without the conditions that imply the Extreme Value Theorem, things become trickier. For instance, if the function is not continuous, then the function might have some unbounded behavior at a vertical asymptote. In this case, we might need to look at the one-sided limits around that asymptote, in order to see if our function tends towards positive or negative infinity on either side of the asymptote. This could tell us that the function doesn't have a global maximum, a global minimum, or that it doesn't have either.

Similarly, if the function is not defined on a closed interval, then we need to investigate what happens to the function's behavior as the function moves towards the "ends" of the interval (which could be a real number but something like positive or negative infinity). These end behavior limits could exist, in which case we need to compare these heights of horizontal asymptotes or open ends of an interval to the heights of any critical numbers.

But we might also find that the function tends towards infinity or negative infinity in the end behavior.

And there are other things to consider about discontinuity of a function (other than vertical asymptotes)!

All in all, it should be evident that if we remove one or both of the conditions on our function that *guarantees* the existence of a global maximum and a global minimum, it becomes much harder to find them, since we have so many different options to consider.

To simplify things, we will look at one case where things align in our favor: a continuous function that only has a single local maximum/minimum on an interval.

Theorem 4.4.4

If $f(x)$ is a continuous function on some interval, and f has only a single critical point (call it $(c, f(c))$) where the direction changes, then if that point is a local maximum of f, then $f(c)$ is the global maximum. Similarly, if $(c, f(c))$ is a local minimum, then $f(c)$ is the global minimum of f.

This is a great result to give us a path forward without having to check all of the edge cases and possibilities mentioned above. There are many functions that might have only a single critical point, or if it does have more than one critical point, only a single one of them acting as a local maximum/minimum.

Note here that we do need to classify the critical point as a local maximum or minimum! We'll use the First Derivative Test or the Second Derivative Test for Local Maximums or Local Minimums for this classification.

Example 4.4.5

For each function, find any global maximums/minimums that may exist.

(a) $f(x) = x\ln(x) - x$ and note that the domain of this function is $(0, \infty)$

(b) $g(x) = xe^{-x}$ and note that the domain of this function is $(-\infty, \infty)$.

4.5 Optimization

How do we use calculus to make decisions? If we're trying to find the best allocation of time between different tasks, or if we're trying to construct some object with limited resources, or if we're trying to find some other solution to the question, "How do we make the best/most/maximum/minimum of..." then we can think about translating these to a calculus context.

Optimization problems are some of the clearest application of calculus concepts to applied problems from any industry or field. While there certainly aren't teams of calculus experts in every office waiting for these kinds of problems to arrive, these problems are routinely solved using calculus, either done by computer software or coming from calculus experts who make sweeping recommendations across a whole field or some other source.

Optimization Framework

We'll start this discussion with a small example. This is *not* an example that shows up in "real life," but will be helpful to build a strategy for approaching these problems.

Here's our problem:

Example 4.5.1

Find two numbers that add to get 14 but give the largest possible product.
That's it. That's the problem. It's not super interesting, right? Let's think about how we'll solve it.

Solution 1. First, we want to think about our two numbers. We'll need variable names for these. You can use whatever letters you'd like here: I'm going to be boring and pick a and b as the names of my two numbers. What do we know about a and b?
The first thing that we know is that $a+b=14$. This isn't a huge amount of information, but it does provide us with a nice connection between our numbers. If we know what one number is, then we automatically know the second one: if one number is 4, then the other number is whatever we need to add to 4 to get 14.
We also know that we're interesting in the product: $P = ab$. We don't, currently, know anything about this product other than the fact that we want it to be large. Ok so automatically, we know that both a and b need to be positive: if we had one number that was positive (like 15) and one number that was negative (like -1), then the product will be negative, and even though they add to 14, we are *not* going to get a big product.
So we know that for both of these numbers $a, b \geq 0$.
This doesn't seem like a lot of information, but we can put it all together really nicely. For instance, I can manipulate the fact that these two numbers add to get 14 into an equation that tells me what the value of one number is based on the other:

$$a + b = 14 \longrightarrow b = 14 - a.$$

I can also write my product as an actual function of a single variable:

$$P = ab \longrightarrow P(a) = a(\underbrace{14 - a}_{b}).$$

Lastly, I actually know a domain for this function: we know that $a \geq 0$, but since $b \geq 0$, then $a \leq 14$. We *can't* have a number larger than 14, since the other number would be negative.

> I just said that we *can't* have a negative product, but nothing in this problem says that. We just know that a negative product will be small. But we don't *have* to limit each number to be between 0 and 14. We could just allow them to be bigger than 14 (and thus also allowed to be smaller than 0) and we'll just get a bunch of negative products that aren't the maximum one. Oh well!
>
> We are making a similar choice by *including* the pair 0 and 14 as a possibility. Do we *really* think that this could give us the biggest product? NO! The product is 0! But it's maybe convenient to have a closed interval as the domain of this function, so why not?
>
> All along the way, we're making choices that guide how we think about this problem. In real life, we'll do the same thing. The only difference is that the choices that we make about what are reasonable values for our inputs to take on are more meaningful, since those inputs represent real things. We have to take these choices seriously.

Ok, so we have a function representing the product, $P(a) = a(14 - a)$ and a domain for that function, $[0, 14]$, and we know we want to find the maximum for it. This is a calculus problem that we can actually do! Take a moment to do it.

Solution 2.

$$\begin{aligned} P(a) &= 14a - a^2 \\ P'(a) &= 14 - 2a \end{aligned}$$

Now we can find the critical points.

$$\begin{aligned} P'(a) &= 0 \\ 14 - 2a &= 0 \\ -2a &= -14 \\ a &= 7 \end{aligned}$$

We also know that the ending points at $a = 0$ and $a = 14$ are critical, since the derivative cannot exist.
There is a point on the product function when $a = 7$ where we have a horizontal tangent line. Does this represent a maximum or a minimum? We have some options for how to do this:

1. *FDT:*

a	$[0, 7)$	7	$(7, 14]$
P'	$\oplus$	0	$\ominus$
P	$\nearrow$ increasing	$(7, 49)$ local max	$\searrow$ decreasing

Since this is the only turning point, it must be the global maximum!

2. *SDT:*

$$P''(a) = -2$$

$$P''(7) < 0$$

Since the function is concave down at this point, then we know that the function reaches a local maximum when $a = 7$. Since this is the only turning point, it must be the global maximum!

3. *EVT:*

a	P	
0	0	global minimum
7	49	global maximum
14	0	global minimum

However we do this, we find that there is a maximum product when $a = 7$. What is the second number, b? Well we know that they add to 14, so:

$$\begin{aligned} b = 14 - a \longrightarrow b &= 14 - 7 \\ &= 7 \end{aligned}$$

So our two numbers are 7 and 7, and they multiply to get 49, the biggest possible product between two numbers that add to 14.

We have accomplished something, even if it's not much. Hurray, we solved a pretty unimportant problem about numbers!

More importantly, though, we set up a process for how we're going to approach optimization problems.

Optimization Process.

1. *Label variables.* What are the quantities that we care about? What are the kinds of measurements that we're given or need to find information about? Label them!

2. *Find a formula to optimize.* What are we trying to find the maximum or minimum of? Is it a formula that we might already know, or is there some other way of constructing that formula? This might involve drawing some geometric shape to get a clue!

3. *Find a constraint.* The **constraint** is really just some other connection between variables that guides their relationship. If we know some of the variables, there might be a way to find another one, since it would then depend on that variable.

4. *Find some domains.* For your variables, what are the smallest and largest possible values that they can reasonably take on? Are there any?

5. *Set up a calculus statement.* A **calculus statement** is a sentence that includes:

   (a) A function you are optimizing. This should be a function with only one input, so we might need to use the constraint to re-write out formula from earlier in order for it to only have a single input variable.

(b) A domain for that function. This should be the interval you found earlier that is relevant for your choice of input variables.

(c) Some indication of whether you're finding a maximum or a minimum.

An example of a calculus statement might be something like: "We want to find the $\boxed{\text{maximum/minimum}}$ of $\boxed{\text{function}}$ on $\boxed{\text{domain}}$."

6. *Do calculus.* We want to find a global maximum or a minimum of a function...we know how to do that! We'll find critical points, and then use the First Derivative Test, the Second Derivative Test for Local Maximums or Local Minimums, or the Extreme Value Theorem (and it's follow-up strategy) to find the global or absolute maximum/minimum!

In our example, we did the following:

1. *Label variables.* Our two numbers were a and b.
2. *Find a formula to optimize.* We said that $P = ab$ was the product.
3. *Find a constraint.* We knew that $a + b = 14$
4. *Find some domains.* We reasoned that $0 \leq a \leq 14$ and $0 \leq b \leq 14$.
5. *Set up a calculus statement.* We wanted to find the maximum of $P(a) = a(\underbrace{14 - a}_{b})$ on $[0, 14]$.
6. *Do calculus.* We showed how we could use three different techniques to solve this.

Balancing Volume and Surface Area

Let's use this new Optimization Process to solve a real problem. This one is a classic problem that (in my opinion) every student should try. I hope you'll see why.

Activity 4.5.1 Constructing a Can.

A typical can of pop is 355 ml, and has around 15 ml of headspace (air). This results in a can that can hold approximately 23 cubic inches of volume.

Let's say we want to construct this can in the most efficient way, where we use the least amount of material. How could we do that? What are the dimensions of the can?

(a) First, let's think of our can and try to translate this to some geometric shape with variable names. Collect as much information as we can about this setup! What is the shape? What are the names of the dimensions?

(b) What is the actual measurement that we are trying to optimize? Are we finding a maximum or a minimum?

(c) What other information about the can do we know? How do we translate this into a constraint, or a connection between our variables?

(d) What kinds of values can our variables take on? Is there a smallest value for either? A largest? Are there other limitations to these?

(e) Now we need to set up a calculus statement. This part mostly relies on us finding a way to build a single-variable function defining the surface area. Build that function, and then write down the calculus statement.

(f) Do some calculus to find the global maximum or minimum, and solve the optimization problem.

(g) What is the relationship between r and h, here? How do they compare? What about the height and diameter of our can?

Is this relationship noticeable in a typical can of pop?

So why, then, do we never see cans that look like this? It is certainly worth thinking about how the setup and assumptions we made in this activity might not be the way things work in real life.

What are some reasons that we might not see these dimensions in a can of pop?

Note 4.5.2

This question (why do the dimensions of cans not match what we found as the optimal solution?) is interesting, and sometime in the in the 1980's, a math professor felt strongly enough about it that they wrote a letter to Carnation, a brand of food products that produces canned goods, asking why reality doesn't match mathematics.
The professor received a response from the Assistant Product Manager of Friskies Buffet with 5 reasons for the non-square dimensions. The full text of the letter can be found here: Appendix A Carnation Letter.

What Other Examples Can We Do?

There's a really important point to make from the Carnation Letter: in real life, our optimization problems are multi-variable problems. We're balancing *many* different aspects of a process or a problem to find an optimal solution. That's hard to do in a calculus course that focuses on single-variable functions!

So what kinds of problems can we actually do?

There are a bunch, but they don't really stand up to intense scrutiny: if we looked carefully at most of the ”classic” optimization problems that we see in calculus texts, they'd fall apart just like the optimal can problem from Activity 4.5.1.

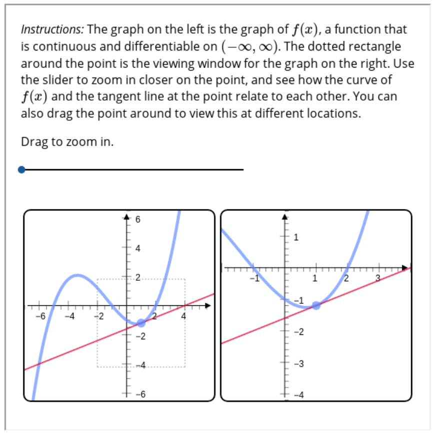
That said, it's good to practice thinking about using constraints and translating physical situations into formulas and functions, so try the practice problems to get used to this part of the optimization process!

4.6 Linear Approximations

We're going to return to a pretty central idea here, one that we've been using and developing and really exploring. But let's think about the very basic version of what we've been looking at over this whole chapter (and more):

The derivative of a function tells us the slope of the line tangent to the function at a point.

But what we'll do is explore how this tangent line and the graph of our function interact and relate to each other. Let's start with just playing with a graph and seeing if we can discover some things to say about the relationship between a tangent line and the function it is lying tangent to!



Standalone Embed

The effect that we are seeing when we zoom in on a function is sometimes described as our function being **locally linear**. What do you think this means? Why is this a good description of what we're looking at, and how these differentiable functions are constructed?

Would this effect be noticeable for every function, even ones that are not differentiable at some points?

Convince yourself that a function will not look locally linear at a point where the function is not differentiable. You might want to remind yourself what it looks like, graphically, when a derivative doesn't exist: When Does a Derivative Not Exist?

Linearly Approximating a Function

The visual above should provide us with a nice framework to think about how we might approximate a function linearly, but we can recap some basic ideas:

- When we say "linear approximation," we're really just referring to the tangent line at some point.
- Our functions only look "locally linear" when we zoom in around some single point. Another way of saying this is that our tangent line only matches the behavior of our function really close to the point where we built the tangent line.
- We have a kind of vague or ambiguous idea of accuracy in approximation. While a tangent line follows the behavior of the function "around" that point where it was built, the actual rate at which it deviates from our function is different. If we move the point in the visual above, we'll see that at some locations, our function is pretty linear and doesn't move away from the tangent line very quickly. In other locations, the function turns quickly away from the tangent line!

Definition 4.6.1 Linear Approximation of a Function.

If $f(x)$ is differentiable at $x = a$, then we say that a **linear approximation** of $f(x)$ centered at $x = a$ is:

$$L(x) = f'(a)(x - a) + f(a).$$

We know, then, that $L(x) \approx f(x)$ for x-values "close" to the center, $x = a$.
Note that the **center** is just the point at which we are building this linear approximation: the point at which we build the tangent line.

Let's see this in action!

Activity 4.6.1 Approximating an Exponential Function.

Let's consider the function $f(x) = e^x$. We're going to build the linear approximation, $L(x)$, but we also want to focus on understanding what the "center" is, and how we think about accuracy of our estimations.

(a) We first need to find a "good" center for our linear approximation. We have two real requirements here:

(a) We need the center to be some x-value that will be "close" to the inputs we're most interested in. We know that $L(x) \approx f(x)$ for x-values "near" the center, so we should keep this in mind. We don't have a specific input that we're interested in (we are not specifically focused on estimating $f(7.35)$ for instance), so we don't need to worry about this for now.

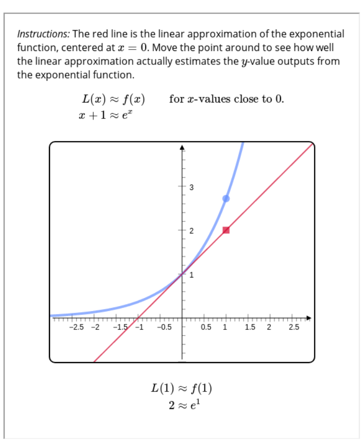
(b) We are going to need to evaluate the function and its derivative at the center: we use $f'(a)$ to find the slope and $f(a)$ to find a y-value for a point on the line. We'd like to choose a center that will make evaluating these functions reasonable, if we can!

We are going to choose a center of 0: why?

(b) Build a linear approximation of $f(x) = e^x$ centered at $x = 0$.

(c) Use your linear approximation to estimate the value of $\sqrt[10]{e} = e^{\frac{1}{10}}$.

(d) Let's visualize this approximation a bit:



Standalone
Embed

Are you confident in your approximation of $\sqrt[10]{e}$? Would you be more or less confident in an approximation of $\frac{1}{e}$? Why?

(e) Is your estimate of $\sqrt[10]{e}$ too big or too small? How can you tell, without even calculating the actual value of $\sqrt[10]{e}$?

How can you tell that *every* estimate that you get out of *any* linear approximation of e^x (no matter what the center is) is going to be too small?

In this activity, we did not have to think much about an appropriate choice of center. We tried to justify our choice, but that's different than having to *make* a choice. Let's approach this a bit differently in our next examples.

Activity 4.6.2 Approximating some Values.

Pick one of the following values to approximate:

- $\sin(-0.023)$
- $\ln(2)$
- $\sqrt{8}$
- $\sqrt[3]{10}$

Throughout the rest of this activity, use your value to build a linear approximation of some relevant function and estimate the value you chose.

(a) To build a linear approximation of some function at some center, we need two things:

(a) A function.

(b) A center.

What function will you be using for $f(x)$? Why that one?

(b) What center are you choosing? Why that one?

(c) Build your linear approximation at your center! You should end up with an actual linear function. It might be helpful to plot this linear function and your actual function to confirm that you have actually built a tangent line.

(d) Use your linear approximation function to estimate your value! Report the estimate, and comment on the accuracy of your estimate. Without calculating the actual value, can you tell if this is close or not? Do you have an overestimate or underestimate?

So far, we have been pretty limited in what we can actually *do* with these linear approximations. A function is only locally linear when we look at a very small interval of x-values. Once we move away from the center far enough (and it's often not that far), then our function curves away from the tangent line and our linear approximation is not at all accurate.

> If you want to see how we can make these linear approximations more accurate, then we can try to think about using a quadratic or cubic function instead: something with some curves built into it that can try to follow the function's behavior a bit! We'll cover that in Section 9.1 Polynomial Approximations of Functions.

Approximating Zeros of a Function

Let's look one really cool application of linear approximations before we finish things up in this section.

In approximately 60 AD, Heron of Alexandria presented a method for approximating square roots (probably...historians know very little about exactly when Heron was born and died, but they think he saw an eclipse that matched one from 62 AD, so it's a good guess). This algorithm was presented along with different formulas for volumes and surface areas of a mixture of objects.

> You might know of Heron from Heron's formula for the area of a triangle!

Over 1000 years later, in the late 1660's, Isaac Newton was one of a long list of mathematicians to re-create this formula in a more general way, where we can use it to approximate roots of polynomials. This method was later extended by several different mathematicians, and is now known as the **Newton-Raphson method**, or sometimes more simply **Newton's method**.

Activity 4.6.3 Walking in the Footsteps of Ancient Mathematicians.

Let's travel all the way back to the first (or maybe second) century AD and recreate Heron's method to approximate the value of $\sqrt{2}$. We'll develop this using modern calculus, and simple linear approximation. We're going to re-frame the problem, and instead we're going to try to use a linear approximation of $f(x) = x^2 - 2$ to approximate the x-value where $f(x) = 0$. We know enough about quadratic functions to know that there are two values: $x = -\sqrt{2}$ and $x = \sqrt{2}$.

(a) We're going to build a linear approximation of $f(x) = x^2 - 2$, and we need a reasonable center. Honestly, any integer will work, since we can evaluate f and f' really easily, but we want to find one that is close to $\sqrt{2}$. Let's center our approximation at $x = 2$.

Find $f'(x)$, and then construct the linear approximation:

$$L(x) = f'(2)(x - 2) + f(2).$$

(b) Now we know that $L(x) \approx f(x)$ for x-values near our center, $x = 2$. What if we estimate the x-value where $f(x) = 0$ by solving $L(x) = 0$ instead? Since $L(x) \approx f(x)$, the x-value where $L(x) = 0$ should make $f(x)$ pretty close to 0 at least.

Solve $L(x) = 0$.

(c) Ok, this might be kind of close to the value of $\sqrt{2}$, right? Let's visualize this.

Hm...so this isn't that good of an approximation yet. We can check this by looking at the actual value of our function at $x = \frac{3}{2}$

and seeing if it's close to 0.

$$
\begin{aligned}
f\left(\frac{3}{2}\right) &= \left(\frac{3}{2}\right)^2 - 2 \\
&= \frac{9}{4} - 2 \\
\frac{1}{4}
\end{aligned}
$$

This...isn't that close to 0.

So let's try this again. This time, though, let's center our *new* linear approximation at $x = \frac{3}{2}$.

(d) Now set this *new* linear approximation equal to 0 and solve $L(x) = 0$ to estimate the solution to $f(x) = 0$.

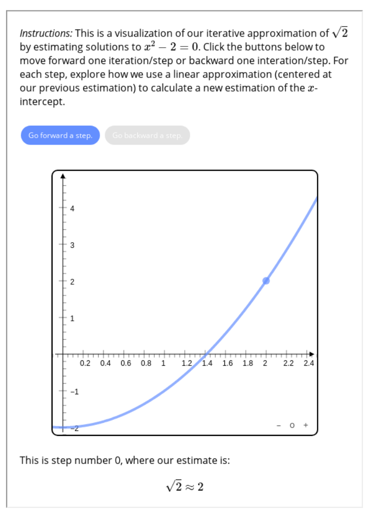
(e) We can keep repeating this process, and that's exactly what the mathematicians we talked about discovered.

Say we've build a linear approximation at some x-value (we'll call it x_{old}).

$$L(x) = f'(x_{\text{old}})(x - x_{\text{old}}) + f(x_{\text{old}}).$$

Set this equal to 0 and solve.

(f) Let's visualize these calculations.



Standalone
Embed

Something kind of strange happens in the last two steps. Why does the value of our estimation not change? What happens to our estimate?

Definition 4.6.2 Newton's Method for Approximating Zeros of Functions.

If x_0 is some initial estimation of a solution to $f(x) = 0$, then we we can iteratively generate more estimations using the following formula:

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}$$

provided that $f'(x_{n-1})$ exists and is non-zero.

A good question to ask is about when this process stops. If we want to estimate some x-intercept of a function, like $\sqrt{2}$ in Activity 4.6.3, then how many steps is enough? There are a couple of ways we can approach this:

1. We can just state at the beginning how many iterations we're going to do. This is what happened in Activity 4.6.3, since this activity was written to only make you calculate a specific number of these estimations. We could have started by saying that we'll calculate this 3 times, or maybe 100 times.

2. We can test to see what $f(x_n)$ is, and then stop when it is within some pre-determined distance from 0. We also did this when we noticed that $f\left(\frac{3}{2}\right) = \frac{1}{4}$ was not very close to 0 (after our first estimation), and so we should calculate this again. We could start by saying that we'll continue until we see a y-value that is within 0.0001 of 0, or some other small distance.

3. We can test to see how close our approximations are to each other, and stop when they're close enough. We saw this happen in the visualization: the last two estimates were the same! They actually weren't, but since the applet only displayed 4 decimal places, the numbers appeared the same after rounding. Maybe we set some criteria there, or we look at the distance between x_n and x_{n-1} (two successive estimates) and stop when they are within some distance from each other.

In reality, we often choose a combination of these. Maybe we set distance threshold for stopping, but use a maximum of 10 iterations as a backup plan. This happens often when we code this algorithm and have a computer run it. It is possible for this code to never give us two successive estimates that are close enough to stop, and so the code would run forever unless we cut it off at 100 iterations or some other value.

A wonderful thing about this small process is that, while it is ancient (dating back to Heron in the first or second century), it is still used today. This is a powerful estimation method that can be used in a variety of areas including statistics and data science.

4.7 L'Hôpital's Rule

We're going to re-visit limits, but with a slightly new problem-solving tool. Specifically, we'll be thinking about Indeterminate Forms. We noticed, back in Section 1.3, that we could evaluate limits for indeterminate forms by swapping out the function with another function that was mostly equivalent, only differing at the x-value we were approaching in the limit (Theorem 1.3.3 Limits of (Slightly) Different Functions).

We ended that section by thinking about a limit where this was difficult, in Activity 1.3.3.

We're now going to build a more systematic (and helpful) way of thinking about these limits using the ideas of Linear Approximation!

Indeterminate Forms

We have given a preliminary definition of Indeterminate Forms already (Definition 1.3.4), but let's remember how these work.

We said that $\frac{0}{0}$ is an indeterminate form, since a limit whose numerator and denominator approach 0 can end up taking on different values or even not exist. For instance, we can notice that the definition of the Derivative at a Point is a limit with this indeterminate form. As long as $f(x)$ is continuous (a necessity of it being differentiable) at $x = a$, then:

$$\lim_{x\to a}\left(\frac{f(x)-f(a)}{x-a}\right) \overset{?}{\to} \frac{f(a)-f(a)}{a-a} = \frac{0}{0}$$

But we have seen so many different values that this limit can end up being! We have spent most of the past two chapters in this text playing with derivatives and evaluating them: all of those values come from this limit! We have also seen that, even for continuous functions, this limit may not exist. A function can be non-differentiable at $x = a$.

We can show the same thing for a second indeterminate form: $\frac{\pm\infty}{\pm\infty}$, which we will simplify by just using the symbol $\frac{\infty}{\infty}$. For us to see that limits with this form can take on different values (or not exist), we just need to think about end behavior limits for rational functions (Subsection).

Let's think about the following limit:

$$\lim_{x\to\infty}\left(\frac{2x^m+1}{1-3x^n}\right).$$

As long as $m, n > 0$, then this limit looks like it's in the form of $\frac{\infty}{\infty}$. Sure, the denominator is really approaching $-\infty$, but we really just mean that there is an infinite numerator and an infinite denominator, regardless of sign.

We also know that the actual limit depends on the degrees m and n! Try to spend a couple of minutes confirming the next few claims:

- If $m < n$, then this limit is 0.
- If $m = n$, then this limit is $-\frac{2}{3}$.
- If $m > n$, then this limit doesn't exist.

All of this to show us that we have some forms of limits where we can't immediately tell what the actual value of the limit is (or if it even exists). L'Hôpital's Rule will be a way for us to navigate these limits a little easier than before, in some cases.

L'Hôpital's Rule

Activity 4.7.1 Building L'Hôpital's Rule.

We're going to take a closer look at the indeterminate form, $\frac{0}{0}$, and use our new ideas of linear approximation to think about how these types of things work.

We're going to be working with the following limit:

$$\lim_{x\to a}\frac{f(x)}{g(x)}$$

where $f(x)$ and $g(x)$ are differentiable at $x = a$ (since we're going to want to build linear approximations of them).

(a) Write out the linear approximations for both $f(x)$ and $g(x)$, both centered at $x = a$. We'll call them $L_f(x)$ and $L_g(x)$.

(b) Describe how well or how poorly these linear approximations estimate the values from our functions $f(x)$ and $g(x)$? What happens to these approximations as we get close to the center $x = a$? What happens in the limit as $x \to a$?

(c) Let's re-write our limit. We can replace $f(x)$ with our formula for its linear approximation, $L_f(x)$ and replace $g(x)$ with its linear approximation, $L_g(x)$:

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \left(\frac{\quad\quad\quad\quad}{\quad\quad\quad\quad} \right)$$

(d) Up until now, we have not thought about indeterminate forms at all. Let's start now.

If this limit is a $\frac{0}{0}$ indeterminate form, then that means that $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$.

Since our functions are, by definition, differentiable at $x = a$, then they also have to be continuous at $x = a$. What does this mean about the values of $f(a)$ and $g(a)$?

(e) Use this new information about the values of $f(a)$ and $g(a)$ to revisit the limit. We re-wrote $\lim_{x\to a} \dfrac{f(x)}{g(x)}$ by replacing each function with its linear approximation. What happens with the algebra when we know this information about $f(a)$ and $g(a)$?

So we have a really nice result here! In the $\frac{0}{0}$ indeterminate form, we can replace the ratio of the y-values from our functions with the ratio of slopes (coming from the first derivatives) of our functions.
In general, we'll put a step in between, where we find $f'(x)$ and $g'(x)$ first before trying to evaluate these derivatives at $x = a$.

Theorem 4.7.1 L'Hôpital's Rule.

If $f(x)$ and $g(x)$ are functions and a is some real number with f and g both being differentiable at a and $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$, then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

Similarly, this holds if $\lim_{x\to a} f(x) = \pm\infty$ and $\lim_{x\to a} g(x) = \pm\infty$.
If f and g are both differentiable as $x \to \infty$ and either:

- $\lim_{x\to\infty} f(x) = 0$ *and* $\lim_{x\to\infty} g(x) = 0$
- $\lim_{x\to\infty} f(x) = \pm\infty$ *and* $\lim_{x\to\infty} g(x) = \pm\infty$

then

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{f'(x)}{g'(x)}.$$

This is also true as $x \to -\infty$.

Example 4.7.2 Some First Limits.

Evaluate the following limits. You should first confirm that they are, actually, indeterminate forms!

(a) $\lim_{x\to 0}\left(\frac{\sin(x)}{x}\right)$

(b) $\lim_{x\to 5}\left(\frac{x^2-6x+5}{x-5}\right)$

(c) $\lim_{x\to 1}\left(\frac{\ln(x)}{x-1}\right)$

(d) $\lim_{x\to \infty}\left(\frac{x^2-6x+5}{x-5}\right)$

(e) $\lim_{x\to \infty}\left(\frac{\ln(x)}{x-1}\right)$

There are more indeterminate forms than these two! In each of the following cases, we mean that a limit with this form can take on different values (or not exist). Other indeterminate forms that we can consider include:

- $f(x)\cdot g(x) \stackrel{?}{\to} 0\cdot\infty$
- $(f(x)-g(x)) \stackrel{?}{\to} \infty-\infty$
- $f(x)^{g(x)} \stackrel{?}{\to} 0^0$
- $f(x)^{g(x)} \stackrel{?}{\to} 1^\infty$
- $f(x)^{g(x)} \stackrel{?}{\to} \infty^0$

The issue with these, though, is that L'Hôpital's Rule only applies to quotients! We needed that quotient for the algebra to work out when we canceled things out to end up with the ratio of slopes.

So our strategies for these other indeterminate forms will all require us to manipulate the product, difference, or expeonential in order to force some division to show up somehow.

Forcing Division

Let's look at each new indeterminate form classified into groups based on the operation between the functions.

Products!

We can re-write $f(x)\cdot g(x)$ as a quotient by dividing by a reciprocal. So either

$$f(x)\cdot g(x) = \frac{f(x)}{1/g(x)}$$

or

$$f(x)\cdot g(x) = \frac{g(x)}{1/f(x)}.$$

Our choice ends up being based on what is most helpful.

Example 4.7.3

Evaluate the limit:

$$\lim_{x\to 0^+} (x\ln(x))$$

Note that since $x \to 0$ and $\ln(x) \to -\infty$, this is a $0 \cdot \infty$ indeterminate form.

Hint. Re-write this limit as:

$$\lim_{x\to 0^+} \left(\frac{\ln(x)}{1/x}\right).$$

Note that this is not an $\frac{\infty}{\infty}$ indeterminate form, and we can use L'Hôpital's Rule.

Solution.

$$\begin{aligned}
\lim_{x\to 0^+} (x\ln(x)) &= \lim_{x\to 0^+} \left(\frac{\ln(x)}{1/x}\right) \\
&= \lim_{x\to 0^+} \left(\frac{\frac{d}{dx}(\ln(x))}{\frac{d}{dx}(1/x)}\right) \\
&= \lim_{x\to 0^+} \left(\frac{1/x}{-1/x^2}\right) \\
&\lim_{x\to 0^+} \left(\frac{1}{x}\right)\left(-\frac{x^2}{1}\right) \\
&= \lim_{x\to 0^+} (-x) \\
&= 0
\end{aligned}$$

So $\lim_{x\to 0^+} x\ln(x) = 0$.

Differences!

We can re-write $f(x) - g(x)$ as a product by factoring something out of the difference. Then, if the product is a $0 \cdot \infty$ indeterminate form, we can divide by a reciprocal to turn it into a quotient.

Choosing what to factor out is sometimes very difficult. But we should note that this is the strategy we used to evaluate Polynomial End Behavior Limits.

Example 4.7.4

Evaluate the following limit:

$$\lim_{x\to\infty} \left(2^x - x^2\right)$$

Note that since $2^x \to \infty$ and $x^2 \to \infty$, this is an $\infty - \infty$ indeterminate form.

Hint. Try to factor out 2^x. You won't be able to actual factor it nicely, but you'll end up with a fraction term $\frac{x^2}{2^x}$ that is an $\frac{\infty}{\infty}$ indeterminate form!

Solution.

$$\lim_{x\to\infty}\left(2^x - x^2\right) = \lim_{x\to\infty} 2^x\left(1 - \frac{x^2}{2^x}\right)$$

Let's focus on the limit $\lim_{x\to\infty}\left(\frac{x^2}{2^x}\right)$, since it is in an $\frac{\infty}{\infty}$ indeterminate form.

$$\begin{aligned}
\lim_{x\to\infty}\left(\frac{x^2}{2^x}\right) &\overset{?}{\to} \frac{\infty}{\infty} \\
&= \lim_{x\to\infty}\frac{\frac{d}{dx}\left(x^2\right)}{\frac{d}{dx}\left(2^x\right)} \\
&= \lim_{x\to\infty}\frac{2x}{2^x\ln(2)} \\
\lim_{x\to\infty}\frac{2x}{2^x\ln(2)} &\overset{?}{\to} \frac{\infty}{\infty} \\
&= \lim_{x\to\infty}\frac{\frac{d}{dx}\left(2x\right)}{\frac{d}{dx}\left(2^x\ln(2)\right)} \\
&= \lim_{x\to\infty}\frac{2}{2^x\ln(2)\ln(2)} \\
\lim_{x\to\infty}\frac{2}{2^x\ln(2)\ln(2)} &\overset{?}{\to} \frac{2}{\infty} = 0
\end{aligned}$$

So then we can go back to our original limit:

$$\begin{aligned}
\lim_{x\to\infty}\left(2^x - x^2\right) &= \lim_{x\to\infty} 2^x\left(1 - \frac{x^2}{2^x}\right) \\
&= \infty(1-0) = \infty
\end{aligned}$$

Exponentials!

We can think about how we approached these types of functions raised to functions when we learned about Logarithmic Differentiation.

We were able to use logarithms to re-write these types of exponentials as products. So we can say that:

$$\begin{aligned}
f(x)^{g(x)} &= e^{\ln\left(f(x)^{g(x)}\right)} \\
&= e^{g(x)\ln(f(x))}
\end{aligned}$$

When we think about limits, the continuity of the exponential function allows us to just focus on the limit of the exponent, $g(x)\ln(f(x))$, which is likely an indeterminate form that we've seen!

Example 4.7.5

Evaluate the following limit:

$$\lim_{x\to 0^+} x^x$$

Note that this is the 0^0 indeterminate form.

Hint. We can re-write x^x as $e^{\ln(x^x)}$ which is the same as $e^{x\ln(x)}$. Now we can evaluate the limit $\lim_{x\to 0^+} x\ln(x)$, and make sure to return the value into the exponent.

Solution. We know from Example 4.7.3 that $\lim_{x\to 0^+} x\ln(x) = 0$. So then:

$$\begin{aligned}
\lim_{x\to 0^+} x^x &= \lim_{x\to 0^+} e^{x\ln(x)} \\
&= e^{\lim_{x\to 0^+} x\ln(x)} \\
&= e^0 = 1
\end{aligned}$$

So $\lim_{x\to 0^+} x^x = 1$.

Chapter 5

Antiderivatives and Integrals

5.1 Antiderivatives and Indefinite Integrals

We've been spending a lot of time thinking about derivatives! We've done this in a couple of different ways:

1. We have thought carefully about what derivatives are, what they measure, and how to interpret them.
2. We have built up a whole list of tools that we can use to actually find or calculate these derivatives. We know how to differentiate most functions (and combinations of functions) that we can think of!
3. We've been able to apply these derivatives to some specific contexts to solve problems or analyze functions and mathematical models.

Let's think about derivatives in a slightly different way!

Activity 5.1.1 Find a Function Where....

For each of the following derivatives, find a function $f(x)$ whose first (or second) derivative matches the listed derivative.

(a) $f'(x) = 4x^3 + 6$

(b) $f'(x) = 8x^7 - x^4 + x$

(c) $f'(x) = \sqrt{x} - \dfrac{1}{x^2}$

(d) $f''(x) = x^3 + \cos(x)$

(e) $f''(x) = e^x - \dfrac{1}{\sqrt[3]{x}}$

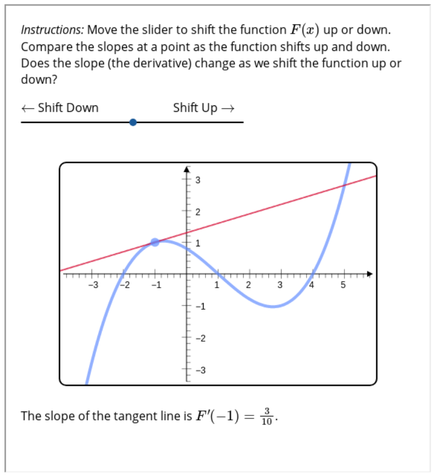
(f) Go back through each of the above derivatives, and find a different option for $f(x)$ that still works. Make sure that it is something completely unique, and not just an equivalent function that is written differently.

Why are able to find multiple answers in these questions, but not when we are given a function and need to find a derivative?

We've done two things here: thought about how we might "undo" differentiation, and discovered a nice property about constants.

Note that we've already discovered this rule! We proved it, back when we were playing with the Mean Value Theorem. We built a related theorem that showed that two functions can have the same derivative, and if they do then they are off by, at most, a constant: Theorem 4.1.7 Equal Derivatives Correspond with Related Functions.

Let's visualize this phenomena!



Standalone
Embed

Antiderivatives

We want to try to define and name these "backwards derivatives." Instead of calling them the "negative first" derivative, we will name them as **antiderivatives**.

Definition 5.1.1 Antiderivative.

For a function $f(x)$, we say that $F(x)$ is an **antiderivative** of $f(x)$ on an interval if $F'(x) = f(x)$ on the interval.
We call $F(x) + C$ the **family of antiderivatives** of $f(x)$, where C represents any real number constant.

Example 5.1.2

For each of the following functions, find the family of antiderivatives.

(a) $f(x) = 7x + \sec^2(x)$

Hint. Do we know a function whose derivative is $\sec^2(x)$?

Solution. $F(x) = \dfrac{7x^2}{2} + \tan(x) + C$

(b) $g(x) = \dfrac{5}{x} - \dfrac{3}{1+x^2}$

Hint. We won't be undoing the Power Rule with either of these! We might try to think about functions whose derivatives are $\frac{1}{x}$ and $\frac{1}{1+x^2}$.

Solution. $G(x) = 5\ln|x| - 3\tan^{-1}(x) + C$

We use absolute values in the logarithm because we want to find a functions whose derivative is $g(x)$ on the whole interval that $g(x)$

is defined. The log function is only defined for positive inputs, but we would like to be able put any non-zero input into our function (since that's the domain of g).

(c) $j(x) = x^5 - 4x + 1 - \dfrac{4}{3x^5}$

Hint. It might help to write the function as $j(x) = x^5 - 4x + 1 - \dfrac{4}{3}x^{-5}$.

Solution. $J(x) = \dfrac{x^6}{6} - 2x^2 + x + \dfrac{1}{3x^4} + C$

Initial Value Problems

Activity 5.1.2 A File Sorting Speed Test.

A computer program is trying to sort a group of computer files based on their size. The program isn't very efficient, and the time that it takes to sort the files increases when it tries to sort more files.
The time that it takes, measured in seconds, based on the total, cumulative size of the files g, measured in gigabytes, is modeled by a function $T(g)$. We don't know the function, but we do know that the time increases at an instantaneous rate of $0.0001g$ seconds when the total size, g increases slightly.

(a) We can build a function for $T'(g)$. What is it?

(b) Find all of the possibilities for the function modeling the time, T, that it takes the computer program to sort files with a total size of g.

(c) What does your constant C represent, here? You can interpret it graphically, interpret it by thinking about derivatives, but you should also interpret it in terms of the time that it takes this program to sort these files by size.

(d) Let's say that we feed some number of files totaling up to 1.4GB in size into this program. It takes 0.24 seconds to sort the files by size.

Find the function, $T(g)$, that models the how quickly this program sorts these files.

We call this type of problem an "initial value problem." Here, we ended up solving for a family of antiderivatives, but then using some more information about that antiderivative (in this case, information about file size and time) to find the specific antiderivative function that was relevant.

Solving Initial Value Problems.

For some function $f(x)$, if we want to find an antiderivative function $F(x)$ and we know some "initial value," $F(a)$, then we can find the exact antiderivative by:

1. Finding the family of antiderivatives: $F(x) + C$.

2. Using the initial value to solve for the constant C, by evaluating $F(x) + C$ at $x = a$ and solving the resulting equation.

Example 5.1.3

(a) For $f(x) = \frac{x^5}{2} + \sin(x)$, find $F(x)$ where $F(0) = 3$.

(b) For $g'(x) = e^x$, find $G(x)$ where $G(0) = 4$ and $g(0) = 2$.

Indefinite Integrals

To finish this out, we'll just build some notation that represents these families of antiderivatives. We can use words to describe them, but it will be helpful to introduce some quick way of writing this using notation.

Definition 5.1.4 Indefinite Integral.

An **indefinite integral** represents a family of antiderivatives:

$$\int f(x)\, dx = F(x) + C$$

where

- $\int$ is a symbol directing us to find a family of antiderivatives (or integrate)
- $f(x)$ is called the integrand
- dx is a differential, and represents both the "end" of the integral as well as an indicator of what the input variable of the integrand should be (or what variable we antidifferentiate "with regard to").
- $F(x)$ is an antiderivative of $f(x)$ (where $F'(x) = f(x)$).
- C is called the "constant of integration" and represents any real number

Example 5.1.5

Find families of antiderivatives according to each of the following indefinite integrals.

(a) $\displaystyle\int \left(\frac{4}{x} - \sqrt{x}\right)\, dx$

(b) $\displaystyle\int (x+4)(x^2-7)\, dx$

Hint. While we do not know how to antidifferentiate products of functions yet, we can just multiply the integrand function!

$$(x+4)(x^2-7) = x^3 + 4x^2 - 7x - 28$$

Antidifferentiate this.

(c) $\displaystyle\int \left(\frac{xe^x - 1}{x}\right)\, dx$

Hint. Similar to the previous problem, we do not know how to antidifferentiate quotients, but we can re-write this function before we antidifferentiate!

$$\begin{aligned}\frac{xe^x - 1}{x} &= \frac{xe^x}{x} - \frac{1}{x}\\ &= e^x - \frac{1}{x}\end{aligned}$$

Antidifferentiate this!

All we have left to do now is to just formalize the antiderivative rules we've been intuitively building and using.

Theorem 5.1.6 Power Rule for Antiderivatives.

$$\int x^n\, dx = \frac{x^{n+1}}{n+1} + C$$

for $n \neq -1$

Theorem 5.1.7 Antiderivatives Related to the Exponential and Log Functions.

$$\begin{aligned}&\int e^x\, dx = e^x + C\\ &\int b^x\, dx = \frac{b^x}{\ln(b)} + C \qquad \text{for } b > 0 \text{ and } b \neq 1\\ &\int \frac{1}{x}\, dx = \ln|x| + C\end{aligned}$$

Theorem 5.1.8 Antiderivatives of Trigonometric Functions.

$$\begin{aligned}\int \sin(x)\, dx &= -\cos(x) + C\\ \int \cos(x)\, dx &= \sin(x) + C\end{aligned}$$

Theorem 5.1.9 Combinations of Indefinite Integrals.

1. Sums: $\displaystyle\int (f(x) + g(x))\, dx = \int f(x)\, dx + \int g(x)\, dx$
2. Differences: $\displaystyle\int (f(x) - g(x))\, dx = \int f(x)\, dx - \int g(x)\, dx$
3. Coefficients: $\displaystyle\int (kf(x))\, dx = k\int f(x)\, dx$

These should all be very familiar, since they are really just restatements of the results from Section 2.3 Some Early Derivative Rules.

We should also be comfortable recognizing derivatives of functions that we know, in order to find more functions that we can antidifferentiate.

If we are following the path set out by us already when we learned about derivatives, then at some point we will need to think about how to interpret these antiderivatives. What does $F(x)$ tell us about $f(x)$?

What does $f(x)$ tell us about $f'(x)$? We're probably so used to thinking about what $f'(x)$ tells us about $f(x)$ that it might be hard to reverse the interpretation. And that's ok!

Instead of worrying about this, we can just present us with the answer, and then spend some time uncovering it more.

Over the next few sections, we'll discover that antiderivatives of $f(x)$ are deeply connected to areas carved out by the graph of $f(x)$.

5.2 Riemann Sums and Area Approximations

One of the last things we said in Section 5.1 was that antiderivatives will be connected to areas. We're going to eventually show this! For now, though, we want to focus on areas defined by curves.

Activity 5.2.1 Approximating Areas.

We're going to consider two different functions, and some areas based on them. Let's think about two functions: $f(x) = 2x + 1$ and $g(x) = x^2 + 1$. For both of these functions, we'll focus on the interval $[0, 2]$. Instead of thinking about the function only, we'll be considering the two-dimensional region bounded between the graph of our function and the x-axis between $x = 0$ and $x = 2$.

(a) Find the area of the region bounded between the graph of $f(x) = 2x + 1$ and the x-axis between $x = 0$ and $x = 2$.

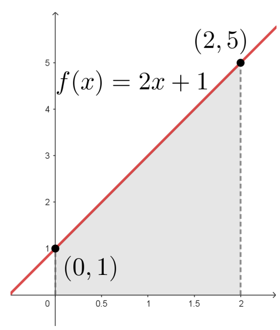


Figure 5.2.1

How did you evaluate this area? What kind(s) of shape(s) did you think about?

(b) Estimate the area of the region bounded between the graph of $g(x) = x^2 + 1$ and the x-axis between $x = 0$ and $x = 2$.

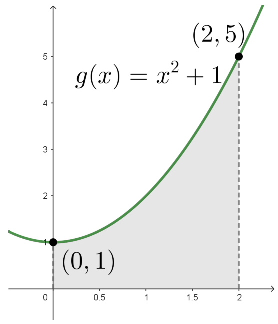


Figure 5.2.2

> Archimedes of Syracuse discovered how to calculate this area exactly, without estimation, around 300 BC, writing his results in the now-famous "Quadrature of the Parabola." This is, notably, before the formalization of Calculus (during the 1600's). It might be unfair to say that Archimedes proved this "without using calculus," though, since his technique, the "Method of Exhaustion," is really a version of what we do in calculus, but without a formal framework of limits.

How did you estimate this area? What kind(s) of shape(s) did you think about?

(c) Come up with an upper and lower bound for this area. In other words, give an underestimate and overestimate for the actual area we would like to know.

How did you come up with these estimates? How "good" do you think your estimates are? Can you come up with "better" (or closer) ones?

Hopefully we've had a chance to think about and compare a couple of different strategies for estimating this area. What we want to do, though, is build a systematic way of estimating this area. We'd like it to have a couple of features:

- Easy area calculations. We don't want to have to spend a lot of time thinking about tricky area formulas, so simple shapes will be nicer to use.

- Flexibility. We want to be able to apply our approach to an area defined by any curve.

> This is the problem with Archimedes' method: it only worked for areas defined by parabolas. Once we change our function to something else, Archimedes would have to come up with a completely new area formula for calculation. The techniques we're looking at now have the advantage of flexibility!

- Precision. We want to be able to make our estimates as precise as we'd like. It's fine to come up with rough estimates, but we would like a method that allows us to increase the accuracy in our estimations.

Rectangular Approximations

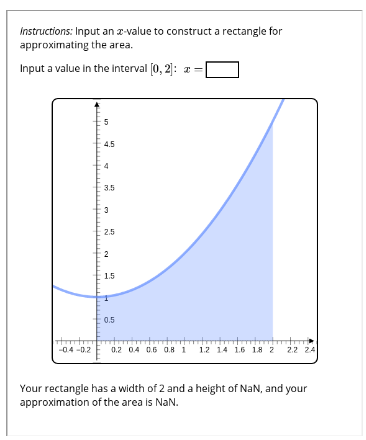
We're going to re-visit the same region as before, but this time we'll outline a process that should help us approximate the area with as much precision as we'd like.

Activity 5.2.2 Approximating the Area using Rectangles.

We're going to stick with the function $g(x) = x^2 + 1$ on the interval $[0, 2]$, and keep thinking about the area bounded by the curve and the x-axis on this interval. We're going to approximate the area in a couple of different tries, each one more accurate than the one before. By the end of this activity, we'll have a pretty good process built!

(a) Let's start with approximating this region with a single rectangle. We're going to define the rectangle by picking some x-value in the interval $[0, 2]$. Then, we'll use the point at that x-value to define the height of our rectangle.

Essentially, we are picking a single point on the our function on the interval and our approximation is pretending that the single point we picked is representative of the whole function on the interval.

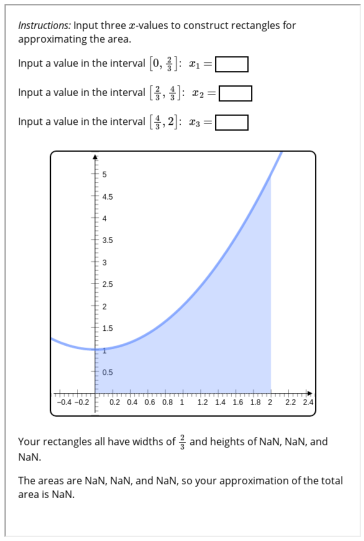


Standalone
Embed

(b) Can you try re-picking an x-value, and trying to find one that gets you an area approximation that is pretty good?

(c) We're going to use more rectangles. Let's jump up to 3 rectangles. If we split up the interval between $x = 0$ and $x = 2$ into 3 rectangles, we can make them all the same width, and pick an x-value that we can use to get a representative point for each of the 3 rectangles.

We'll need to pick 3 x-values this time.



Standalone
Embed

(d) Can you try re-picking your x-values, and trying to find one that gets you an area approximation that is pretty good?

(e) Let's scale this up a bit. Pick a good number for your number of rectangles. We'll call this value n.

(If you're working in a classroom, maybe it would be good to pick the number of groups or the number of students, or some other number between 10 and 20 or something like that.)

For your value n, we're going to divide up the interval between $x = 0$ and $x = 2$ into n pieces. These will be the intervals that we pick from to get our rectangles. What are the subintervals? What are the widths of each subinterval (and then the widths of the rectangles)? Call this with Δx.

(f) For each subinterval, pick an x-value in the subinterval to represent it.

(g) Evaluate the function $f(x) = x^2 + 1$ at each of the x-values you picked. These are the heights of your rectangles!

(h) Find the areas of each rectangle by multiplying the height of each rectangle by Δx, the width of each rectangle.

(i) Add these areas up to get a total approximation of the actual area!

What do you think: is it worth fiddling with what x-value to pick from each subinterval to try to get a better approximation? If n is large, do you think it matters how we pick the x-values from each subinterval?

This is our process! We'll refer to it often as the **slice-and-sum process**, since we are slicing out region into a bunch of pieces, approximating the area on each piece (by using one point to represent the whole slice), and then summing the areas back up.

More formally, we can call this the Riemann Sum process, because the sum of the areas is a special form of summation.

Definition 5.2.3 Riemann Sum.

For a closed interval $[a, b]$ with a partition $\{x_0, x_1, ..., x_n\}$ with $a = x_0 < x_1 < ... < x_n = b$, consider some x_k^*, any x-value in the interval $[x_{k-1}, x_k]$ and Δx_k, the length of the interval $[x_{k-1}, x_k]$. If f is a function that is defined on the interval $[a, b]$, then we call the sum

$$\sum_{k=1}^{n} f(x_k^*)\Delta x_k = f(x_1^*)\Delta x_1 + f(x_2^*)\Delta x_2 + ... + f(x_n^*)\Delta x_n$$

a **Riemann Sum** for f on $[a, b]$.

Note 5.2.4

In practice, we typically choose a *Regular Partition*, where each subinterval $[x_{k-1}, x_k]$ is equally-wide, and so $\Delta x_k = \frac{b-a}{n}$ for every $k = 1, 2, ..., n$. We then normally write our Riemann sum as

$$\sum_{k=1}^{n} f(x_k^*)\Delta x = f(x_1^*)\Delta x + f(x_2^*)\Delta x + ... + f(x_n^*)\Delta x$$

where Δx is the value of the widths of all of the equally-sized subintervals.

Selection Strategies

This is great! We have a scalable way to approximate areas, and it seems like we can pretty easily increase the precision of our approximations by increasing n, the number of slices/rectangles that we use. And the great thing about this is that when we do increase n, we don't increase the complexity of our calculations!

Sure, it would be tedious to calculate and add 100 areas of rectangles by hand, but those area calculations don't get more difficult: there are just more of them.

The only real downside is that when we increase the number of slices/rectangles, we are really increasing the number of decisions that we have to make: we have to choose an x_k^* for each subinterval, and so while it isn't hard to just calculate a bunch of areas and add them up, it is difficult, on a human level, to make a bunch of decisions about which x-value to choose from each subinterval. But this decision isn't even that important!

We use the "star" notation on the x_k^* to represent the fact that it really doesn't matter which x-value gets chosen from the subinterval: as long as we pick one, we get an approximation! And when n increases, it matters less and less what the actual x-value is: as long as our function $f(x)$ is continuous, then there will be not much variation among the y-value outputs for any x-values in each (small) interval!

All of this to say: let's make a single decision about picking n x-values from n subintervals instead of having to make n decisions (one for each x-value).

Left, Right, and Midpoint Riemann Sums.

When we build a Riemann sum, we can make a choice to systematically choose the values for x_k^* (for $k = 1, 2, ..., n$). There are many ways of doing this, but here are three:

- *Left Riemann Sum:* We pick the left-most x-value from each subinterval. That is, if the partition is $\{a = x_0, x_1, x_2, ..., b = x_n\}$, then we choose $\{a, x_1, x_2, ..., x_{n-1}\}$ as our x-values to evaluate f at for the rectangle heights.

  We refer to these as L_n, a Left Riemann sum with n rectangles.

- *Right Riemann Sum:* We pick the right-most x-value from each subinterval. That is, if the partition is $\{a = x_0, x_1, x_2, ..., b = x_n\}$, then we choose $\{x_1, x_2, ..., b\}$ as our x-values to evaluate f at for the rectangle heights.

  We refer to these as R_n, a Right Riemann sum with n rectangles.

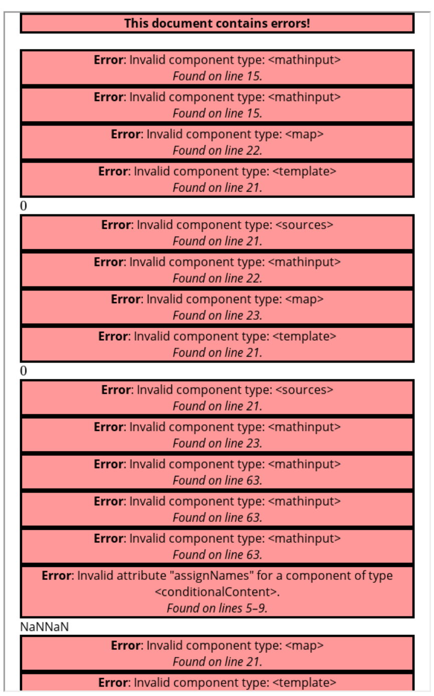
- *Midpoint Riemann Sum:* We pick the x-value that is in the middle of each subinterval. That is, if the partition is $\{a = x_0, x_1, x_2, ..., b = x_n\}$, then we choose $\left\{\frac{a+x_1}{2}, \frac{x_1+x_2}{2}, ..., \frac{x_{n-1}+b}{2}\right\}$ as our x-values to evaluate f at for the rectangle heights.

  We refer to these as M_n, a Midpoint Riemann sum with n rectangles.

None of this is a requirement for a Riemann sum, but we will consistently find that when we limit the number of decisions that we have to make, the complexity of the calculation decreases.

Notice that we've already made a similar choice with how we calculate Δx: it is not required that each rectangle have the same width, but it is very nice to not have to think about n different widths!

Lastly, we'll finish with a nice interactive Riemann sum calculator. Feel free to explore some different graphs and see how the Riemann sums work when we change how we select the values for x_k^* as well as when we change the number of rectangles, n.



Standalone
Embed

5.3 The Definite Integral

The big result from our last section on Riemann sums is not just that we can approximate areas by thinking about a bunch of small (thin) rectangles. The big result is that this strategy is scalable: we can increase n, the number of slices/rectangles, and essentially guarantee that, eventually, our approximations will be very accurate.

Now, we move from a concrete process for building rectangles to calculate areas to a more conceptual framework: what happens when $n \to \infty$?

Evaluating Areas (Instead of Approximating Them)

Our goal is to move from approximating area to evaluating areas: calculating the real value of the area of these regions bounded between curves and the x-axis. We have already decided (when we built the framework for Riemann sums and made scalability and precision in our estimates a focus) that the area we're interested in is the result of some limiting process: we increase the number of slices, n, and in turn decrease the width of each slice, Δx.

Definition 5.3.1 Definite Integral.

If $f(x)$ is some function defined on the interval $[a, b]$ and $\sum_{k=1}^{n} f(x_k^*)\Delta x$ is a Riemann sum with n slices and $\Delta x = \frac{b-a}{n}$, then we say that the **definite integral** of $f(x)$ from $x = a$ to $x = b$ is:

$$\int_{x=a}^{x=b} f(x)\,dx = \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k^*)\Delta x$$

if this limit exists. When this limit exists, we say that $f(x)$ is **integrable** on the interval $[a, b]$.

We call $x = a$ and $x = b$ the **limits of integration** for this definite integral, and we read $\int_{x=a}^{x=b} f(x)\,dx$ as "the integral from $x = a$ to $x = b$ of $f(x)$ with regard to x," or sometimes we might just say "of $f(x)\,dx$" for short.

Note 5.3.2

This is assuming we're using a *Regular Partition* (Note 5.2.4). If we are not, and each slice has its own width called Δx_k, then the definition of a definite integral requires that as $n \to \infty$ we see $\Delta x_k \to 0$ for all $k = 1, 2, ..., n$. Essentially, we need all of the widths to eventually get tiny: we can't let one slice take up half of the width and then let all of the other slices get tiny, since that would still be an approximation of the area we want.

We don't need to worry about this, though, since we'll always just choose to make all of the Δx_k's the same size: $\Delta x = \frac{b-a}{n}$.

Let us make something *very* clear: we will absolutely not calculate these areas this way. Let's see why not.

Let's say we want to calculate $\int_{x=0}^{x=2} (x^2 + 1)\,dx$. This is the area we were estimating in Activity 5.2.2 Approximating the Area using Rectangles. How many slices did you pick at the end of this activity? How annoying was it to add up those areas?

Whatever you did, it's not enough: even if we decided to divide this region up into $n = 1000$ pieces, this is merely an approximation of the limit we want:

$$\int_{x=0}^{x=2} (x^2+1)\, dx = \lim_{n\to\infty} \sum_{k=1}^{n} \underbrace{\left((x_k^*)^2+1\right)}_{f(x_k^*)} \underbrace{\left(\frac{2}{n}\right)}_{\Delta x}$$

There are some ways of evaluating this specific limit using some known formulas for sums of squares and end behavior limits of rational functions. But these techniques are extremely limited: we might get lucky being able to fiddle with this limit of this sum for this function, but we won't be so lucky in general.

Instead, let's just think about these areas, focus on what types of functions are integrable, and then build towards our end goal of connecting these areas to antiderivatives.

Signed Area

We're going to now deal with the consequences of our decisions. A truth about mathematics, sometimes not an obvious truth, is that every time we state a definition what we are actually doing is making a decision. We are deciding on some common way of classifying and describing an object. These classifications and descriptions are choices that we are making: choices to prioritize some property or aspect over a different one, choices to include or exclude a type of object into the group of things we're interested in, choices that come with downstream effects.

We chose to define the area bounded between a curve defined by the function $f(x)$ and the x-axis between $x = a$ and $x = b$ as:

$$\int_{x=a}^{x=b} f(x)\, dx = \lim_{n\to\infty} \sum_{k=1}^{n} f(x_k^*)\Delta x.$$

We are going to stand by this definition. It's a good one, for the reasons we described at the beginning of Section 5.2 Riemann Sums and Area Approximations.

But there are some weird things to notice. Let's notice them!

Activity 5.3.1 Weird Areas.

Let's think about a simple linear function, $f(x) = 4 - 3x$. We'll both approximate and evaluate the area bounded between $f(x)$ and the x-axis from $x = 0$ to $x = 3$:

$$\int_{x=0}^{x=3} (4-3x)\, dx$$

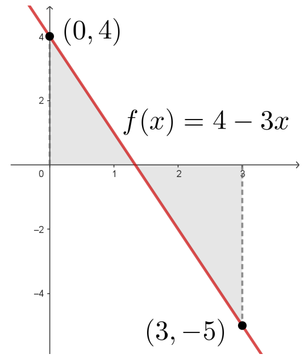


Figure 5.3.3

(a) Explain why we do not need to think about Riemann sums in order for us to calculate the shaded in area. How would you calculate this without using calculus?

Calculate the area!

(b) Let's approximate this area using a Riemann sum. Calculate L_3, the Left Riemann sum with $n = 3$ rectangles.

(c) Let's approximate this area a second time, but with a different selection strategy for our x-values. Calculate R_3, the Right Riemann sum with $n = 3$ rectangles.

(d) Compare your answers for L_3 and R_3. They should be *very* different. Why? What is happening that makes R_3 specifically such a weird value?

(e) Do you need to go back and re-calculate the area geometrically (from the first part of this activity)? Explain why your answer for $\int_{x=0}^{x=3} (4 - 3x)\, dx$ *should* be negative, based on the Riemann sums we calculated.

(f) Find a value $x = b$ such that:

$$\int_{x=0}^{x=b} (4 - 3x)\, dx = 0.$$

(g) Find a *different* value $x = b$ such that:

$$\int_{x=0}^{x=b} (4 - 3x)\, dx = 0.$$

Is there a second way of making this area 0?

Weird areas, right? Negative? That's not how we normally think about areas. So we have to be slightly careful with how we describe this new object, the definite integral, that we've built. We don't need to go back and change anything about the object itself: we just need to change how we talk about it.

It's common to think about $\int_{x=a}^{x=b} f(x)\, dx$ as the "area under the curve $f(x)$ from $x = a$ to $x = b$," but we know that's not really true. Instead, we'll think about it as a **signed area** of the region bounded between the curve

$f(x)$ and the x-axis from $x = a$ to $x = b$. When we say "signed area," we're just referring to the consequence of using y-values to define "heights" of the rectangles: when the curve is under the x-axis, we end up with negative values for heights, and so those rectangles have negative area.

Activity 5.3.2 Weird Areas - Part 2.

We're going to think about the same region, kind of.

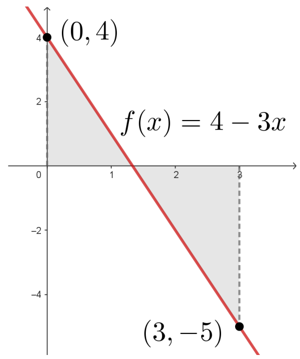


Figure 5.3.4

Let's think about the same linear function, $f(x) = 4-3x$, but this time we'll approximate and evaluate the area bounded between $f(x)$ and the x-axis from $x = 3$ to $x = 0$:

$$\int_{x=3}^{x=0} (4 - 3x)\, dx$$

(a) Use geometry to calculate the area. Compare this to the result from Activity 5.3.1.

(b) Let's approximate this using a Riemann sum. Calculate M_3, the Midpoint Riemann sum with $n = 3$ rectangles.

(c) Do you need to go back and re-calculate the area geometrically (from the first part of this activity)? Explain why your answer for $\displaystyle\int_{x=3}^{x=0} (4-3x)\, dx$ *should* be positive, based on the Riemann sums we calculated.

Ok so we have some intuition about how the signs of these areas work, and we've also built up some nice properties that we can talk through. Let's finish this section by just summarizing some of the things we've done and thinking about what kinds of functions this works for!

Properties of Definite Integrals

First, this result should be reasonable: we can always calculate these areas for continuous functions!

Theorem 5.3.5 Continuous Functions are Integrable.

If $f(x)$ is continuous on the interval $[a, b]$, then $f(x)$ is integrable on $[a, b]$. That is, the limit $\displaystyle\lim_{n\to\infty}\sum_{k=1}^{n} f(x_k^*)\Delta x$ exists and so we can evaluate *the definite integral:*

$$\int_{x=a}^{x=b} f(x)\, dx.$$

We'll come back to this, but first, let's summarize some properties that we've discovered.

Theorem 5.3.6 Properties of Definite Integrals.

If a. b, *and* c *are real numbers and* $f(x)$ *is a function that is continuous on the intervals* $[a, b]$ *and* $[b, c]$, *then:*

- *The signed area under a single point is 0:*

$$\int_{x=a}^{x=a} f(x)\, dx = 0$$

- *We can cut a region into pieces and evaluate the areas separately:*

$$\int_{x=a}^{x=c} f(x)\, dx = \int_{x=a}^{x=b} f(x)\, dx + \int_{x=b}^{x=c} f(x)\, dx$$

- *When we integrate a function "backwards" through an interval, we an area with an opposite sign:*

$$\int_{x=a}^{x=b} f(x)\, dx = -\int_{x=b}^{x=a} f(x)\, dx$$

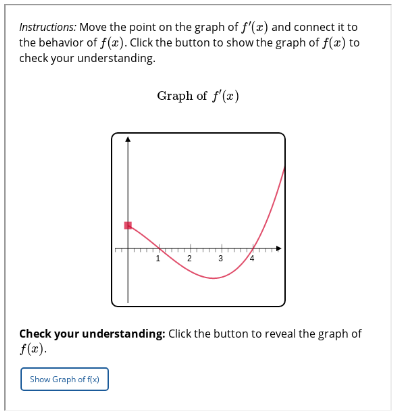
Ok, that's enough of this: let's get to the point and try to figure out how to actually calculate these areas without relying on our functions being "nice" enough that we can use geometry!

5.4 The Fundamental Theorem of Calculus

Let's remind ourselves of how we interpret derivatives. We are going to repeat a task that we did in Activity 4.2.3 First Derivative Test Graphically. It should feel familiar, which is good! We're going to use the intuition to make the big connection we've been forecasting so far.

Activity 5.4.1 Interpreting the Graph of a Derivarive.

Let's look at a picture of a graph of the first derivative, $f'(x)$, and try to get some information about $f(x)$ from it. Use the following graph of $f'(x)$, the first derivative, to answer the questions about $f(x)$.



Standalone
Embed

Since we don't have a huge amount of detail, you'll likely have to estimate the x-values for intervals and points in the following questions, but that's ok! Estimate away! Just make sure you know what you're looking for in the graph of $f'(x)$ to answer these questions.

(a) List the intervals on which $f(x)$ is increasing. What about decreasing?

(b) Find the x-values of any local maximums and/or local minimums of $f(x)$.

(c) List the intervals on which $f(x)$ is concave up. What about concave down?

(d) Find the x-values of any inflection points of $f(x)$.

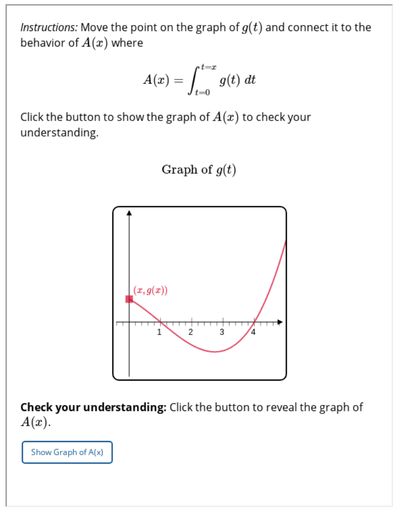
Areas and Antiderivatives

Activity 5.4.2 Interpreting Area.

First, we're going to define a bit of a weird function. Sometimes it's called the **Area function**:

$$A(x) = \int_{t=0}^{t=x} g(t)\ dt.$$

This is a strange function, because we're defining the function as an integral of another function. Specifically, note that the *input* for our area function $A(x)$ is the ending limit of integration: we're calculating the signed area "under" the curve of $g(t)$ from $t = 0$ up to some variable ending point $t = x$.
We can visualize this function by looking at the areas we create as we change x. For now, get used to just seeing the area "under" g when we move the point around. The areas themselves are the outputs of the function $A(x)$.



Standalone
Embed

Now we can think about this area function, and try to connect it to the graph of $g(t)$.

(a) List the intervals on which $A(x)$ is increasing. What about decreasing?

(b) Find the x-values of any local maximums and/or local minimums of $A(x)$.

(c) List the intervals on which $A(x)$ is concave up. What about concave down?

(d) Find the x-values of any inflection points of $A(x)$.

(e) Compare your answers here to your answers about the behavior of $f(x)$ based on the (same) graph of $f'(x)$ in Activity 5.4.1.

What does this mean about the connection between areas and derivatives, or areas and antiderivatives?

There it is! The way that we can interpret antiderivatives of functions! We found that the derivative of the function that tells us the signed area trapped between a curve and the x-axis between a fixed starting point and a variable ending point is the curve itself.

Another way of saying this, though, is that the function that tells us the signed area trapped between a curve and the x-axis between a fixed starting point and a variable ending point is an antiderivative of the curve itself! This is the Fundamental Theorem of Calculus, or at least half of it.

Theorem 5.4.1 Fundamental Theorem of Calculus (Part 1).

For a function f that is continuous on an interval $[a, b]$, and a function $A(x) = \displaystyle\int_{t=a}^{t=x} f(t)\, dt$ defined for x-values in $[a, b]$, then $A'(x) = f(x)$. That is:

$$\frac{d}{dx}\left(\int_{t=a}^{t=x} f(t)\, dt\right) = f(x).$$

Proof.

The proof of this theorem is one of the most delightful proofs we'll see. This is a "connector" theorem: a theorem that brings together several big ideas or objects from one common area of math and links them together. Let's enjoy the proof together.

Let $f(t)$ be a function that is continuous on the interval $a \leq t \leq b$. Then, we'll define the area function as $A(x) = \displaystyle\int_{t=a}^{t=x} f(t)\, dt$ for $a \leq x \leq b$. We are interested in $A'(x)$.

From Definition 2.1.2, we know:

$$A'(x) = \lim_{\Delta x \to 0} \left(\frac{A(x+\Delta x) - A(x)}{\Delta x} \right)$$

If we just focus on the numerator, $A(x+\Delta x) - A(x)$, we have:

$$\begin{aligned} A(x+\Delta x) - A(x) &= \left(\int_{t=a}^{t=x+\Delta x} f(t)\, dt \right) - \left(\int_{t=a}^{t=x} f(t)\, dt \right) \\ &= \int_{t=x}^{t=x+\Delta x} f(t)\, dt \end{aligned}$$

Let's approximate this integral with a Riemann sum with $n = 1$ rectangle.

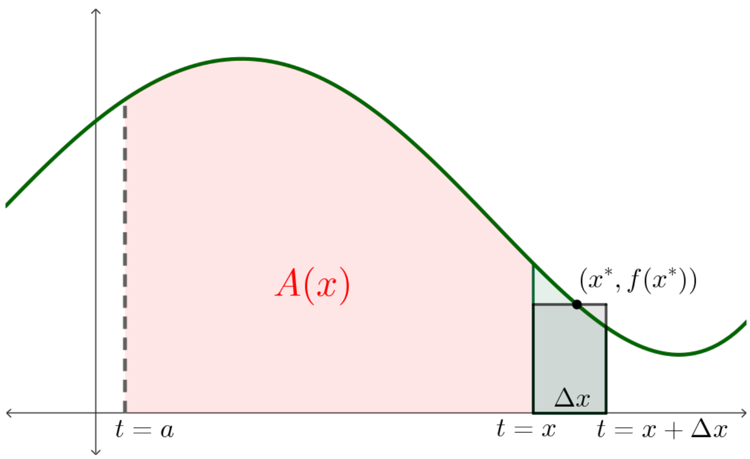


Figure 5.4.2

The total width of our interval is Δx, so we have that

$$\int_{t=x}^{t=x+\Delta x} f(t)\, dt \approx f(x^*)\Delta x$$

where x^* is some x-value in $[x, x+\Delta x]$. Note that we don't have a sum, as we normally would, since we are only "adding" a single area of a single rectangle.

This is only an approximation of the difference $A(x+\Delta x) - A(x)$, and so we can say, for small values of Δx,

$$\begin{aligned} A'(x) &\approx \left(\frac{A(x+\Delta x) - A(x)}{\Delta x} \right) \\ A'(x) &\approx \left(\frac{f(x^*)\Delta x}{\Delta x} \right) \\ A'(x) &\approx f(x^*) \end{aligned}$$

All that is left to do is to convince ourselves of two facts:

1. This approximation gets better as Δx gets smaller, and as $\Delta x \to 0$ we have $A'(x) \to f(x^*)$.

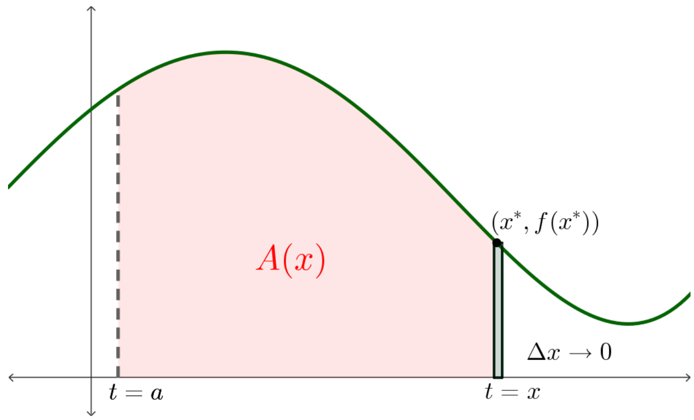


Figure 5.4.3

2. As $\Delta x \to 0$, the options for x^* in $[x, x + \Delta x]$ reduce to just x, since the interval collapses towards the single value. So as $\Delta x \to 0$, we have $x^* \to x$.

To be convinced that $A'(x) \to f(x^*)$, we just have to rely on the fact that, while our Riemann sum only has $n = 1$ rectangle, as $\Delta x \to 0$ the width(s) of "all" of our rectangles (our only one) approach 0, and so we end up with the definition of a definite integral in the limit:

$$\begin{aligned}\lim_{\Delta x \to 0} f(x^*)\Delta x &= \lim_{\Delta x \to 0} \sum_{k=1}^{1} f(x_k^*)\Delta x \\ &= \lim_{\Delta x \to 0} \int_{t=x}^{t=x+\Delta x} f(t)\ dt\end{aligned}$$

Hopefully it is easy to see that $x^* \to x$, since $[x, x + \Delta x]$ collapses on x.

Once we are convinced of these two facts, then it is clear that $A'(x) = f(x)$, since:

$$\begin{aligned}A'(x) &= \lim_{\Delta x \to 0} \left(\frac{A(x + \Delta x) - A(x)}{\Delta x} \right) \\ &\lim_{\Delta x \to 0} \left(\frac{f(x^*)\Delta x}{\Delta x} \right) \\ &= \lim_{\Delta x \to 0} f(x^*) \\ &= f(x)\end{aligned}$$

This completes the proof! Most of the proofs that you might see for this theorem use the Mean Value Theorem to help, since we can see a connection between the derivative $A'(x)$ and the average rate of change of the area function:

$$\left(\frac{A(x + \Delta x) - A(x)}{\Delta x} \right)$$

The Mean Value Theorem really is behind many of the most important results in calculus!

This theorem is going to be the big result that we use to show how to actually evaluate an area, and so it is easy to think of it as purely support for a "more important" result coming next. But we should pause and think about what this result tells us.

What we've done here is come up with a way of:

1. Guaranteeing that every continuous function has an antiderivative family. We have found a function whose derivative is whatever continuous function we want!

2. Generating antiderivatives. Until now, we have had to rely on being able to recognize functions as derivatives of other things, or be able to

"undo" derivative rules. And this will continue to be an important way for us to antidifferentiate functions. But now we have a way of constructing antiderivatives, albeit weird looking ones—we are not yet used to thinking about a function that is defined as a definite integral with a variable ending point.

We will play with this idea more later (in Section 6.1), and so for now we will push forward towards our goal of evaluating a definite integral without directly calculating a limit of Riemann sums.

Evaluating Definite Integrals

Activity 5.4.3 Evaluating Areas and Antiderivatives.

In this short activity, we'll just collect information about antiderivatives and this new area function,

$$A(x) = \int_{t=a}^{t=x} f(t)\, dt$$

for a function $f(t)$ that is continuous on the interval $a \leq t \leq x$.
For our purposes in this activity, let's say that $f(x) = x + \cos(x)$.

(a) From the Fundamental Theorem of Calculus (Part 1), we know that $A(x)$ is an antiderivative of $f(x)$, since $A'(x) = f(x)$.

Write out the function $A(x)$, and then name/write out one *other* antiderivative of $f(x)$, some $F(x)$.

(b) We know that all of the antiderivatives of a function are connected to each other.

Describe the connection between $A(x)$ and your $F(x)$.

(c) What is the value of $A(a)$? What is the value of $F(a)$? How are they different from each other?

(d) What is the value of $A(b)$? What is the value of $F(b)$? How are they different from each other?

(e) What about the differences: $A(b)-A(a)$ compared to $F(b)-F(a)$?

Theorem 5.4.4 Fundamental Theorem of Calculus (Part 2).

For a function $f(x)$ continuous on the closed interval $[a, b]$ and some $F(x)$, an antiderivative of $f(x)$, then

$$\int_{x=a}^{x=b} f(x)\, dx = F(x)\Big|_{x=a}^{x=b} = F(b) - F(a).$$

The vertical bar means "evaluated," and $F(x)\Big|_{x=a}^{x=b}$ is typically read as "$F(x)$ evaluated from $x = a$ to $x = b$."

Phew, this was a lot! Let's sit back a bit and enjoy the fruits of all of this deep, mathematical thinking: we have a relatively straight-forward way of evaluating definite integrals!

1. Find an antiderivative of the integrand. (Any antiderivative will do, so we can just choose the one with 0 as the constant term!)

2. Evaluate that antiderivative at the end points of the interval we're integrating over, and subtract.

Example 5.4.5

Evaluate the following definite integrals. Interpret the answers.

(a) $\displaystyle\int_{x=0}^{x=2} \left(x^2+1\right)\, dx$

Solution.

$$\begin{aligned}
\int_{x=0}^{x=2} (x^2+1)\, dx &= \underbrace{\left(\frac{x^3}{3}+x\right)}_{F(x)}\Bigg|_{x=0}^{x=2} \\
&= \underbrace{\left(\frac{2^3}{3}+2\right)}_{F(2)} - \underbrace{\left(\frac{0^3}{3}+0\right)}_{F(0)} \\
&= \left(\frac{8}{3}+2\right) - (0) \\
&= \frac{14}{3}
\end{aligned}$$

This is the area we were approximating in Section 5.2!

(b) $\displaystyle\int_{x=0}^{x=2\pi} (\sin(x)-\cos(x))\, dx$

Solution.

$$\begin{aligned}
\int_{x=0}^{x=2\pi} (\sin(x)-\cos(x))\, dx &= (-\cos(x)-\sin(x))\Big|_{x=0}^{x=2\pi} \\
&= (-\cos(2\pi)-\sin(2\pi)) - (-\cos(0)-\sin(0)) \\
&= (-1-0) - (-1-0) \\
&= 0
\end{aligned}$$

Why is this area 0? What does that mean about the region trapped between $y=\sin(x)-\cos(x)$ and the x-axis between $x=0$ and $x=2\pi$?

(c) $\displaystyle\int_{x=1}^{x=4} \left(\sqrt{x}-e^x\right)\, dx$

Solution.

$$\begin{aligned}
\int_{x=1}^{x=4} \left(\sqrt{x}-e^x\right)\, dx &= \int_{x=1}^{x=4} \left(x^{1/2}-e^x\right)\, dx \\
&= \left(\frac{x^{3/2}}{3/2}-e^x\right)\Bigg|_{x=1}^{x=4} \\
&= \left(\frac{2(4)^{3/2}}{3}-e^4\right) - \left(\frac{2(1)^{3/2}}{3}-e^1\right)
\end{aligned}$$

$$= \frac{14}{3} - e^4 + e$$

This value is $\frac{14}{3} - e^4 + e \approx -47.21$. Why is this value negative? What does that mean about the region we're looking at, and the function we're looking at?

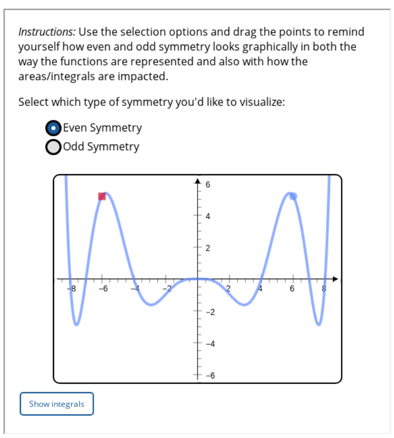
5.5 More Results about Definite Integrals

We'll end this chapter by looking a bit more closely at definite integrals and pulling a couple of small results out of our understanding of them, as well as some prior knowledge.

Symmetry

Activity 5.5.1 Symmetry in Functions and Integrals.

First, let's take a moment to remind ourselves (or see for the first time) what two types of "symmetry" we'll be considering. We call them "even" and "odd" symmetry, but sometimes we think of them as a "reflective" symmetry and a "rotational" symmetry in the graphs of our functions.



Standalone
Embed

(a) Convince yourself that you know what we mean when we say that a function is **even symmetric** on an interval if $f(-x) = f(x)$ on the interval.

Similarly, convince yourself that you know what we mean when we say that a function is **odd symmetric** on an interval if $f(-x) = -f(x)$ on the interval.

(b) Now let's think about areas. Before we visualize too much, let's start with a small question: How does the height of a function impact the area defined by a definite integral? It should be helpful to think about Riemann sums and areas of rectangles here.

The important question then, is how does a function being even or odd symmetric tell us information about areas defined by definite integrals of that function?

Theorem 5.5.1 Definite Integrals of Symmetric Functions.

If $f(x)$ is a continuous function on $[-a, a]$ for some real number $a > 0$, then:

- *If $f(x)$ is even symmetric on $[-a, a]$, then:*

$$\int_{x=-a}^{x=0} f(x)\, dx = \int_{x=0}^{x=a} f(x)\, dx.$$

- *If $f(x)$ is odd symmetric on $[-a, a]$, then:*

$$\int_{x=-a}^{x=0} f(x)\, dx = -\int_{x=0}^{x=a} f(x)\, dx.$$

Activity 5.5.2 Connecting Symmetric Integrals.

We're going to do some sketching here, and I want you to be clear about something: your sketches can be absolutely terrible. It's ok! They just need to embody the kind of symmetry we're talking about. You will probably sketch something and notice that your areas aren't to scale (or maybe even the wrong sign!), and that's fine.
It might be helpful to practice sketching graphs accurately, but don't worry if that part is a struggle.

(a) Sketch a function $f(x)$ with the following properties:

- $f(x)$ is *even symmetric* on the interval $[-6, 6]$
- $\displaystyle\int_{x=0}^{x=6} f(x)\, dx = 4$
- $\displaystyle\int_{x=-6}^{x=-2} f(x)\, dx = -1$

(b) Find the values of the following integrals:

- $\displaystyle\int_{x=0}^{x=2} f(x)\, dx$
- $\displaystyle\int_{x=-6}^{x=6} f(x)\, dx$

(c) Sketch a function $g(x)$ with the following properties:

- $g(x)$ is *odd symmetric* on the interval $[-9, 9]$
- $\displaystyle\int_{x=0}^{x=4} g(x)\, dx = 5$
- $\displaystyle\int_{x=-9}^{x=0} g(x)\, dx = 2$

(d) Find the values of the following integrals:

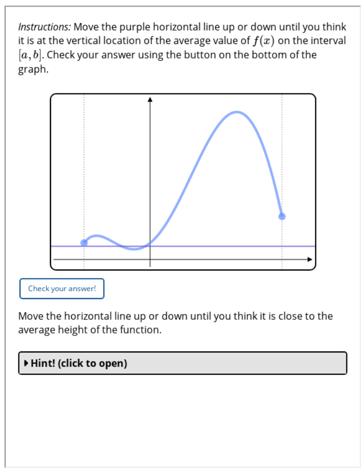
- $\displaystyle\int_{x=-9}^{x=-4} g(x)\, dx$
- $\displaystyle\int_{x=-4}^{x=9} g(x)\, dx$

Average Value of a Function

Activity 5.5.3 Visualizing the Average Height of a Function.

We are going to build a formula to find the "average height" or "average value" of a function $f(x)$ on the interval $[a, b]$. We're going to look at a function and try to find the average height. Along the way, we'll think a bit about areas!

(a) Consider the following function. Find the average height of the function on the interval pictured!



Standalone
Embed

(b) How does the area "under" the curve $f(x)$ on the interval compare to the area of the rectangle formed by the average height line?

(c) How do you define the two areas?

(d) Set up an equation connecting the two areas, and solve for the average height of $f(x)$.

Theorem 5.5.2 Average Value of a Function.

If a function $f(x)$ is continuous on the interval $[a, b]$, then the average value of $f(x)$ on $[a, b]$ is:

$$\frac{\displaystyle\int_{x=a}^{x=b} f(x)\, dx}{b-a}.$$

5.6 Introduction to u-Substitution

We have spent some time thinking about integrals, both indefinite and definite integrals, and what they represent. Let's end this discussion of integration with some classifications of indefinite integrals.

This classification is completely unserious, but maybe helpful. Let's say that we had to classify each of the following integrals as Easy, Medium, or Hard. Again, these are completely ambiguous and not at all defined, but I hope that we can see the spirit of what we're thinking of. Here are the three integrals that we want to classify:

$$\int x^5\,dx \qquad \int \sec^2(x)\,dx \qquad \int \frac{x^3}{1+x^2}\,dx$$

Which one could have the Easy? What about a Medium difficulty integral? What makes this one a bit harder, but not so hard to be classified as Hard? Can you even solve the Hard right now?

Let's think about a legitimately difficult integral, but one that we *can* actually think about.

Activity 5.6.1 A Hard Integral.

We're going to be thinking about two integrals here, but before we do, we should remind ourselves about how we can "re-phrase" an integration question.

If we are asked to find $\int f(x)\,dx$, then we are really being asked to find some function $F(x)$ whose derivative is $f(x)$. Of course, we're actually being asked to find *all* of the possible functions that fit this requirement, but we know that the constant of integration covers all of the differences.

This means that we can (and should?) check our answers pretty consistently: just find a derivative and check that it matches the integrand!

(a) Find $\int 3x^2 e^{x^3}\,dx$.

(b) Explain the role of $3x^2$ in the integral. Explain why the positioning of x^3 matters.

(c) Let's try another one.

$$\int 7x^6 \cos(x^7)\,dx$$

(d) How can you tell, in general, that the derivative you're looking at is one that was produced through the Chain Rule and not the Product Rule?

Undoing the Chain Rule

Let's try to formalize this process of "undoing the Chain Rule" that we noticed in Activity 5.6.1. It might be helpful to, first, think about the Chain Rule and how the differentiation process works. Let's look at the differentiating $\sin(x^7)$ using the Chain Rule.

The first thing we should do when finding $\frac{d}{dx}\left(\sin(x^7)\right)$ is recognize and identify the composition. We might call x^7 the "inside" function, or re-label it as u. Then we know that the Chain Rule will tell us to differentiate the "outside" function with regard to the "inside" function (or u if we re-label things) and then multiply by the derivative of the inside function (or $\frac{du}{dx}$).

$$\begin{aligned}\frac{d}{dx}\left(\sin(\underbrace{x^7}_{u})\right) &= \frac{d}{dx}\left(\sin(u)\right)\\ &= \frac{d}{du}\left(\sin(u)\right)\left(\frac{du}{dx}\right)\end{aligned}$$

$$= \cos(u)\frac{d}{dx}\left(x^7\right)$$
$$= \cos(x^7)7x^6$$

From here, we might re-write things to make it look nicer (coefficients look weird when they're not in the front) and write:

$$\frac{d}{dx}\left(\sin(x^7)\right) = 7x^6\cos(x^7)$$

When we work through this process backwards, we'll need to identify the "inside" function, but also find the derivative of that "inside" function, $\frac{du}{dx}$. This derivative gets introduced in the Chain Rule, and so it will have to be picked out when we undo the Chain Rule.

Let's build this process and go through the integral $\int 7x^6\cos(x^7)\ dx$ from Activity 5.6.1.

Process for u-Substitution.

1. Identify an "inside" function and/or a "function-derivative pair." We'll label the "inside" function, or the "function" part of the "function-derivative pair," as u.

   Example: In our integral $\int 7x^6\cos(x^7)\ dx$ we can see the "inside" function is x^7, and also we have x^7 and $7x^6$ as the function-derivative pair. Let $u = x^7$.

2. Define the substitution for the differential, $du = u'\ dx$. We can think about this as a result from knowing that $\frac{du}{dx} = u'$, or by thinking that a small change in u corresponds with u' multiplied by a small change in dx.

   Example: In our integral $\int 7x^6\cos(x^7)\ dx$, we labeled $u = x^7$. This means that $du = 7x^6\ dx$, since this is the derivative of x^7.

3. Substitute! Re-write the integral, replacing the parts that you've labeled as u and du.

   Example: We can re-write our integral to make this a bit easier to see:

$$\begin{aligned}\int 7x^6\cos(x)\ dx &= \int \cos(\underbrace{x^7}_{u})\,\underbrace{7x^6\ dx}_{du}\\ &= \int \cos(u)\ du\end{aligned}$$

4. Antidifferentiate! We should have an integral that is "written in terms of u," and so we can antidifferentiate the function as if u was our input variable. Notice that what we've done with our substitution is to undo the "multiply by the derivative of the inside function" step of the Chain Rule. Now we can antidifferentiate the "outside" function!

   Example: $\int \cos(u)\ du = \sin(u) + C$

5. Substitute "back" to have our antiderivative family written in terms of our original input variable, x in this case. We'll replace u with whatever we had defined our substitution to be in step 1.

   Example: We defined $u = x^7$, so:

$$\sin(u) + C = \sin(x^7) + C$$

Let's practice this with some more examples.

Activity 5.6.2 Picking the Pieces of a Substitution.

We're going to look at three integrals. Instead of working through them one-at-a-time, we'll look at all three simultaneously, where we can practice identifying, substituting, and antidifferentiating all at the same time.

(a) Let's consider these three integrals:

- $\displaystyle\int \frac{3x^2+1}{(x^3+x-2)^2}\,dx$
- $\displaystyle\int \cos(x)\sqrt{\sin(x)}\,dx$
- $\displaystyle\int \frac{(\ln(x))^3}{x}\,dx$

For each of these integrals, identify the substitution: define u as some function of x.

(b) For each substitution, define $du = u'\,dx$.

(c) For each integral, use your substitution (for both u and the differential du) to re-write the integral.

(d) Antidifferentiate each integral, and then use your substitution to write each integral back in terms of x.

Let's try to explain a little bit of what is happening. This style of problem solving is really useful in mathematics, and shows up in many places.

The first time I saw Figure 5.6.1 was in a Differential Equations class. We were learning about Laplace transformations, a technique that is very useful for solving a variety of problems in the field of differential equations. My professor was explaining why and how Laplace transformations were so powerful, and drew a version of the figure I've included to explain u-substitution. It was so helpful for me to understand what was happening, but the most helpful thing was when the professor said, offhand, "But that's exactly the same type of thing that u-substitution does, too." So many things fell into place for me because of that comment! You are guaranteed to see different versions of this picture throughout this textbook, but you can also keep an eye out for this in different problem-solving techniques in mathematics.

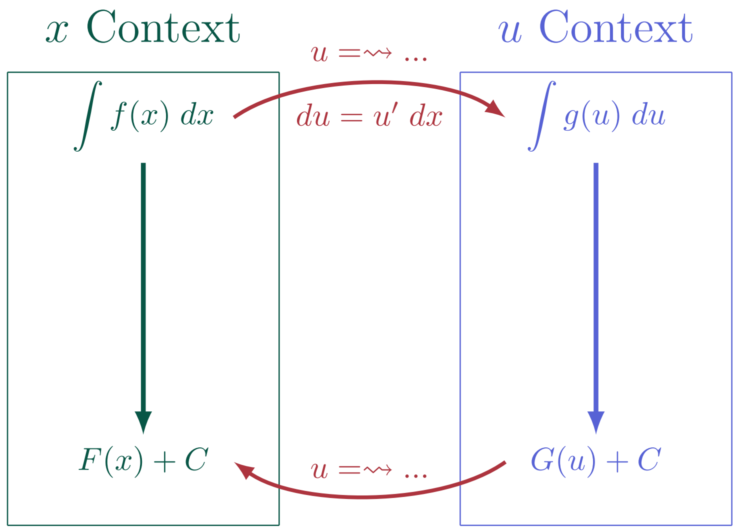


Figure 5.6.1 General idea of how a variable substitution in an integral works.

Let's explain what is happening in this picture. First, we typically are presented with integrals in some context. Our goal is to antidifferentiate. So for $\displaystyle\int f(x)\, dx$, we hope to antidifferentiate and end up with $F(x) + C$, the family of antiderivative of $f(x)$. These two things (the indefinite integral and the family of antiderivatives) exist in the same context (since they're defined with the same variable). We have spent some time moving directly from $f(x)\, dx$ to $F(x) + C$. But now we're seeing that this isn't always as direct of a path as we might wish: some integrals are *hard*.

In this case, we can try to identify what the problem is (in this case, composition) and find some transformation to apply to our integral. We choose a variable substitution, and we translate our integral to a different context (in this case, by writing it in terms of a different variable). In order for us to do this, we ned to define some substitution u, and then translate dx into du using the relationship $du = u'\, dx$.

Once we have this new integral, of a different function $g(u)$ in a different context with u acting as the input, we try to antidifferentiate again. If we pick our substitution carefully and we know what kind of problem we're trying to fix (in this case, getting around the composition), then this new integral in the u context could be "easier" to antidifferentiate. So our goal is to antidifferentiate the integral, but we can antidifferentiate it after translating it to a different context.

Once we have this antiderivative, we can translate that antiderivative family back to the original context (in this case, we write it in terms of x). We do this by utilizing the same translation or substitution that we defined earlier: we have something defined to link u and x. We can notice that, since the object we're translating is not an integral anymore, we do not have a differential to translate.

And there we have it: u-substitution works by identifying a problem that makes our task hard, translating our object to a friendlier context based on what we know about our problem, solving the problem in this friendlier context, and then finally translating the solution back to our original context.

Doesn't this feel like what we do with Logarithmic Differentiation? Use logs to translate our function to a friendlier form, differentiate in the log-context, and then try to translate back by solving for y'?

Let's practice this new strategy!

Activity 5.6.3 Compare Two Integrals.

Let's compare two integrals, and use this to build a more general strategy for performing u-substitution.

(a) Consider the following integral:

$$\int -4x^3 \sec^2(1 - x^4)\, dx$$

Select and justify a choice for u.

(b) Perform the u-substitution and antidifferentiate, and then substitute back to write your antiderivative in terms of x.

(c) Compare that integral to this one:

$$\int x^3 \sec^2(1 - x^4)\, dx$$

What is different about this new integral? What has remained the same? How does that impact your choice for u, or *does* it impact your choice for u?

(d) Has that changed what du should be?

(e) Ok so we've noticed an issue here. There are *plenty* of good ways of solving this problem, where du doesn't "show up" perfectly in our integral. In this case, we have that we're missing a necessary coefficient. We have the x^3 part, but we are missing the -4.

Try to re-write our integral with a -4 coefficient in there. We'll do that by multiplying the integrand function by 1, disguised as $\frac{-4}{-4}$ or $\left(-\frac{1}{4}\right)(-4)$.

(f) Now we can use the same u-substitution as before, and integrate in a similar way! Notice, though, that we will retain the coefficient of $-\frac{1}{4}$.

(This should be reasonable: our integral is $-\frac{1}{4}$ of the original one, since our coefficient was 1 to the original's -4.)

Go ahead and integrate!

This gives us a general strategy that we can use: if we pick a u-substitution, but we cannot find du in our integral, we can try to manipulate the integrand function to find it! We normally do this by applying some operation and its inverse (like multiplying by -4 and dividing by it as well, as in Activity 5.6.3).

Fixing Coefficients for du.

If we choose to perform a u-substitution in the integral $\int f(x)\, dx$, but we require a coefficient k in our definition of du, we can "fix" the coefficient in our integral:

$$\frac{1}{k} \int k f(x)\, dx$$

This strategy works well for coefficients, since we can factor out the $\frac{1}{k}$ from the integral.

Substitution for Definite Integrals

How would we evaluate the following definite integral?

$$\int_{x=0}^{x=2} \left(\frac{x+1}{x^2+2x+1} \right) dx$$

We can think back to what the Fundamental Theorem of Calculus (Part 2) says about evaluating a definite integral. We need to do two things:

1. Find an antiderivative of our function, $F(x)$. Any antiderivative will do, and we often pick the one where the constant term is 0.
2. Evaluate the antiderivative at the end points of the interval and subtract: $F(b) - F(a)$.

So for us to evaluate this definite integral, we can split the work into two parts.

Antidifferentiate, then Evaluate

Part 1: Antidifferentiation. We can think about the function $f(x) = \dfrac{x+1}{x^2+2x+1}$ and find the family of antiderivatives. Then, we can disregard the constant term (by selecting the antiderivative where the constant is 0 for convenience).

So we'll use u-substitution on the integral $\displaystyle\int \left(\frac{x+1}{x^2+2x+1} \right) dx$.

We can use $u = x^2 + 2x + 1$, which gives us $du = (2x+2)\, dx$ or $du = 2(x+1)\, dx$.

$$\begin{aligned}
\int \left(\frac{x+1}{x^2+2x+1} \right) dx &= \frac{1}{2} \int \left(\frac{2(x+1)}{x^2+2x+1} \right) dx \\
&= \frac{1}{2} \int \frac{1}{u}\, du \\
&= \frac{1}{2} \ln|u| + C \\
&= \frac{1}{2} \ln|x^2+2x+1| + C
\end{aligned}$$

So let's choose $F(x) = \frac{1}{2} \ln|x^2+2x+1|$ as the antiderivative we'll use.

Part 2: Evaluate at the End Points. For the integral $\displaystyle\int_{x=0}^{x=2} \left(\frac{x+1}{x^2+2x+1} \right) dx$, our ending points are $x = 2$ and $x = 0$, so let's evaluate!

$$\begin{aligned}
F(2) &= \frac{1}{2} \ln|(2)^2 + 2(2) + 1| \\
&= \frac{1}{2} \ln(9)
\end{aligned}$$

$$
\begin{aligned}
F(0) &= \frac{1}{2}\ln\left|(0)^2 + 2(0) + 1\right| \\
&= \frac{1}{2}\ln(1) = 0 \\
F(2) - F(1) &= \frac{1}{2}\ln(9) - 0 \\
&= \frac{1}{2}\ln(9) \\
&= \ln(3)
\end{aligned}
$$

Great, so we have a way of evaluating this integral!

$$\int_{x=0}^{x=2} \left(\frac{x+1}{x^2+2x+1}\right)\, dx = \ln(3)$$

A More Wholistic Substitution

Hang on, wait.

When we substituted our integral, we were substituting the *indefinite* integral. What if we applied our substitution to the *definite* integral.

The only difference is the limits of integration (other than the interpretation of area vs. family of antiderivatives, of course). So let's substitute the limits of integration.

Consider the definite integral. Really think about it.

$$\int_{x=0}^{x=2} \left(\frac{x+1}{x^2+2x+1}\right)\, dx$$

Here, we label the limits of integration as x-values: $x = 0$ and $x = 2$.

Can't we use our substitution rule to find corresponding u-values? What happens then? Let's approach this definite integral using the same substitution: we will think of $u = x^2 + 2x + 1$ again. But now we can find corresponding values of u when $x = 0$ and $x = 2$. All we need to do is evaluate the formula for u at those x-values!

$$
\begin{aligned}
\int_{x=0}^{x=2} \left(\frac{x+1}{x+2x+1}\right) &= \frac{1}{2}\int_{x=0}^{x=2} \left(\frac{2(x+1)}{x^2+2x+1}\right)\, dx \\
&= \frac{1}{2}\int_{u=(0)^2+2(0)+1}^{u=(2)^2+2(2)+1} \frac{1}{u}\, du \\
&= \frac{1}{2}\int_{u=1}^{u=9} \frac{1}{u}\, du \\
&\left(\frac{1}{2}\ln|u|\right)\Bigg|_{u=1}^{u=9} \\
&= \frac{1}{2}\ln(9) - \frac{1}{2}\ln(1) \\
&= \frac{1}{2}\ln(9) \\
&= \ln(3)
\end{aligned}
$$

So notice that we end up with the same thing here...we can substitute the limits of integration, and this matches the same value that we would get when we evaluate our antiderivatives at the endpoints of the x-interval.

We can amend our picture from Figure 5.6.1 to include definite integrals: in this case, we can evaluate the definite integral in either context, as long as we translate the limits of integration as well.

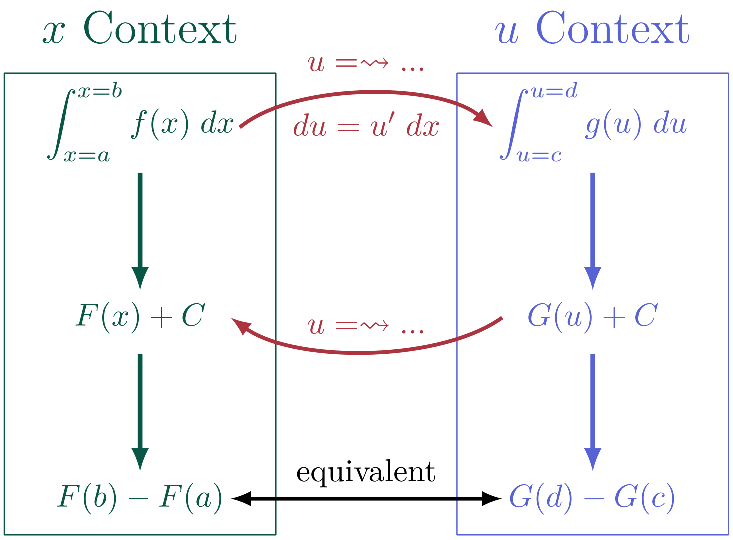
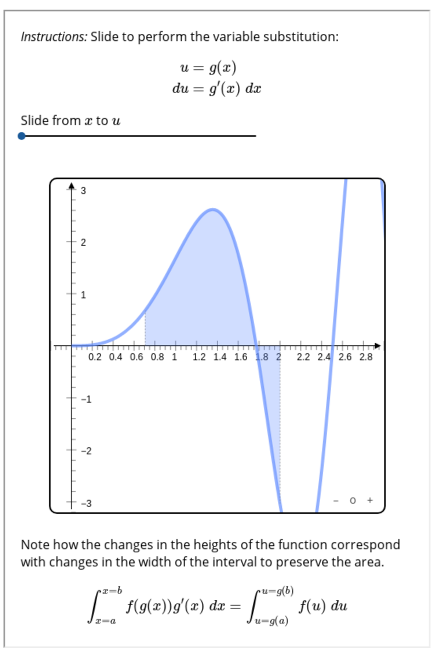


Figure 5.6.2 General idea of how a variable substitution in a definite integral works.

To see a visualization of what is happening, we can look below: move the slider to see the continuous deformation of the integral as we apply the variable substitution:



Standalone
Embed

Something we can say is that, since the area doesn't change when we do a variable substitution, then area is **invariant** under the transformation we're applying with that variable substitution.

More to Translate

There are more little tricks and nuances that we can think about with u-substitution: in general, this is an extremely flexible integration technique that we'll use in a variety of ways. For now, let's leave things off with one more interesting example.

In this example, we'll see a similar kind of issue to the one we saw in Activity 5.6.3: when we pick our substitution, there will be some issues "finding" du.

Example 5.6.3

Integrate the following, making sure to translate the whole integrand function to be written in terms of u.

$$\int \left(\frac{x^3}{\sqrt{x^2+1}} \right) \, dx$$

Hint 1. First, notice that $u = x^2 + 1$ is a great choice: we really want to focus on that composition. If this is the case, though, then $du = 2x \, dx$.

Hint 2. We can write x^3 as $x^2 \cdot x$, or if you *really* want to, we can write it as $\frac{1}{2}x^2 \cdot (2x)$

Hint 3. Our u-substitution formula can be written in a whole bunch of different ways!

$$\begin{aligned} u &= x^2 + 1 \\ x^2 &= u - 1 \\ x &= \pm\sqrt{u-1} \end{aligned}$$

These are all equivalent, but the first two might be the most helpful:

- Anywhere in our integral that we can see an x^2+1, we can replace that with u.
- We can also replace any *extra* x^2 pieces with $u - 1$!

Solution.

$$\int \left(\frac{x^3}{\sqrt{x^2+1}} \right) \, dx u = x^2 + 1$$

$$du = 2x \, dx$$

$$\begin{aligned} \int \left(\frac{x^3}{\sqrt{x^2+1}} \right) \, dx &= \frac{1}{2} \int \left(\frac{x^2 \cdot (2x)}{\sqrt{x^2+1}} \right) \, dx \\ &= \frac{1}{2} \int \frac{(u-1)}{\sqrt{u}} \, du && u = x^2 + 1 \leftrightarrow x^2 = u - 1 \\ &= \frac{1}{2} \int \frac{u}{\sqrt{u}} - \frac{1}{\sqrt{u}} \, du \\ &= \frac{1}{2} \int u^{1/2} - u^{-1/2} \, du \\ &= \frac{1}{2} \left(\frac{2u^{3/2}}{3} - 2u^{1/2} \right) + C \\ &= \frac{(x^2+1)^{3/2}}{3} - \sqrt{x^2+1} + C \end{aligned}$$

We'll spend more time thinking about u-substitution (as well as other variable substitutions) later on in Chapter 7. For now, this is a good stopping point, and should give us enough of a handle on u-substitution to integrate some difficult integrals!

Appendix A

Carnation Letter

Carnation

Corporate Offices

5045 Wilshire Boulevard
Los Angeles, California 90036
Telephone (213) 932 6000

April 13, 1987

[redacted]
Professor of Mathematics
Dept. of Mathematics
[redacted]

Dear Professor [redacted]

We appreciate the interest you expressed in examining the height-to-diameter relationship of containers used in our food products. A 1:1 ratio of height versus diameter is the most efficient use of material, if only the surface area of material is considered. However, there are many other factors which must be considered when designing a can for a particular product. Listed below are some of these other factors:

1) Thermal Processing — There is an inverse relationship between the most efficient design for cans relative to surface area and the amount of processing time required to sterilize the product contained within. In other words, a tall thin can or short wide can will require considerably less processing time and energy to achieve commercial sterility than a can which is nearly equal in height and diameter.

2) Strength Requirements — During thermal processing, considerable internal pressure develops. This pressure can cause the ends of the can to become permanently distorted. Because of this, ends on most cans are made of metal which is substantially thicker than that used in the can cylinder. Therefore, there is not a simple cost-to-surface area relationship relative to metal. As this can becomes taller and the end becomes smaller, thinner metal can be used in both the cylinder and the ends.

3) Can Manufacturing Line Changeover Time — Virtually all can lines run a variety of can sizes. The time required to change over from one can size to another is considerably less if only can height is changed, rather than height and diameter. In addition, since the same ends can be used if only the height is changed, the machinery used to manufacture ends does not have to be changed over to a different diameter. Reduced changeover time translates into reduced downtime and increased line efficiency.

4) Scrap Loss — Generally, more metal scrap is generated as the diameter is increased.

5) Warehouse and Shipping Efficiency — Smaller diameter cans make more efficient use of packaging and shipping space.

As you can see, cost and efficiency of a container are related to factors other than just the amount of material used. These are just a few of the factors which must be taken into consideration when designing a can. We hope that you now better understand that container design is not quite as simple as minimizing surface area.

Once again, thank you for your genuine interest.

Sincerely,

[redacted]

Vince [illegible]
Assistant Product Manager
Friskies Buffet

Figure A.0.1 Response letter from Carnation.

Full Text of the Carnation Letter. April 13, 1987

[REDACTED]
Professor of Mathematics
Dept. of Mathematics
[REDACTED]

Dear Professor [REDACTED],

We appreciate the interest you expressed in examining the height-to-diameter relationship of containers used in our food products. A 1:1 ratio of height versus diameter is the most efficient use of material, if only the surface area of material is considered. However, there are many other factors which must be considered when designing a can for a particular product. Listed below are some of these other factors:

1) Thermal Processing — There is an inverse relationship between the most efficient design for cans relative to surface area and the amount of processing time required to sterilize the product contained within. In other words, a tall thin can or short wide can will require considerably less processing time and energy to achieve commercial sterility than a can which is nearly equal in height and diameter.

2) Strength Requirements — During thermal processing, considerable internal pressure develops. This pressure can cause the ends of the can to become permanently distorted. Because of this, ends on most cans are made of metal which is substantially thicker than that used in the can cylinder. Therefore, there is not a simple cost-to-surface area relationship relative to metal. As this can becomes taller and the end becomes smaller, thinner metal can be used in both the cylinder and the ends.

3) Can Manufacturing Line Changeover Time — Virtually all can lines run a variety of can sizes. The time required to change over from one can size to another is considerably less if only can height is changed, rather than height and diameter. In addition. since the same ends can be used if only the height is changed, the machinery used to manufacture ends does not have to be changed over to a different diameter. Reduced changeover time translates into reduced downtime and increased line efficiency.

4) Scrap Loss — Generally, more metal scrap is generated as the diameter is increased.

5) Warehouse and Shipping Efficiency — Smaller diameter cans make more efficient use of packaging and shipping space.

As you can see, cost and efficiency of a container are related to factors other than just the amount of material used. These are just a few of the factors which must be taken into consideration when designing a can. We hope that you now better understand that container design is not quite as simple as minimizing surface area.

Once again, thank you for your genuine interest.

Sincerely,
[REDACTED]
Vince [Illegible]
Assistant Product Manager
Friskies Buffet

Colophon

This book was authored in PreTeXt.