1 Functions and Limits

1.1 Quick Review Facts

Functions

Recall that a *function* is a rule that takes an input and assigns a specific output. Note that a function always gives exactly one output, and always gives the same output for a given input. Here we remember some facts about common functions.

Polynomials: You should remember the quadratic formula, which says that if $ax^2 + bx + c = 0$ then

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

It is also useful to recall that

- $(a+b)^2 = a^2 + 2ab + b^2$
- $(a+b)(a-b) = a^2 - b^2$
- $(a^2 + ab + b^2)(a-b) = a^3 - b^3$.

Rational functions are the ratio of two polynomials.

Trigonometric functions: In this course we will *always* use radians, because they are unitless and thus easier to track (especially when using the chain rule). Useful facts include:

- The most important trigonometric identity, and really the only one you probably need to remember, is $\cos^2(x) + \sin^2(x) = 1$.
- From this you can derive the fact that $1 + \tan^2(x) = \sec^2(x)$.
- $\sin(-x) = -\sin(x)$. We call functions like this "odd".
- $\cos(-x) = \cos(x)$. We call functions like this "even."
- $\sin(x + \pi/2) = \sin(\pi/2 - x) = \cos(x)$
- A fact that we will probably use exactly twice is the sum of angles formula for sine: $\sin(x+y) = \sin(x)\cos(y) + \cos(x)\sin(y)$.
- Similarly, $\cos(x+y) = \cos(x)\cos(y) + \sin(x)\sin(y)$

Set and interval notation

We write $\{x : \text{condition}\}$ to represent the set of all numbers x that satisfy some condition. We will sometimes write $\mathbb{R}$ to refer to all the real numbers. We will also refer to various intervals:

$(a,b) = \{x : a < x < b\}$	open interval	$[a,b] = \{x : a \leq x \leq b\}$	closed interval
$[a,b) = \{x : a \leq x < b\}$	half-open interval	$(a,b] = \{x : a < x \leq b\}$	half-open interval

1.2 Review of functions

Definition 1.1. A *function* is a rule that takes an input and assigns a specific output. Note that a function always gives exactly one output, and always gives the same output for a given input.

In the abstract, a function can take any type of input and give any type of output. In this class we will primarily study functions whose inputs and outputs are all real numbers.

Definition 1.2. The *domain* of a function is the set of possible valid inputs. The *range* or *image* is the set of possible outputs.

Example 1.3.

1. The function $f(x) = x^2$ has all real numbers in its domain, and its image is the set of non-negative real numbers.
2. The function $f(x) = \sqrt{x}$ has all non-negative real numbers as its domain, and non-negative real numbers as its image.
3. The function $f(x) = \frac{1}{x^2-1}$ has all real numbers except 1 and -1 in its domain, and all real numbers greater than zero or less than or equal to -1 in its image. We can write this set as $\{x : x > 0 \text{ or } x \leq -1\}$, or equivalently as $\{x : x > 0\} \cup \{x : x \leq -1\}$ or $(-\infty, -1] \cup (0, +\infty)$.

Remark 1.4. The word "range" is sometimes used to refer to the type of output a function can have; in this context people also use the word "codomain". In this class we will always use "range" to refer to an output a function can actually produce.

Functions can be described many ways: a verbal description, an algebraic rule, a graph, or a list of possible inputs and the corresponding outputs.

Example 1.5. What are the domain and range of $f(x) = x^3$?

The domain of the function is all real numbers, since we can cube any number. Less obviously, the range is also all reals: if we cube a negative number, we get a negative number, and if we cube a positive number we get a positive number.

Example 1.6. What are the domain and range of $\frac{1}{x-1}$?

The domain is all reals except 1, because we can't divide by zero. (In general, the domain is often "everywhere nothing goes wrong.") The image is all reals except 0, since we can divide 1 by any number except 0 and thus get the reciprocal of any non-zero number.

In other notation, the domain is $\{x : x \neq 1\}$ and the range is $\{x : x \neq 0\}$.

Definition 1.7. A *piecewise function* is a function defined by breaking its domain up into pieces and giving a rule for each piece.

Example 1.8. 1.

$$f(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

is a piecewise function, given by the rule "If the input is negative, the output is zero; otherwise the output is 1." The domain is all reals and whose range is $\{0, 1\}$.

2.

$$g(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x \geq 0 \end{cases}$$

is not a function because it does not give a clear output when given 0 as input.

3.

$$h(x) = \begin{cases} x^2 + 1 & x < 0 \\ 3x - 2 & x > 0 \end{cases}$$

is a piecewise function whose domain does not include 0. The domain is $\{x : x \neq 0\}$ and the range is $(-2, +\infty)$.

4.

$$f(x) = \begin{cases} x + 2 & x \geq 1 \\ x^2 + 2 & x \leq 1 \end{cases}$$

This function might concern you since it appears to have two values for 1; but after looking a bit more closely we see that both pieces define $f(x) = 3$ so we're okay. This is a function whose domain is all reals and whose image is $[2, +\infty)$.

1.2.1 Function Catalog

We will now present a list of functions; we should be familiar with these functions, their graphs, and often their domains and images.

1. A *constant function* is given by $f(x) = c$ for some real number c. It's domain is all real numbers, and its range is the set with one point $\{c\}$.

2. A *linear function* is given by $f(x) = ax + b$. Its domain and range are both all real numbers.

3. A *polynomial function* is given by $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$, where n is some positive integer and the a_i are all real numbers. A polynomial is a sum of terms, where each term is some real number multiplied by x raised to a positive integer power.

   The domain of any polynomial is all real numbers.

   (3a) A *quadratic polynomial* is a polynomial whose highest term has exponent 2, given by $f(x) = ax^2 + bx + c$. It has image $\{x : x \geq C\}$ or $\{x : x \leq C\}$ for some real number C.

   It will be useful to recall the quadratic formula; if $f(x) = ax^2 + bx + c$ then $f(x) = 0$ precisely when
   $$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

   (3b) A *cubic polynomial* has 3 as its highest exponent, given by $f(x) = ax^3 + bx^2 + cx + d$. Its image is all real numbers.

4. A *rational function* is given by the ratio of two polynomial functions (note the similarity between "ratio" and "rational"). Thus a rational function is of the form
   $$f(x) = \frac{a_0 + a_1 x + \cdots + a_n x^n}{b_0 + b_1 x + \cdots + b_m x^m}.$$

   A rational function has domain all real numbers, except for the finite collection of points where the denominator is zero.

   Example 1.9.
   - $f(x) = \frac{x^2+1}{x-1}$ is a rational function with domain $\{x : x \neq 1\}$.
   - $g(x) = \frac{1}{x^4+7}$ is a rational function with domain all reals, since the denominator is never zero for any real number. (The range is $(0, 1/7]$).

5. The function
$$|x| = \begin{cases} x & x \geq 0 \\ -x & x \leq 0 \end{cases} = \sqrt{x^2}$$
is well-defined since both rules give the same output for 0. This function is called the *absolute value* of x. The piecewise definition is usually more useful. The domain is all reals, and the image is $[0, +\infty)$; in fact, the point of this function is to "sanitize" all your real number inputs into positive numbers.

We will now discuss the exponential functions.

1. The *n-th root function* is given by $f(x) = x^{1/n}$. The number $x^{1/n}$ is the unique positive number y such that $y^n = x$. If n is even then this function has all non-negative numbers in its domain and image; if n is odd then all real numbers are in the domain and image.
2. The *reciprocal function* is given by $f(x) = x^{-1} = \frac{1}{x}$. This function has domain and range $\{x : x \neq 0\}$. It also has the interesting property that $f(f(x)) = x$ for any $x \neq 0$; that is, applying the rule twice gets you back where you started.
3. We can define a general exponential function $f(x) = x^{m/n}$ where m and n are any integers by combining the previous two rules with the rules that
   - $x^a x^b = x^{a+b}$
   - $(x^a)^b = x^{ab}$
   - $x^a y^a = (xy)^a$

Example 1.10. If we wish to calculate $8^{-5/3}$, we can rewrite this as
$$(8^{5/3})^{-1} = ((8^{1/3})^5)^{-1} = (2^5)^{-1} = 32^{-1} = \frac{1}{32}.$$

Example 1.11. Compute $27^{-2/3}$.
$$27^{-2/3} = ((27^{1/3})^2)^{-1} = (3^2)^{-1} = 9^{-1} = \frac{1}{9}.$$

Example 1.12. What is the domain of $f(x) = \dfrac{x^2 - 4}{x^2 + 5x + 6}$?

The domain is all reals except where the denominator is zero. $x^2+5x+6 = (x+2)(x+3)$ is zero when $x = -2$ or $x = -3$, so the domain is $\{x : x \neq -2, -3\}$.

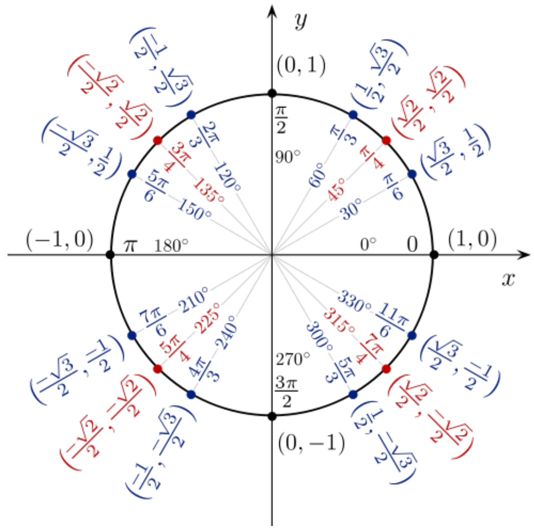


Figure 1.1: The Unit Circle

Now we discuss the trigonometric functions. In calculus we essentially always use radians. Recall that $\sin(x)$ and $\cos(x)$ are given by the unit circle: if we start from the point $(1, 0)$ and rotate x radians counterclockwise, then our x coordinate will be $\cos(x)$ and our y coordinate will be $\sin(x)$. We can also recall that if θ is the measure of a non-right angle of a right triangle, then $\sin(\theta)$ is the ratio of the length of the opposite side to the length of the hypotenuse, and $\cos(\theta)$ is the ratio of the length of the adjacent side to the length of the hypotenuse.

There is one important trigonometric identity we must remember, which is that $\sin^2(x) + \cos^2(x) = 1$; this is just the Pythagorean theorem applied to triangles with hypotenuse of length one.

We can see that sin and cos both have domain all reals, and image $[-1, 1]$.

We also have four other trigonometric functions:

1. $\tan(x) = \frac{\sin(x)}{\cos(x)}$ has domain $\{x : x \neq n\pi + \pi/2\}$ since the function isn't defined when $\cos(x) = 0$, and has image all reals.

2. $\cot(x) = \frac{\cos(x)}{\sin(x)}$ has domain $\{x : x \neq n\pi\}$ since the function isn't defined hwen $\sin(x) = 0$, and has image all reals.

3. $\sec(x) = \frac{1}{\cos(x)}$ has domain $\{x : x \neq n\pi + \pi/2\}$ and image $(-\infty, -1] \cup [1, +\infty)$.

4. $\csc(x) = \frac{1}{\sin(x)}$ has domain $\{x : x \neq n\pi\}$ and image $(-\infty, -1] \cup [1, +\infty)$.

The trigonometric functions also have a few important symmetries:

- $\sin(-x) = -\sin(x)$. Functions with this property are called *odd functions*.
- $\cos(-x) = \cos(x)$. Functions with this proprty are called *even functions*.
- $\sin(\pi/2 - x) = \cos(x)$. The sin function is a *reflection* of the cos function around the line $x = \pi/4$.
- $\sin(x + \pi/2) = \cos(x)$. The sin function is a *translation* of the cos function along the x axis.

This leads into our next topic, which is to ask how we can turn some functions into other functions.

1.2.2 Deriving functions from other functions

We can't possibly list every function we will ever use. Instead, let's talk about how to start with a few functions—the ones above—and use them to construct more functions.

Example 1.13. What must I do to graph A to get graph B?

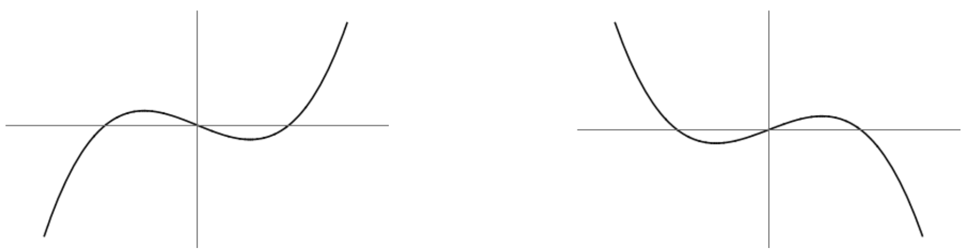


Figure 1.2: Left: graph A, Right: graph B

Example 1.14. What must I do to graph C to get graph D?

Figure 1.3: Left: graph C, Right: graph D

Now we can move on to the main event: various operations we can apply to a function to get a new function.

Assume that c is a positive real number.

We can *shift* the graph of a function up, down, left, or right:

- The graph of $y = f(x) + c$ is the graph of $y = f(x)$ shifted up by c units.
- The graph of $y = f(x) - c$ is the graph of $y = f(x)$ shifted down by c units.
- The graph of $y = f(x - c)$ is the graph of $y = f(x)$ shifted right by c units.
- The graph of $y = f(x + c)$ is the graph of $y = f(x)$ shifted left by c units.

Note the perhaps-counterintuitive directions on the last two.

Example 1.15. The first graph is the graph of x^2. What is the second graph?

Figure 1.4: The graphs of x^2 and $x^2 - 1$

Answer: $x^2 - 1$. (Since there's no axis labels, $x^2 - c$ would also be reasonable).

Example 1.16. What do I need to do to the graph of x^3 to get the graph of $(x + 3)^3$?

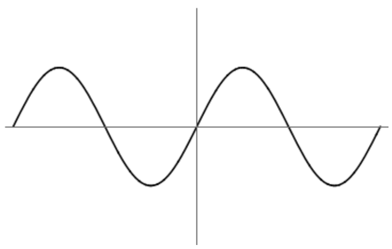
Figure 1.5: The graphs of x^3 and $(x + 3)^3$

Answer: shift it to the left by three units.

We can also *stretch* the graph of a function vertically or horizontally.

- The graph of $y = c \cdot f(x)$ is the graph of $y = f(x)$ stretched vertically by a factor of c. Note c can be less than one here, in which case the graph is shrunk.
- The graph of $y = f(x/c)$ is the graph of $y = f(x)$ stretched horizontally by a factor of c. Note again that c can be less than one, in which case the graph is shrunken.

Example 1.17. If I stretch the function $\sin(x)$ to be twice as tall, what function do I get?



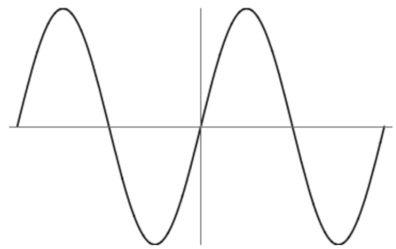
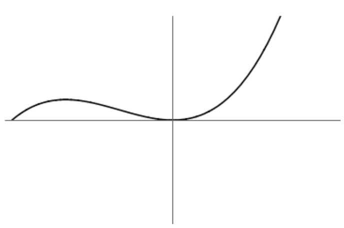


Figure 1.6: The graphs of $\sin(x)$ and $2\sin(x)$

We can also *reflect* a graph about the x axis or y axis (or, with a little creativity, some other axis).

- The graph of $y = -f(x)$ is the graph of $y = f(x)$ reflected about the x-axis, that is, flipped top-to-bottom.
- The graph of $y = f(-x)$ is the graph of $y = f(x)$ reflected about the y-axis, that is, flipped left-to-right.

Example 1.18. Here is an example of what a function looks like reflected.



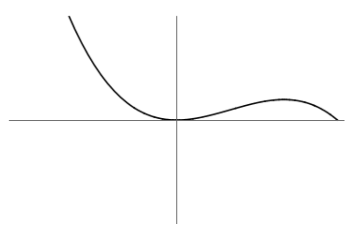


Figure 1.7: The graphs of $x^3 + 2x^2$ and $-x^3 + 2x^2$

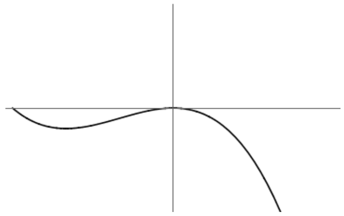


Figure 1.8: The graphs of $-x^3 - 2x^2$ and $x^3 - 2x^2$

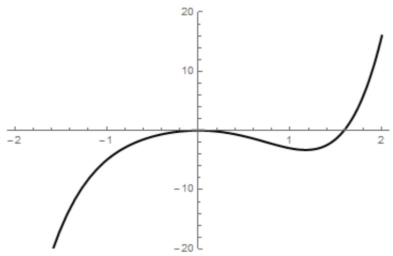


Figure 1.9: The graph of $x^5 - 4x^2$

Example 1.19. Figure 1.9 is the graph of $x^5 - 4x^2$. What would the graph of $(x+1)^5 - 4(x+1)^2$ look like? What would the graph of $(2x)^5 - 4(2x)^2$ look like?

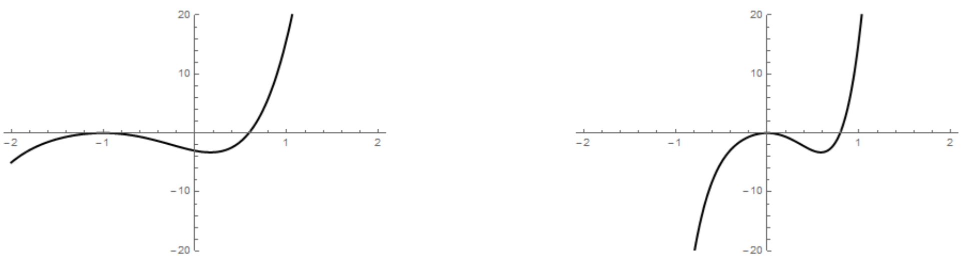


Figure 1.10: The graphs of $(x+1)^5 - 4(x+1)^2$ and $(2x)^5 - 4(2x)^2$

Example 1.20. Which of the functions $f(x) = x^2+1$, $f(x) = x^3+3$, $f(x) = x^4$, $f(x) = x^5+x$ is even?

Example 1.21. Which of the functions $f(x) = x^2+1$, $f(x) = x^3+3$, $f(x) = x^4$, $f(x) = x^5+x$ is odd?

In general a polynomial with only even-degree terms will be even, and a polynomial with only odd-degree terms is odd. (Hopefully this will be easy to remember!) A polynomial with both even-degree and odd-degree terms is generally neither even nor odd.

Finally, we can combine two functions.

- The function $f + g$ is defined by $(f + g)(x) = f(x) + g(x)$.
- The function $f \cdot g$ is defined by $(f \cdot g)(x) = f(x)g(x)$.
- The function $f \circ g$ is defined by $(f \circ g)(x) = f(g(x))$.

This last rule will be very important, and is called *composition of functions*. $f \circ g$ corresponds to putting our input into the function g, and then taking the output and feeding that into the function f. This only makes sense if the image of g is in the domain of f.

Remark 1.22. $f \circ g$ and $g \circ f$ are not the same thing. For instance, if $f(x) = x^2$ and $g(x) = x+1$, then $(f \circ g)(x) = f(x+1) = (x+1)^2 = x^2+2x+1$, but $(g \circ f)(x) = g(x^2) = x^2+1$.

Example 1.23. If $f(x) = \sqrt{x}$ and $g(x) = 3x^2$ then what is$(f \circ g)(x)$? What is the domain? What about $(g \circ f)(x)$?

$(f \circ g)(x) = \sqrt{3x^2}$. This is the same as $\sqrt{3}|x|$. The domain is all reals.

$(g \circ f)(x) = 3\sqrt{x}^2$. This is the same as $3|x|$ but the domain is only $[0, +\infty)$ since we can't plug a negative number into f.

Example 1.24. Can we write $x^2 + 1$ as the composition of two simple functions?

Answer: Let $f(x) = x^2$ and $g(x) = x + 1$. Then $g(f(x)) = x^2 + 1$

Can we write $\sqrt{x^3 - 1}$ as the composition of three simple functions?

Answer: Let $f(x) = x^3$, $g(x) = x - 1$, and $h(x) = \sqrt{x}$. Then $h(g(f(x))) = \sqrt{x^3 - 1}$.

1.3 Informal Continuity and Limits

Let's start with an easy question:

Question 1.25. What is the square root of four?

Everyone can probably tell me that the answer is "two". So now let's do a harder one:

Question 1.26. What is the square root of five?

Without a calculator, you probably can't tell me the answer. But you should be able to make a pretty good guess. Five close to four; so $\sqrt{5}$ should be close to two.

We call this sort of estimate a *zeroth-order approximation*. In a zeroth-order approximation, we only get to use one piece of information: the value of our function at a specific number. Then we use that information to estimate its value at nearby numbers.

We can only do so good a job with that limited amount of information, but we can still do a surprising amount.

Example 1.27. Suppose $f(1) = 36, f(2) = 35, f(3) = 38, f(4) = 38$. What can we say to estimate $f(5)$?

From looking at the data we have, it seems like $f(5)$ should be 38 or 39, probably. But it's actually 45. These are the low temperatures in Pasadena for the first five days of this year.

Often tomorrow's temperature will be similar to today's temperature. But there's no guarantee.

This example shows that we can't always do what we did with $\sqrt{5}$. Some functions jump around too much for this sort of approximation thing to work; values of similar inputs don't have similar outputs.

We don't like these functions, precisely because they're hard to think about or understand. So we're mostly going to look at functions that we *can* approximate effectively.

Definition 1.28 (Informal)**.** We say a function f is *continuous* at a number a if whenever x is close to a, then $f(x)$ is close to $f(a)$.

In other words, for a continuous function, when x and a are close together, then $f(a)$ is a decent approximation for $f(x)$.

Another way to think of this is that the function f is continuous at a if it doesn't "jump" at a.

There are a few different ways for a function to not be continuous at a given number. I will categorize these more carefully in a couple days, but right now I want to show you a few different things that can happen.

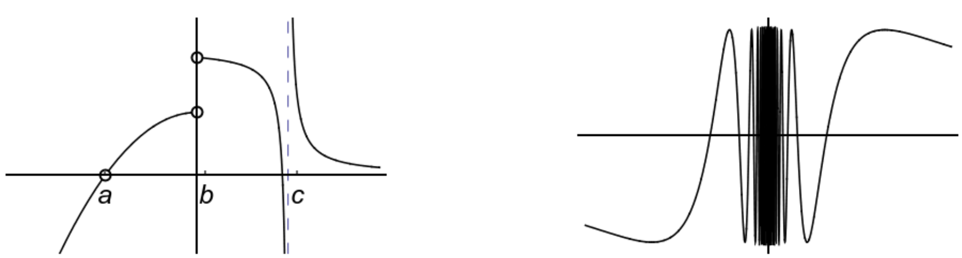


Figure 1.11: Left: a: removable discontinuity; b: jump discontinuity; c: infinite discontinuity. Right: bad discontinuity

Some functions get even worse than that. My two favorite discontinuous functions are:

$$T(x) = \begin{cases} 1/q & x = p/q \text{ rational} \\ 0 & x \text{ irrational} \end{cases} \qquad \chi(x) = \begin{cases} 1 & x \text{ rational} \\ 0 & x \text{ irrational} \end{cases}$$

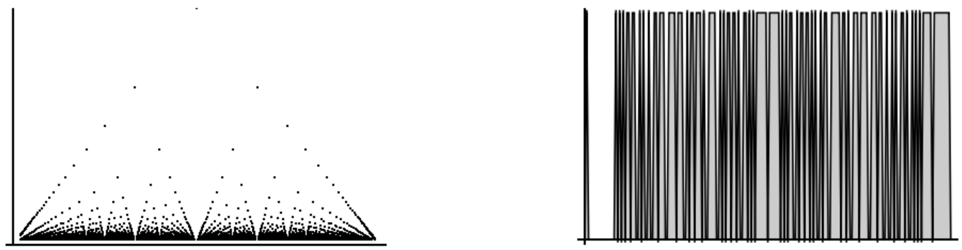


Figure 1.12: Left: $T(x)$ is really discontinuous. Right: $\chi(x)$ is really really discontinuous

In fact, in some sense "most functions" aren't at all continuous. If you found away to choose $f(x)$ completely at random for each real number x, you would get a spectacularly discontinuous function. But you would never actually be able to describe it sensibly.

But for the most part this isn't a problem. Most of the functions that we can easily describe are continuous most of the time. And so when approximating functions we don't understand, we often assume it's reasonably continuous.

Fact 1.29. *Any reasonable function given by a reasonable single formula is continuous at any number for which it is defined.*

In particular, any function composed of algebraic operations, polynomials, exponents, and trigonometric functions is continuous at every number in its domain.

If a function is continuous at every number in its domain, we just say that it is continuous. Note, importantly, that a continuous function doesn't have to be continuous at every real number.

Example 1.30. The function

$$f(x) = \frac{x^3 - 5x + 1}{(x-1)(x-2)(x-3)}$$

is "reasonable", so it is continuous. This means that it is continuous exactly on its domain, which is $\{x : x \neq 1, 2, 3\}$.

Example 1.31. Where is $\sqrt{1 + x^3}$ continuous?

Answer: Root functions are continuous on their domains. $1 + x^3 \geq 0$ when $x \geq -1$ so the function is continuous on its domain, $[-1, +\infty)$.

Remark 1.32. Sometimes we might also talk about functions that are "continuous from the right" at a. This means that $f(a)$ is a good approximation of $f(x)$ if x is close to a and also bigger than—and thus to the right of—a.

In order to understand continuity better, it's helpful to turn the question around and look at things from the opposite direction. (This is a trick that's often useful in math). So instead of asking whether we can estimate $f(x)$ given $f(a)$, we'll turn this around. If we know $f(x)$ for every x near a, what can we say about $f(a)$?

Definition 1.33. Suppose a is a real number, and f is a function which is defined for all x "near" the number a. We say "The *limit* of $f(x)$ as x approaches a is L," and we write

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

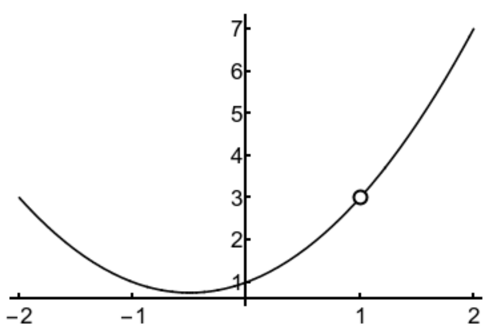
if we can make $f(x)$ get as close as we want to L by picking x that are very close to a.

Graphically, this means that if the x coordinate is near a then the y coordinate is near L. Pictorially, if you draw a small enough circle around the point $(a, 0)$ on the x-axis and look at the points of the graph above and below it, you can force all those points to be close to L.

Notice that we're trying to use knowing $f(x)$ to tell us what happens near a. So we specifically ignore the value of $f(a)$ even if we already know it.

Example 1.34. Let's consider the function $f(x) = \frac{x^3-1}{x-1}$. We can see the graph below. Notice that the function isn't defined at $a = 1$, so $f(1)$ is meaningless and we can't compute it.

But f is defined for all x near 1, so we can compute the limit. Looking at the graph and estimating suggests that when x gets close to 1, then $f(x)$ gets close to 3, and so we can say that $\lim_{x \to 1} f(x) = 3$.



That last example worked, but we basically just eyeballed it. We want a way to actually justify our claims. We can do that using two core principles. The first is what I call the Almost Identical Functions property.

Lemma 1.35 (Almost Identical Functions)**.** *If* $f(x) = g(x)$ *on some open interval* $(a-d, a+d)$ *surrounding* a*, except possibly at* a*, then* $\lim_{x \to a} f(x) = \lim_{x \to a} g(x)$ *whenever one limit exists.*

This tells us that two functions have the same limit at a if they have the same values near a. This makes sense, because the limit only depends on the values near a.

How does this help us? Ideally, we take a complicated function and replace it with a simpler function.

Example 1.36. Above, we looked at the function $f(x) = \frac{x^3-1}{x-1}$. You may know that we can factor the numerator; thus we in fact have $f(x) = \frac{(x-1)(x^2+x+1)}{x-1}$.

At this point you probably want to cancel the $x-1$ term on the top and the bottom. But in fact that would change the function! For $f(1)$ isn't defined. But the function $g(x) = x^2+x+1$ is perfectly well-defined at $a = 1$. Thus $f(1) \neq g(1)$, and so f and g can't be the same function.

However, they do give the same value if we plug in any number other than 1. If $y \neq 1$ then $y - 1 \neq 0$, so we have

$$f(y) = \frac{(y-1)(y^2+y+1)}{y-1} = y^2 + y + 1 = g(y).$$

Thus f and g aren't the same, but they are *almost* the same. So lemma 1.100 tells us that $\lim_{x\to 1} f(x) = \lim_{x\to 1} g(x)$.

However, this doesn't actually do everything we want it to do. We've replaced a complicated function $f(x) = \frac{x^3-1}{x-1}$ with a simpler function $g(x) = x^2 + x + 1$, but we still haven't figured out what to do with that function.

This leads to our second principle. We started off talking about continuous functions, and said that if f is continuous at a, then $f(a)$ is a good estimate for $f(x)$ when x is near to a. In other words, when x is near a then $f(x)$ is near $f(a)$—so $\lim_{x\to a} f(x) = f(a)$.

This really is the same as the less formal definition we gave at the beginning of this section. There, we said that f is continuous if $f(a)$ is a good approximation for $f(x)$; here we say that f is continuous if $f(x)$ is a good approximation for $f(a)$. This also clarifies *how good* the approximation needs to be. For f to be continuous, the approximation needs to get perfect as x gets close to a.

Example 1.37. The *Heaviside Function* or *step function* is given by

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

It is often used in electrical engineering applications to describe the current running through a switch before and after it has been flipped.

We can ask: what is $\lim_{x\to 0} H(x)$?

There isn't one: no matter how close x gets to 0, sometimes $H(x)$ will be 0 and sometimes it will be 1. So there is no one value that approximates $H(x)$ for any x near a.

However, the Heaviside function clearly behaves well if look only at one side or the other of it. And just as we could talk about continuity to one side or the other, we can talk about *one-sided limits*.

Definition 1.38. Suppose a is a real number, and f is a function which is defined for all $x < a$ that are "near" the number a. We say "The limit of $f(x)$ as x approaches a from the left is L," and we write

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

if we can make $f(x)$ get as close as we want to L by picking x that are very close to (but less than) a.

Suppose a is a real number, and f is a function which is defined for all $x > a$ that are "near" the number a. We say "The limit of $f(x)$ as x approaches a from the right is L," and we write

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

if we can make $f(x)$ get as close as we want to L by picking x that are very close to (but greater than) a.

Under this definition, we see that $\lim_{x\to 0^-} H(x) = 0$ and $\lim_{x\to 0^+} H(x) = 1$.

Example 1.39. What is $\lim_{x\to 1^-} f(x)$ if $f(x) = \begin{cases} x^2+2 & x > 1 \\ x-3 & x < 1 \end{cases}$?

Answer: -2.

1.4 A Formal Definition of Limits

1.4.1 The $\epsilon - \delta$ definition

We start by giving a rigorous, formal, and intimidating-looking definition of a limit.

Definition 1.40. Suppose a is a real number, and f is a function defined on some open interval containing a, except possibly for at a. We say the *limit* of $f(x)$ as x approaches a is L, and write

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

if for every real number $\epsilon > 0$ there is a real number $\delta > 0$ such that whenever $0 < |x-a| < \delta$ then $|f(x) - L| < \epsilon$.

This looks scary, but you should notice that this is *exactly the same thing we said before* in Definition 1.33. The letter ϵ represents "how close we want $f(x)$ to get to L" and δ represents "how close x needs to get to a".

Then this definition says that if we pick any margin of error $\epsilon > 0$, then there is some distance δ such that if x is within distance δ of a, then $f(x)$ is within our margin of error ϵ of L.

Remark 1.41. The Greek letter epsilon (ϵ) became the letter "e", and stands for "error". The Greek letter delta (δ) became the letter "d", and stands for "distance". This isn't just a mnemonic for you; this is actually why those letters were chosen.

Example 1.42. 1. If $f(x) = 3x$ then prove $\lim_{x\to 1} f(x) = 3$.

Let $\epsilon > 0$ and set $\delta = \underline{\epsilon/3}$. Then if $|x-1| < \delta$ then

$$|f(x) - 3| = |3x - 3| = 3|x-1| < 3\delta = \epsilon.$$

2. If $f(x) = x^2$ then prove $\lim_{x\to 0} f(x) = 0$.

Let $\epsilon > 0$ and set $\delta = \underline{\sqrt{\epsilon}}$. Then if $|x - 0| < \delta$, then

$$|f(x) - 0| = |x^2| = |x|^2 < (\sqrt{\epsilon})^2 = \epsilon.$$

3. If $f(x) = \frac{x^2-1}{x-1}$ then $\lim_{x\to 1} f(x) = 2$.

This is harder to see at first, until we recall or notice that this function is mostly the same as $x + 1$.

Let $\epsilon > 0$ and let $\delta = \underline{\epsilon}$. Then if $0 < |x - 1| < \delta$, we have

$$\begin{aligned} |f(x) - 2| &= \left| \frac{x^2 - 1}{x - 1} - 2 \right| \\ &= |x + 1 - 2| && \text{since } x \neq 1 \\ &= |x - 1| < \delta = \epsilon. \end{aligned}$$

Remark 1.43. Despite the fact that we set δ as the first thing we do in the proof, we often figure out what it should be last. I strongly recommend beginning your proof by writing "And set $\delta =$ ____" and then working out the proof. By the time you get to the end you'll know what δ needs to be and you can go back and fill in th blank.

Example 1.44. If $f(x) = 4x - 2$ then find (with proof!) $\lim_{x\to -2} f(x)$.

We first need to generate a "guess". This is a nice function, so it seems like the answer should be close to $f(-2) = -10$.

Let $\epsilon > 0$ and set $\delta = \underline{\epsilon/4}$. Then if $|x - (-2)| < \delta$ we compute

$$|f(x) + 10| = |4x - 2 + 10| = |4x + 8| = 4|x + 2| < 4\delta = \epsilon.$$

Example 1.45. If $f(x) = x^2$ find (with proof!) $\lim_{x \to 3} f(x)$.

We first need to generate a "guess". This is a nice, should-be-continuous function, so it seems like the answer should be close to $f(3) = 9$.

Let $\epsilon > 0$ and set $\delta \leq \underline{\epsilon/7, 1}$. Then if $|x - 3| < \delta$ we compute

$$|x^2 - 9| = |x + 3| \cdot |x - 3| < |x + 3|\delta$$

but this is kind of a problem because we still have an x floating around. But logically, we know that if δ is small enough, x will be close to 3 and thus $|x + 3|$ will be close to 6.

To guarantee that $|x + 3|$ is actually close to 6, we'll require $\delta \leq 1$ as well. Then we compute

$$\begin{aligned}
|x^2 - 9| < |x + 3|\delta &= |(x - 3) + 6| \cdot \delta \\
&\leq (|x - 3| + |6|)\,\delta && \text{by the triangle inequality} \\
&< (1 + 6)\delta = 7\delta.
\end{aligned}$$

Notice we said that $|x + 3|$ would be close to 6, and what we actually showed is that $|x + 3| \leq 7$–which of course it is if it is close to 6.

So now we just need to make sure δ is small enough that $7\delta \leq \epsilon$, so in addition to letting $\delta \leq 1$ we also let $\delta \leq \epsilon/7$, so we have

$$|x^2 - 9| < 7\delta = 7\epsilon/7 = \epsilon.$$

Remark 1.46.

- We often use an approach of isolating all our xs and turning them into an $x - 3$ or $x - a$ or whatever we *know how to control*. Since in example 1.65 we know that $|x - 3| < \delta$ we want to turn all our xs into $|x - 3|$s. Then we can deal with whatever is left over.

- Notice that here we didn't actually say what δ is; we just listed some properties it needs to have, by saying that $\delta \leq \epsilon/12, 1$. If we want to pick out a specific number, we can write $\delta = \min(\epsilon/12, 1)$, but this isn't actually necessary.

Example 1.47. If $f(x) = x^2 + x$, find (with proof) $\lim_{x \to 2} f(x)$.

This is a continuous function, so it seems like the answer should be close to $f(2) = 6$.

Let $\epsilon > 0$ and set $\delta < \underline{\sqrt{\epsilon/2}, \epsilon/10}$. Then if $0 < |x - 2| < \delta$ we have

$$\begin{aligned}
|f(x) - 6| &= |x^2 + x - 6| = |(x^2 - 4) + (x - 2)| \\
&\leq |x^2 - 4| + |x - 2| && \text{(triangle inequality)} \\
&= |x - 2| \cdot |x + 2| + |x - 2| = |x - 2| (|x + 2| + 1) \\
&= |x - 2| (|x - 2 + 4| + 1) \leq |x - 2| (|x - 2| + 5) && \text{(triangle inequality)} \\
&< \delta(\delta + 5) = \delta^2 + 5\delta.
\end{aligned}$$

You could try to figure out exactly when $\delta^2 + 5\delta = \epsilon$, and after some quadratic formula-ing you'd find you need $\delta \leq \frac{-5+\sqrt{25+4\epsilon}}{2}$. But that's tedious and actually way too much work. (But if you prefer this approach it's perfectly acceptable).

It's easier to instead list two conditions: we let $\delta \leq \sqrt{\epsilon/2}, \epsilon/10$. Then $\delta^2 \leq \epsilon/2$ and $5\delta \leq \epsilon/2$, and we have

$$|f(x) - 6| < \delta^2 + 5\delta \leq \epsilon/2 + \epsilon/2 = \epsilon.$$

Example 1.48. Now suppose

$$g(x) = \begin{cases} x^2 + x & x \neq 2 \\ 0 & x = 2 \end{cases}$$

What is $\lim_{x \to 2} g(x)$?

This looks really nasty, but is actually easy after we already did Example 1.47.

The limit doesn't care about what happens at any one specific point, and especially doesn't care about what happens at 2. So for our purposes, this function is the same as $f(x) = x^2 + x$, and thus the limit is, as before, 6.

Let $\epsilon > 0$, and let $\delta < \sqrt{\epsilon/2}, \epsilon/10$. Then if $0 < |x - 2| < \delta$ we have

$$|g(x) - 6| = \left|x^2 + x - 6\right| < \epsilon$$

as computed in Example 1.47. (This is a completely valid proof as written!)

1.4.2 Limit Laws

We now hopefully have a good understanding of what we want limits to *mean*. But this sort of proof process would be super cumbersome if we needed to use it every time we wanted to compute a limit. Fortunately, we can make things much simpler. In this (sub)section we'll introduce basic ideas that we use to make computing limits reasonable; in the next couple of sections we'll see how we do this in practice.

Our approach to computing limits begins with three basic principles, the most important of which we've already seen.

Lemma 1.49 (Identity)**.** *Let a be a real number. Then* $\lim_{x\to a} x = a$.

Proof. Let $\epsilon > 0$ and let $\delta = \epsilon$. If $|x-a| < \delta$, then $|x-a| < \delta = \epsilon$. □

Lemma 1.50 (Constants)**.** *Prove that if a, c are real numbers, then* $\lim_{x\to a} c = c$.

Proof. Let $\epsilon > 0$, and set $\delta = 1$. Then if $0 < |x-a| < \delta$ we have $|f(x) - c| = |c-c| = 0 < \epsilon$. □

Lemma 1.51 (Almost Identical Functions)**.** *If $f(x) = g(x)$ on some open interval $(a-d, a+d)$ surrounding a, except possibly at a, then $\lim_{x\to a} f(x) = \lim_{x\to a} g(x)$ whenever one limit exists.*

Proof. Suppose $\lim_{x\to a} f(x) = L$. Let $\epsilon > 0$; then there is some δ_1 such that if $0 < |x-a| < \delta_1$ then $|f(x) - L| < \epsilon$. Then let $\delta < d, \delta_1$. If $0 < |x-a| < \delta$ then $g(x) = f(x)$, and thus

$$|g(x) - L| = |f(x) - L| < \epsilon.$$

□

But by themselves, these results aren't terribly interesting; all of those functions are boring! But importantly, we can also learn how limits interact with basic algebraic operations, which allows us to break complicated expressions up into these simple parts.

Proposition 1.52. *Suppose c is a constant real number, and f and g are functions such that $\lim_{x\to a} f(x) = L_1$ and $\lim_{x\to a} g(x) = L_2$ exist. Then*

1. *(Additivity)* $\lim_{x\to a} (f(x) \pm g(x)) = \lim_{x\to a} f(x) \pm \lim_{x\to a} g(x)$.

   Proof. Let $\epsilon > 0$. Then there exist $\delta_1, \delta_2 > 0$ such that if $0 < |x-a| < \delta_1$ then $|f(x) - L_1| < \epsilon/2$, and if $0 < |x-a| < \delta_2$ then $|g(x) - L_2| < \epsilon/2$.

   Let $\delta \leq \delta_1, \delta_2$. Then if $0 < |x-a| < \delta$, we compute

   $$|f(x)+g(x)-(L_1+L_2)| = |(f(x)-L_1)+(g(x)-L_2)| \leq |f(x)-L_1|+|g(x)-L_2| < \epsilon/2+\epsilon/2 = \epsilon.$$

   □

2. *(Scalar multiples)* $\lim_{x\to a}(cf(x)) = c\lim_{x\to a} f(x)$

Proof. If $c = 0$ then the left hand side is $\lim_{x\to a} 0 = 0$ and the right hand side is $0L_1 = 0$ so the equality holds.

If $c \neq 0$, then let $\epsilon > 0$. Then by definition of limit, there exists some δ so that if $0 < |x-a| < \delta$ then $|f(x) - L_1| < \epsilon/c$.

Then if $0 < |x-a| < \delta$, we have

$$|cf(x) - cL_1| = c|f(x) - L_1| < c(\epsilon/c) = \epsilon,$$

which is what we wanted to show. □

3. *(Products)* $\lim_{x\to a}(f(x)g(x)) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x)$.

Proof. Let $\epsilon > 0$. Then there exist δ_1, δ_2 such that

- if $0 < |x-a| < \delta_1$ then $|f(x) - L_1| < \epsilon/(2|L_2|), 1$,
- and if $0 < |x-a| < \delta_2$ then $|g(x) - L_2| < \epsilon/(2|L_1| + 2)$.

Set $\delta \leq \delta_1, \delta_2$. Then if $0 < |x-a| < \delta$, we compute

$$\begin{aligned}
|f(x)g(x) - L_1L_2| &= |f(x)g(x) - f(x)L_2 + f(x)L_2 - L_1L_2| \\
&\leq |f(x)g(x) - f(x)L_2| + |f(x)L_2 - L_1L_2| \\
&= |f(x)| \cdot |g(x) - L_2| + |L_2| \cdot |f(x) - L_1| \\
&= |f(x) - L_1 + L_1| \cdot |g(x) - L_2| + |L_2| \cdot |f(x) - L_1| \\
&\leq (|f(x) - L_1| + |L_1|) \cdot |g(x) - L_2| + |L_2| \cdot |f(x) - L_1| \\
&< (1 + |L_1|)\,(\epsilon/(2|L_1| + 2)) + |L_2| \cdot \epsilon/(2Ł_2|) \\
&= \epsilon/2 + \epsilon/2 = \epsilon.
\end{aligned}$$

□

4. *(Quotients) That last rule also works with division if that makes sense: if* $\lim_{x\to a} g(x) \neq 0$*, then*

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

Proof. I'm not going to prove this because it's really long and annoying and not very informative. It's a lot like the last proof except more tedious. If you're feeling masochistic you can probably prove it yourself. □

5. *(Exponents) The rule for multiplication extends to exponentials:* $\lim_{x\to a}(f(x)^n) = (\lim_{x\to a} f(x))^n$. *Also roots:* $\lim_{x\to a} \sqrt[n]{f(x)} = \sqrt[n]{\lim_{x\to a} f(x)}$, *assuming all the functions make sense.*

Proof. We're only going to prove this for the case of $f(x)^n$ where n is a positive integer. The other proofs are basically the same, but this has less bookkeeping.

$$\begin{aligned}
\lim_{x\to a} f(x)^n &= \lim_{x\to a} f(x)\cdot f(x)^{n-1} \\
&= \left(\lim_{x\to a} f(x)\right)\left(\lim_{x\to a} f(x)^{n-1}\right) && \text{by the rule on products} \\
&\vdots \\
&= \left(\lim_{x\to a} f(x)\right)\cdot\left(\lim_{x\to a} f(x)\right)\cdot\cdots\cdot\left(\lim_{x\to a} f(x)\right) \\
&= \left(\lim_{x\to a} f(x)\right)^n
\end{aligned}$$

Formally we should write this up as a "proof by induction", which you can learn about in Math 2971. ☐

Example 1.53. 1.

$$\begin{aligned}
\lim_{x\to 1} x^3 &= \left(\lim_{x\to 1} x\right)^3 && \text{Exponents} \\
&= 1^3 && \text{Identity} \\
&= 1
\end{aligned}$$

2.

$$\begin{aligned}
\lim_{x\to 1}(x+1)^3 - 2 &= \lim_{x\to 1}(x+1)^3 - \lim_{x\to 1} 2 && \text{Additivity} \\
&= \left(\lim_{x\to 1}(x+1)\right)^3 - 2 && \text{Exponents and Constants} \\
&= \left(\lim_{x\to 1} x + \lim_{x\to 1} 1\right)^3 - 2 && \text{Additivity} \\
&= (1+1)^3 - 2 && \text{Identity and Constants} \\
&= 2^3 - 2 = 8 - 2 = 6.
\end{aligned}$$

3.

$$\begin{aligned}
\lim_{x\to 1} \frac{x^2}{x} &= \frac{\lim_{x\to 1} x^2}{\lim_{x\to 1} x} && \text{Quotients} \\
&= \frac{\left(\lim_{x\to 1} x\right)^2}{\lim_{x\to 1} x} && \text{Exponents} \\
&= \frac{1^2}{1} && \text{Identity} \\
&= 1/1 = 1.
\end{aligned}$$

We can also approach this problem a different way, since this function is just the same as x everywhere except at 0:

$$\begin{aligned}
\lim_{x\to 1} \frac{x^2}{x} &= \lim_{x\to 1} x && \text{Almost Identical Functions} \\
&= 1 && \text{Identity}
\end{aligned}$$

4.

$$\begin{aligned}
\lim_{x\to 0} \frac{x^2}{x} &= \lim_{x\to 0} x && \text{Almost Identical Functions} \\
&= 0
\end{aligned}$$

Unlike the previous problem, we *cannot* use the Quotient property here because the bottom approaches zero. Compare:

$$\begin{aligned}
\lim_{x\to 0} \frac{x}{x^2} &= \lim_{x\to 0} \frac{1}{x} && \text{Almost Identical Functions} \\
&\neq \frac{\lim_{x\to 0} 1}{\lim_{x\to 0} x}
\end{aligned}$$

The last step doesn't work because now we're dividing by zero, which we can never do. This limit is in fact $\pm\infty$, and we'll look at how to show that without a proof from the definition soon.

Of course, even showing all these steps gets tedious, so you don't have to do that unless I explicitly ask you to. (However, it will be a topic on a mastery quiz.) It's useful to be able to do this when you want to check your work carefully, or when you're working with something particularly tricky.

1.5 Continuity and Computing Limits

Now that we understand limits, we can return to continuity.

Definition 1.54 (Formal)**.** We say that f is continuous at a if $\lim_{x\to a} f(x) = f(a)$.

This definition works in both directions. If we want to know whether a function is continuous, we can check its limits; and if we want to know the limit of a continuous function, we can find it by plugging in.

This really is the same as the less formal definition we gave in section 1.3. There, we said that f is continuous if $f(a)$ is a good approximation for $f(x)$; here we say that f is continuous if $f(x)$ is a good approximation for $f(a)$. This also clarifies *how good* the approximation needs to be. For f to be continuous, the approximation needs to get perfect as x gets close to a.

The definition of continuity says that $\lim_{x\to a} f(x) = f(a)$. This secretly actually requires three distinct things to happen:

1. The function is defined at a; that is, a is in the domain of f.

2. $\lim_{x\to a} f(x)$ exists.

3. The two numbers are the same.

There are a few different ways for a function to be discontinuous at a point:

1. A function f has a *removable discontinuity* at a if $\lim_{x\to a} f(x)$ exists but is not equal to $f(a)$.

2. A function f has a *jump discontinuity* at a if $\lim_{x\to a^-} f(x)$ and $\lim_{x\to a^+} f(x)$ both exist but are unequal.

3. A function f has a *infinite discontinuity* if f takes on aribtrarily large or small values near a. We'll talk about this more soon.

4. It's also possible for the one-sided limits to not exist, but this doesn't have a special name. We'll see this with $\sin(1/x)$ when we study trigonometric functions in section 1.6. In this class, I'll just call a function like this *really bad*. But we'll mostly avoid talking about them.

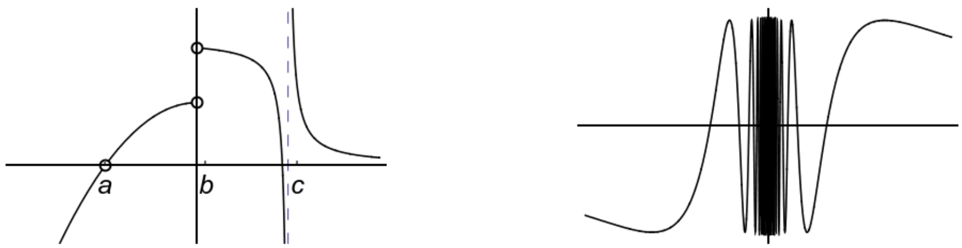


Figure 1.13: We saw this picture in section 1.3, but now we have language to talk about it.

A common informal definition is that a continuous function is one whose we can draw without lifting our pencil from the paper. Once we make this precise, this is another way to think about continuous functions. And we make it precise via the Intermediate Value Theorem

Theorem 1.55 (Intermediate Value Theorem)**.** *Suppose f is continuous (and defined!) on the closed interval $[a, b]$ and y is any number between $f(a)$ and $f(b)$. Then there is a c in (a, b) with $f(c) = y$.*

Example 1.56. Suppose $f(x)$ is a continuous function with $f(0) = 3, f(2) = 7$. Then by the Intermediate Value Theorem there is a number c in $(0, 2)$ with $f(c) = 5$.

Example 1.57. Let $g(x) = x^3 - x + 1$. Use the Intermediate Value Theorem to show that there is a number c such that $g(c) = 4$.

To use the intermediate value theorem, we need to check that our function is continuous, and then find one input whose output is less than 4, and another whose output is greater than 4. g is a polynomial and thus continuous. Testing a few values, we see $g(0) = 1, g(1) = 1, g(2) = 7$. Since $g(1) = 1 < 4 < 7 = g(2)$, by the Intermediate Value Theorem ther is a c in $(1, 2)$ with $g(c) = 4$.

Example 1.58. Show that there is a θ in $(0, \pi/2)$ such that $\sin(\theta) = 1/3$.

We know that sin is a continuous function, and that $\sin(0) = 0$ and $\sin(\pi/2) = 1$. Since $0 < 1/3 < 1$, by the Intermediate Value Theorem there is a θ in $(0, \pi/2)$ such that $\sin(\theta) = 1/3$.

Remark 1.59. The converse of this theorem is not true. It is possible to have a function that satisfies the conclusions of the Intermediate Value Theorem, but is not continuous; these functions are called Darboux Functions.

For example, let $f(x) = \begin{cases} \sin(1/x) & x \neq 0 \\ 0 & x = 0 \end{cases}$. Then f satisfies the conclusion of the intermediate value theorem: it's continuous except at zero, so the theorem works on any interval that doesn't contain zero. Any interval containing zero contains every value in $[-1, 1]$, so if $a < 0 < b$ and y is between $f(a)$ and $f(b)$, then $-1 \leq y \leq 1$ and so there is a c in (a, b) such that $f(c) = y$. Thus f is Darboux.

Historically, the main reason we didn't take this as the definition of continuous, instead of the limit definition that we actually use, is that we didn't want to treat functions like this as "continuous".

1.5.1 Limits of Continuous Funtions

This definition does a few things for us:

1. It gives us a clear rule for when a function is continuous. In particular, it will resolve questions about edge-case "weird" functions like $\sin(1/x)$, as we'll discuss in section 1.6.
2. If we know a function is continuous, we can easily compute its limit just by plugging in the value.
3. The conclusion of our discussion of limit laws in section 1.4.2 is that when functions are made up of algebraic operations, they are continuous whenever they are defined.

Example 1.60. 1. The function $f(x) = 3x$ is continuous at 1, so $\lim_{x \to 1} f(x) = f(1) = 3$.

2. The function $f(x) = x^2$ is continuous at 0, so $\lim_{x \to 0} f(x) = f(0) = 0$.
3. The function $f(x) = \frac{x^2-1}{x-1}$ is definitely not continuous at 1, because it's not defined there. But we can use almost identical functions:

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

Example 1.61. If $f(x) = \frac{x-1}{x^2-1}$ then what is $\lim_{x \to 1} f(x)$?

Answer: $1/2$. If $x \neq 1$, then

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

We know that $\frac{1}{x+1}$ is continuous, and that it is defined at $a = 1$. Thus $\lim_{x \to 1} f(x) = \lim_{x \to 1} \frac{1}{x+1} = \frac{1}{2}$.

Example 1.62. $\lim_{x \to -2} \frac{(x+1)^2-1}{x+2} = \lim_{x \to -2} \frac{x^2+2x+1-1}{x+1} = \lim_{x \to -2} \frac{x(x+2)}{x+2} = \lim_{x \to -2} x = -2.$

Note that $\frac{x(x+2)}{x+2} \neq x$, but their limits at 0 are the same because the functions are the same near 0 (and in fact everywhere except at 0).

Example 1.63. What is $\lim_{x\to 0} \frac{\sqrt{9+x}-3}{x}$?

We use a trick called multiplication by the conjugate, which takes advantage of the fact that $(a+b)(a-b) = a^2 - b^2$. This trick is used *very often* so you should get comfortable with it.

$$\begin{aligned}
\lim_{x\to 0} \frac{\sqrt{9+x}-3}{x} &= \lim_{x\to 0} \frac{\sqrt{9+x}-9}{x} \frac{\sqrt{9+x}+3}{\sqrt{9+x}+3} \\
&= \lim_{x\to 0} \frac{(9+x)-3}{x(\sqrt{9+x}+3)} = \lim_{x\to 0} \frac{x}{x(\sqrt{9+x}+3)} \\
&= \lim_{x\to 0} \frac{1}{\sqrt{9+x}+3} = \frac{1}{\lim_{x\to 0} \sqrt{9+x}+3} = \frac{1}{6}.
\end{aligned}$$

Example 1.64. What is $\lim_{x\to 1} \frac{x-1}{\sqrt{5-x}-2}$?

$$\begin{aligned}
\lim_{x\to 1} \frac{x-1}{\sqrt{5-x}-2} &= \lim_{x\to 1} \frac{x-1}{\sqrt{5-x}-2} \frac{\sqrt{5-x}+2}{\sqrt{5-x}+2} \\
&= \lim_{x\to 1} \frac{(x-1)(\sqrt{5-x}+2)}{(5-x)-4} \\
&= \lim_{x\to 1} \frac{(x-1)(\sqrt{5-x}+2)}{-(x-1)} \\
&= \lim_{x\to 1} -(\sqrt{5-x}+2) = -4.
\end{aligned}$$

Example 1.65. The *Heaviside Function* or *step function* is given by

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

It is often used in electrical engineering applications to describe the current running through a switch before and after it has been flipped.

We can ask: what is $\lim_{x\to 0} H(x)$?

There isn't one: no matter how close x gets to 0, sometimes $H(x)$ will be 0 and sometimes it will be 1. So there is no one value that approximates $H(x)$ for any x near a.

However, the Heaviside function clearly behaves well if look only at one side or the other of it. And just as we could talk about continuity to one side or the other, we can talk about *one-sided limits*.

Definition 1.66. Suppose a is a real number, and f is a function which is defined for all $x < a$ that are "near" the number a. We say "The limit of $f(x)$ as x approaches a from the left is L," and we write

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

if we can make $f(x)$ get as close as we want to L by picking x that are very close to (but less than) a.

Suppose a is a real number, and f is a function which is defined for all $x > a$ that are "near" the number a. We say "The limit of $f(x)$ as x approaches a from the right is L," and we write

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

if we can make $f(x)$ get as close as we want to L by picking x that are very close to (but greater than) a.

Under this definition, we see that $\lim_{x \to 0^-} H(x) = 0$ and $\lim_{x \to 0^+} H(x) = 1$.

Example 1.67. What is $\lim_{x \to 1^-} f(x)$ if $f(x) = \begin{cases} x^2 + 2 & x > 1 \\ x - 3 & x < 1 \end{cases}$?

Answer: -2.

Example 1.68. The Heaviside function of example 1.65 is not continuous, since there's a jump at 0.

It is continuous from the right at 0, since $\lim_{x \to 0^+} H(x) = 1 = H(0)$. This function is not continuous from the left, since $\lim_{x \to 0^-} H(x) = 0 \neq H(0)$.

Definition 1.69. A function is *continuous from the right at* a if $\lim_{x \to a^+} f(x) = f(a)$.

A function is *continuous from the left at* a if $\lim_{x \to a^-} f(x) = f(a)$.

Proposition 1.70. *A function is continuous at* a *if and only if it is continuous from the left and from the right at* a.

Remark 1.71. At a jump discontinuity, a function will often be continuous from one side but not the other. This is not necessarily the case, though: consider the function

$$f(x) = \begin{cases} 2 & x > 0 \\ 1 & x = 0 \\ 0 & x < 0 \end{cases}$$

Limits exist from the right and the left, but the function is not continuous from either side.

1.5.2 Function Extensions

Recall we like continuous functions because we can use their values at one point to approximate the values they should have at nearby points. And we observed that this is really unhelpful at any point where the function isn't defined. So if we have a function that's continuous everywhere it's defined, we'd like to replace it with a function that is continuous—and defined—everywhere.

Definition 1.72. We say that g is an *extension* of f if the domain of g contains the domain of f, and $g(x) = f(x)$ whenever $f(x)$ is defined.

In general, we can only extend a function to be continuous at all real numbers if the only discontinuities were removable. This is why we call discontinuities like that "removable".

Example 1.73. Let $f(x) = \frac{x^2-1}{x-1}$. Can we define a function g that agrees with f on its domain, and is continuous at all reals?

f is continuous everywhere on its domain, and is undefined at $x = 1$. We can see that $g(x) = x + 1$ will give the same value as f everywhere on f's domain, and it is continuous since it is a polynomial. Thus g is a continuous extension of f to all reals.

Alternatively, we could compute that $\lim_{x \to 1} f(x) = 2$. Then we define

$$h(x) = \begin{cases} \frac{x^2-1}{x-1} & x \neq 1 \\ 2 & x = 1. \end{cases}$$

The function $h(x)$ is defined at all reals, and since it is continuous at 1 by our computation, it is continuous everywhere. It also must extend f since it is just defined to be f everywhere in the domain of f. So h is a continuous extension of f to all reals.

Importantly, g and h are actually the same function, since they give the same output for every input. There is at most one continuous extension of any given function; but there are multiple ways to describe that extension.

Example 1.74. The function $f(x) = 1/x$ is continuous on its domain, but we cannot extend it to a function continuous at all reals, because the limit at 0 does not exist.

Example 1.75. Let $f(x) = \frac{x^2-4x+3}{x-3}$. Can we extend f to a function continuous at all reals?

Answer: f is continuous at all reals except $x = 3$. But the function $g(x) = x - 1$ is the same everywhere except for 3, and is continuous at 3.

Example 1.76. Let

$$g(x) = \begin{cases} x^2 + 1 & x > 2 \\ 9 - 2x & x < 2 \end{cases}$$

Can we extend this to a continuous function on all reals?

Answer: $\lim_{x\to 2^-} f(x) = \lim_{x\to 2^-} 9-2x = 5$, and $\lim_{x\to 2^+} f(x) = \lim_{x\to 2^+} x^2+1 = 5$, so the limit at 2 exists. Thus we can extend g to

$$g_f(x) = \begin{cases} x^2+1 & x \geq 2 \\ 9-2x & x \leq 2 \end{cases}$$

which is continuous at all reals.

1.6 Trigonometry and the Squeeze Theorem

We now want to look at limits of trigonometric functions. Fortunately, they behave *mostly* how we want them to.

Proposition 1.77. *If a is a real number, then $\lim_{x\to a}\sin(x) = \sin(a)$ and $\lim_{x\to a}\cos(x) = \cos(a)$.*

In fact, since trigonometric functions are just ways of combining sine and cosine, essentially all trigonometric functions behave this way where they are defined.

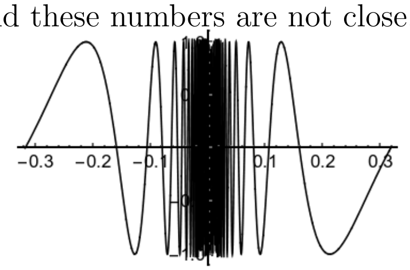
Example 1.78. $\lim_{x\to\pi}\cos(x) = -1$.

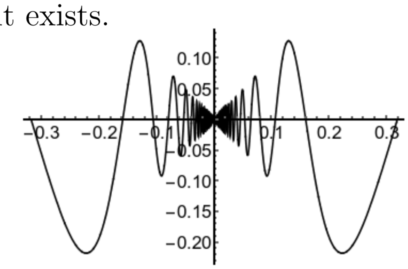
$\lim_{x\to\pi}\tan(x) = 0$.

But where the functions are not defined, sometimes very odd things can happen. We've seen a graph of $\sin(1/x)$ before, in section 1.3. We said that the function wasn't continuous at 0. In fact, no limit exists there.

Suppose a limit does exist at zero; specifically, let's suppose that $\lim_{x\to 0}\sin(1/x) = L$. Then if x is close to 0, it must be the case that $\sin(1/x)$ is close to L.

But however close we want x to be to 0, we can find a $x_1 = \frac{1}{(2n+1/2)\pi}$, and then $\sin(1/x_1) = \sin((2n+1/2)\pi) = \sin(\pi/2) = 1$. But we can also find an $x_2 = \frac{1}{(2n+3/2)\pi}$ so that $\sin(1/x_2) = \sin(2n\pi + 3\pi/2) = \sin(3\pi/2) = -1$. So L must be really close to 1 and really close to -1, and these numbers are not close. So no limit exists.





Left: graph of $\sin(1/x)$, Right: graph of $x\sin(1/x)$

In contrast, from the graph it appears that $\lim_{x\to 0} x\sin(1/x)$ does exist. We can't possibly prove this by replacing $x\sin(1/x)$ with an almost identical function and plugging values in:

the function is gross and complicated, and any almost identical function will also be gross and complicated.

But we can easily see that $\lim_{x\to 0} x = 0$. This doesn't mean that $\lim_{x\to 0} xf(x) = 0$ for any $f(x)$; if $f(x)$ gets really big then it can "cancel out" the x term getting very small. (A good example of this is $\lim_{x\to 0} x\frac{1}{x}$, which is of course 1).

But if we can prove that the second term, which in this case is $\sin(1/x)$, does *not* get really big, then the entire limit will have to go to zero. We make this intuition precise with the following important theorem:

Theorem 1.79 (Squeeze Theorem)**.** *If* $f(x) \leq g(x) \leq h(x)$ *near* a *(except possibly at* a*), and* $\lim_{x\to a} f(x) = \lim_{x\to a} h(x) = L$*, then* $\lim_{x\to a} g(x) = L$*.*

To use the Squeeze Theorem, we need to do two things:

1. Find a lower bound and an upper bound for the function we're interested in; and
2. show that their limits are equal.

We usually do this by factoring the function we care about into two pieces, where one goes to zero and the other is bounded, and thus doesn't get infinitely big.

In this case, we know that $-1 \leq \sin(1/x) \leq 1$ by properties of $\sin(x)$. We "want" to multiply both sides of the equation by x to get $-x \leq x\sin(1/x) \leq x$, but this is actually incorrect when x is negative. In general, it's hard to reason about inequalities when negative numbers are involved, so we use absolute values to make sure we don't have to worry about it:

$$-|x| \leq x\sin(1/x) \leq |x|$$

Then we can compute that $\lim_{x\to 0}(-|x|) = \lim_{x\to 0}|x| = 0$ and so by the squeeze theorem, $\lim_{x\to 0} x\sin(1/x) = 0$.

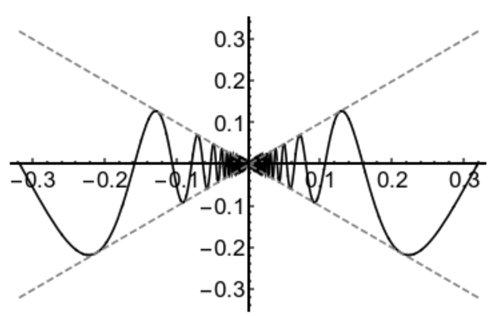


Figure 1.14: A graph of $x\sin(1/x)$ with $|x|$ and $-|x|$

This means that we can extend the function $x\sin(1/x)$ to be continuous at all reals, by defining

$$f(x) = \begin{cases} x\sin(1/x) & x \neq 0 \\ 0 & x = 0. \end{cases}$$

Remark 1.80. There is an argument people make sometimes that looks like the squeeze theorem, but is actually wrong. People reason:

$$\begin{aligned} -|x| &\leq x\sin(1/x) \leq |x| \\ \lim_{x\to 0} -|x| &\leq \lim_{x\to 0} x\sin(1/x) \leq \lim_{x\to 0} |x| \\ 0 &\leq \lim_{x\to 0} x\sin(1/x) \leq 0 \end{aligned}$$

and conclude that $\lim_{x\to 0} x\sin(1/x) = 0$.

However, this reasoning only works if you already know the limit exists. Compare:

$$\begin{aligned} -1 &\leq \sin(1/x) \leq 1 \\ \lim_{x\to 0} -1 &\leq \lim_{x\to 0} \sin(1/x) \leq \lim_{x\to 0} 1 \\ -1 &\leq \lim_{x\to 0} \sin(1/x) \leq 1. \end{aligned}$$

This uses the same reasoning, but the third statement doesn't actually make any sense because the limit doesn't exist. (Imagine writing that $-1 \leq \text{green} \leq 1$, for instance).

Example 1.81. Using the Squeeze Theorem, show that $\lim_{x\to 3}(x-3)\frac{x^2}{x^2+1} = 0$.

We could in fact do this without the squeeze theorem, but we also can use squeeze.

We divide the function into two parts. We see that $(x-3)$ approaches zero, so we need to bound the other factor.

We know that $0 \leq x^2 \leq x^2+1$ and so $0 \leq \frac{x^2}{x^2+1} \leq 1$ for any x. We want to multiply through by $x-3$, but that only works if $x > 3$. So we use absolute values to keep everything correct and get

$$0 \leq \left|(x-3)\frac{x^2}{x^2+1}\right| \leq |x-3|.$$

Then $\lim_{x\to 3} 0 = \lim_{x\to 3} -|x-3| = 0$, and so by the squeeze theorem $\lim_{x\to 3}(x-3)\frac{x^2}{x^2+1} = 0$.

Example 1.82. What is

$$\lim_{x\to 1} \frac{x-1}{2+\sin\left(\frac{1}{x-1}\right)}?$$

The top goes to zero and the bottom is bounded, so this looks like a squeeze theorem problem. If you have trouble seeing this, it may help to rewrite the problem as

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

We know that $-1 \leq \sin\left(\frac{1}{x-1}\right) \leq 1$ and so $1 \leq 2+\sin\left(\frac{1}{x-1}\right) \leq 3$, and thus

$$\begin{aligned}
1 &\geq \frac{1}{2+\sin\left(\frac{1}{x-1}\right)} \geq \frac{1}{3} \\
|x-1| &\geq \frac{|x-1|}{2+\sin\left(\frac{1}{x-1}\right)} \geq \frac{|x-1|}{3} \\
|x-1| &\geq \left|\frac{x-1}{2+\sin\left(\frac{1}{x-1}\right)}\right| \geq \frac{|x-1|}{3}
\end{aligned}$$

since the denominator is always positive. But $\lim_{x\to 1}|x-1| = \lim_{x\to 1}\frac{|x-1|}{3} = 0$, so by the squeeze theorem

$$\lim_{x\to 1}\frac{x-1}{2+\sin\left(\frac{1}{x-1}\right)} = 0.$$

Example 1.83. Prove that $\lim_{x\to 3}(x-3)\left(5\sin\left(\frac{1}{x-3}\right)-2\right) = 0$.

We know that

$$\begin{aligned}
-1 &\leq \sin\left(\frac{1}{x-3}\right) \leq 1 \\
-5 &\leq 5\sin\left(\frac{1}{x-3}\right) \leq 5 \\
-7 &\leq 5\sin\left(\frac{1}{x-3}\right) - 2 \leq 3.
\end{aligned}$$

We want to multiply through by $x-3$, but this causes problems when $x<3$ and thus $x-3<0$. So first we put absolute values on everything.

But there's a subtlety here. We know our bad term is between -7 and 3. But when we take absolute values, that doesn't make it larger than $|-7|$ and smaller than $|3|$—no numbers satisfy those rules. Instead, we know that since we've added absolute values, everything will be bigger than zero. This gives us a lower bound.

For the upper bound, we care about how far away from zero we can get. One way to see this is that if $5\sin\left(\frac{1}{x-3}\right)-2>0$, we know that it must be less than 3; but if $5\sin\left(\frac{1}{x-3}\right)-2<0$, we know it must be bigger than -7, so the absolute value is <7. So overall we get the bounds

$$0 \leq \left|(x-3)\left(5\sin\left(\frac{1}{x-3}\right)-2\right)\right| \leq |7(x-3)|.$$

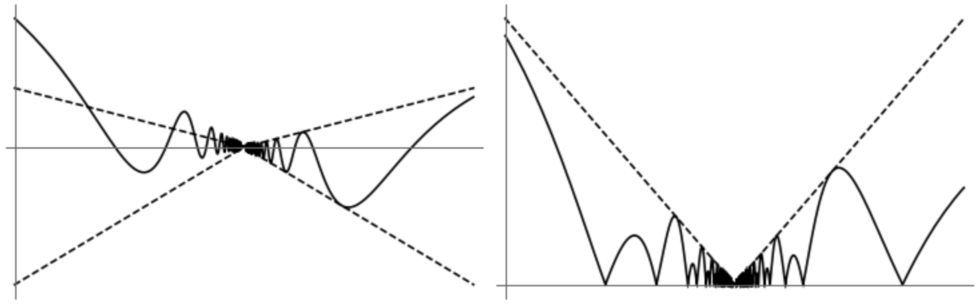


Figure 1.15: Left: $-7|x-3|$ is a fine lower bound, but $3|x-3|$ isn't an upper bound. Right: After we take absolute values, we see that $7|x-3|$ has the smallest coefficient we could possibly use and still get an upper bound.

Now we can compute that $\lim_{x\to 3} 0 = 0$ and $\lim_{x\to 3} |7(x-3)| = 0$, so by the squeeze theorem we know that $\lim_{x\to 3}(x-3)\left(5\sin\left(\frac{1}{x-3}\right)\right) = 0$.

Example 1.84. What is $\lim\limits_{x\to -1}(x+1)\cos\left(\dfrac{x^5-3x^2+e^x-1700+(2+x)^{(1+x)^x}}{(x+1)^{27.2}}\right)$?

This looks complicated but is actually quite simple. $-1 \leq \cos(y) \leq 1$ for any y, including $y = x^5-3x^2+e^x-1700+x^{x^x}$. Thus we have

$$\begin{aligned}0 &\leq |\cos(y)| \leq 1\\ 0 &\leq |(x+1)\cos(y)| \leq |x+1|.\end{aligned}$$

Then we know that $\lim_{x\to -1} 0 = \lim_{x\to -1}|x+1| = 0$. Thus by the squeeze theorem,

$$\lim_{x\to -1}|(x+1)\cos(x^5-3x^2+e^x-1700+x^{x^x})| = 0,$$

and thus

$$\lim_{x\to -1}(x+1)\cos(x^5-3x^2+e^x-1700+x^{x^x}) = 0.$$

Example 1.85. What is

$$\lim_{x\to 0}\frac{x-1}{2+\sin\left(\frac{1}{x-1}\right)}?$$

This is a trick question. Here we have no concerns about zeroes in the denominator or points outside of the domain, we can repeatedly apply limit laws:

$$\begin{aligned}\lim_{x\to 0}\frac{x-1}{2+\sin\left(\frac{1}{x-1}\right)} &= \frac{\lim_{x\to 0}(x-1)}{\lim_{x\to 0} 2+\sin\left(\frac{1}{x-1}\right)}\\ &= \frac{-1}{2+\sin\left(\lim_{x\to 0}\frac{1}{x-1}\right)}\\ &= \frac{-1}{2+\sin(-1)} = \frac{-1}{2-\sin(1)}.\end{aligned}$$

Remark 1.86. Notice that we don't conclude that since $f(x) \leq g(x) \leq h(x)$ then $\lim_{x\to a} f(x) \leq \lim_{x\to a} g(x) \leq \lim_{x\to a} h(x)$. This is in fact not always true; it's only true if the middle limit exists, which is what we're trying to prove! So we just compute the outer two limits, and then invoke the squeeze theorem.

Example 1.87. $\lim_{x\to+\infty} \frac{\sin(x)}{x}$ exists, by the squeeze theorem.

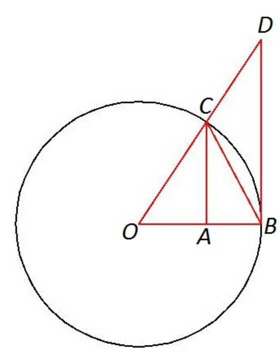
For large x we have $\frac{-1}{x} \leq \frac{\sin(x)}{x} \leq \frac{1}{x}$, and $\lim_{x\to+\infty} \frac{-1}{x} = \lim_{x\to+\infty} \frac{1}{x} = 0$. So by the squeeze theorem $\lim_{x\to+\infty} \frac{\sin(x)}{x} = 0$.

You might notice this is *exactly the same proof* we gave for $\lim_{x\to 0} x \sin(1/x)$. This is not a coincidence, since the two functions are the same after the substitution $y = 1/x$.

There is one more important limit involving sin:

Proposition 1.88 (Small Angle Approximation)**.**

$$\lim_{x\to 0} \frac{\sin x}{x} = 1$$



Proof. We'll assume x is small and positive; this all still works if x is small and negative, with different signs. Our diagram is of a circle with radius 1.

Let x be the measure of angle AOC in our diagram. Observe that $\sin x$ is precisely the length of the line segment AC by definition, and so triangle BOC has area $\sin x/2$. The area of the entire circle is π and so the area of the wedge from B to C is $\pi x/2\pi = x/2$. Since the triangle is contained in the wedge, we have $\sin x/2 \leq x/2$ and thus $\sin x/x \leq 1$.

Note that AC is $\sin x$ and AO is $\cos x$, so AC over AO is $\sin(x)/\cos(x) = \tan(x)$. By similarity, we have DB $= \tan x$, and the area of triangle BOD is $\tan x/2$. Since the wedge from B to C is contained in this triangle, we have $x/2 \leq \tan x/2$ and thus $\cos x \leq \sin x/x$.

Thus $\cos x \leq \frac{\sin x}{x} \leq 1$. But $\lim_{x\to 0} \cos x = 1$, so by the squeeze theorem we have

$$1 \leq \lim_{x\to 0} \frac{\sin x}{x} \leq 1$$

and thus get the desired result. □

Remark 1.89. This means that the function

$$f(x) = \begin{cases} \sin(x)/x & x \neq 0 \\ 1 & x = 0 \end{cases}$$

is a continuous extension of $\sin(x)/x$ to all reals.

Example 1.90. $\lim_{x\to 0} \frac{\sin(2x)}{2x} = 1$.

Example 1.91. What is $\lim_{x\to 0} \frac{\sin(4x)\sin(6x)}{\sin(2x)x}$?

We can write

$$\begin{aligned}\lim_{x\to 0} \frac{\sin(4x)\sin(6x)}{\sin(2x)x} &= \lim_{x\to 0} \frac{\sin(4x)/4x \cdot \sin(6x)/6x \cdot 24x^2}{\sin(2x)/2x \cdot 2x \cdot x} \\ &= \lim_{x\to 0} \frac{\sin 4x}{4x} \cdot \frac{\sin 6x}{6x} \cdot \frac{2x}{\sin(2x)} \cdot \frac{24x^2}{2x^2} \\ &= 1 \cdot 1 \cdot 1 \cdot 12 = 12.\end{aligned}$$

Here we are simply pairing off the $\sin(y)$'s with ys and then collecting the remainder into the last term.

Example 1.92. What is $\lim_{x\to 0} \frac{x}{\cos(x)}$?

This problem is actually easy. We can just plug in 0 for x and get $\lim_{x\to 0} \frac{x}{\cos(x)} = \frac{0}{1} = 0$.

In contrast, $\lim_{x\to 0} \frac{\cos(x)}{x}$ is mildly tricky, and we're not ready to do it yet. We'll discuss this sort of limit in section 1.7.1.

Example 1.93. What is $\lim_{x\to 0} \frac{x\sin(2x)}{\tan(3x)}$?

When we see a tangent in a problem, it is often helpful to rewrite it in terms of sin and cos. We can then collect terms:

$$\begin{aligned}\lim_{x\to 0} \frac{x\sin(2x)}{\tan(3x)} &= \lim_{x\to 0} \frac{x\sin(2x)}{\sin(3x)/\cos(3x)} \\ &= \lim_{x\to 0} \frac{3x}{\sin 3x} \cdot \frac{\sin(2x)\cos(3x)}{3} = 1 \cdot \frac{0}{3} = 0.\end{aligned}$$

Example 1.94. What is $\lim_{x\to 3} \frac{\sin(x-3)}{x-3}$?

This is a small angle approximation again, since $x - 3$ is approaching zero. Thus the limit is 1.

Example 1.95. What is $\lim_{x\to 3} \frac{\sin(x^2-9)}{x-3}$?

We have a $\sin(0)$ on the top and a 0 on the bottom, but the 0s don't come from the same form; we need to get a $x^2 - 9$ term on the bottom. Multiplication by the conjugate gives

$$\begin{aligned}\lim_{x\to 3} \frac{\sin(x^2-9)}{x-3} &= \lim_{x\to 3} \frac{\sin(x^2-9)}{x-3} \cdot \frac{x+3}{x+3} = \lim_{x\to 3} \frac{\sin(x^2-9)(x+3)}{x^2-9} \\ &= \lim_{x\to 3} \frac{\sin(x^2-9)}{x^2-9} \cdot \lim_{x\to 3} x+3 = 1 \cdot (3+3) = 6.\end{aligned}$$

Example 1.96. What is $\lim_{x\to 0} \frac{1-\cos x}{x}$?

We can see that the limits of the top and the bottom are both 0, so this is an indeterminate form. We can't use the small angle approximation directly because there is no sin here at all. But we can fix that by multiplying by the conjugate.

$$\begin{aligned}\lim_{x\to 0} \frac{1-\cos x}{x} &= \lim_{x\to 0} \frac{1-\cos x}{x} \cdot \frac{1+\cos(x)}{1+\cos(x)} = \lim_{x\to 0} \frac{1-\cos^2(x)}{x(1+\cos(x))} = \lim_{x\to 0} \frac{\sin^2(x)}{x(1+\cos(x))} \\ &= \lim_{x\to 0} \frac{\sin(x)}{1+\cos(x)} = \frac{0}{2} = 0.\end{aligned}$$

1.7 Infinite Limits

A few times in the past couple sections we've talked about vertical asymptotes, or functions going to infinity. In this section we want to look at exactly what that means. Some limits deal with infinity as an output, and others deal with it as an input (or both).

Remark 1.97. Recall that infinity is not a number. Sometimes while dealing with infinite limits we might make statements that appear to treat infinity as a number. But it's not safe to treat ∞ like a true number and we will be careful of this fact.

1.7.1 Limits To Infinity

Definition 1.98. We write

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

to indicate that as x gets close to a, the values of $f(x)$ get arbitrarily large (and positive).

We write

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

to indicate that as x gets close to a, the values of $f(x)$ get arbitrarily negative.

We write

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

to indicate that as x gets close to a, the values of $f(x)$ get arbitrarily positive or negative. We usually use this when both occur.

Remark 1.99. Important note: If the limit of a function is infinity, the limit *does not exist.* This is utterly terrible English but I didn't make it up so I can't fix it. All the theorems that say "If a limit exists" are not including cases where the limit is infinite.

Lemma 1.100. *Let $f(x), g(x)$ be defined near a, such that $\lim_{x\to a} f(x) = c \neq 0$ and $\lim_{x\to a} g(x) = 0$. Then*

$$\lim_{x\to 0} \frac{f(x)}{g(x)} = \pm\infty.$$

Further, assuming $c > 0$ then the limit is $+\infty$ if and only if $g(x) \geq 0$ near a, and the limit is $-\infty$ if and only if $g(x) \leq 0$ near a. If $c < 0$ then the opposite is true.

Remark 1.101. If the limit of the numerator is zero, then this lemma is *not useful*. That is one of the "indeterminate forms" which requires more analysis before we can compute the limit completely.

Example 1.102. What is $\lim_{x\to 3} \frac{-1}{\sqrt{x-3}}$? We see the top goes to 1 and the bottom goes to 0, so the limit is $\pm\infty$. Since the denominator is always positive and the numerator is negative, the limit is $-\infty$.

We have to be careful while working these problems: the limit laws that work for finite limits don't always work here, since the limit laws assume that the limits exist, and these do not. In particular, adding and subtracting infinity *does not work*. Instead, we need to arrange the function into a form where we can use lemma 1.100.

Example 1.103. We already know that $\lim_{x\to 0} 1/x = \pm\infty$.

1. If we take $\lim_{x\to 0} 1/x - 1/x$, we could say the limit is $\pm\infty - \pm\infty$, but this is silly—the limit is actually 0.

2. In contrast, $\lim_{x\to 0} 1/x + 1/x = \lim_{x\to 0} 2/x = \pm\infty$. We don't add the infinities together.

3. And $\lim_{x\to 0} 1/x + 1/x^2$ is the trickiest. We have a $\pm\infty$ plus a $+\infty$. But again we can't add infinities—we need to combine them into one term.

   $$\lim_{x\to 0} \frac{1}{x} + \frac{1}{x^2} = \lim_{x\to 0} \frac{x+1}{x^2} = +\infty$$

   since the numerator approaches 1 and the denominator approaches 0, but is always positive.

   We could heuristically say that $\frac{1}{x^2}$ goes to $+\infty$ "faster" than $\frac{1}{x}$ goes to $\pm\infty$, and so it wins out; but this is really vague and handwavy so we try to replace it with more precise arguments like this one.

We organize our thinking about these situations in terms of the "indeterminate forms", which are: $\frac{0}{0}, \frac{\infty}{\infty}, 0 \cdot \infty, \infty \pm \infty, 1^\infty, \infty^0$. Notice that none of these are actual numbers, and they can never be the correct answer to pretty much any question.

More importantly, indeterminate forms don't even tell us what the answer should be; if plugging in gives you one of those forms, the true limit could potentially be pretty much anything. We have to do more work to get our functional expression into a determinate form. As a general rule, we use algebraic manipulations to get a form of $\frac{0}{0}$, then factor out and cancel $(x - a)$ until either the numerator or the denominator is no longer 0.

Remark 1.104. Neither $\frac{0}{1}$ nor $\frac{1}{0}$ is an indeterminate form. $\frac{0}{1}$ is just a number, equal to 0. $\frac{1}{0}$ is not a number and is never the correct answer to a question, but it's also not indeterminate. By lemma 1.100, if $\lim f(x) = 1$ and $\lim g(x) = 0$ then $\lim f(x)/g(x) = \pm\infty$.

Similarly, $\frac{0}{\infty}$ and $\frac{\infty}{0}$ are also not numbers but not indeterminate. The first suggests the limit is 0; the second suggests the limit is $\pm\infty$.

The form $\infty \cdot \infty$ mostly works fine, and gives you another ∞ whose sign depends on the signs of the ∞s you're multiplying. But again, $\infty \cdot \infty$ is never the actual answer to any actual question.

Example 1.105. What is $\lim_{x \to -2} \frac{1}{x+2} + \frac{2}{x(x+2)}$? This looks like $\infty + \infty$ so we have to be careful. We have

$$\begin{aligned} \lim_{x \to -2} \frac{1}{x+2} + \frac{2}{x(x+2)} &= \lim_{x \to -2} \frac{x}{x+2} + \frac{2}{x(x+2)} \\ &= \lim_{x \to -2} \frac{x+2}{x(x+2)} = \lim_{x \to -2} \frac{1}{x} = \frac{-1}{2}. \end{aligned}$$

Example 1.106. $\lim_{x \to 3^+} \frac{1}{(x-3)^3} = +\infty$: the limit of the top is 1, and the limit of the bottom is 0, so the limit is $\pm\infty$. But when $x > 3$ the denominator is ≥ 0, so the limit is in fact $+\infty$. Conversely $\lim_{x \to 3^-} \frac{1}{(x-3)^3} = -\infty$ since when $x < 3$ we have $(x-3)^3 \leq 0$.

$\lim_{x \to -1^+} \frac{1}{(x+1)^4} = +\infty$. And $\lim_{x \to -1^-} \frac{1}{(x+1)^4} = +\infty$. Thus $\lim_{x \to -1} \frac{1}{(x+1)^4} = +\infty$.

1.7.2 Limits at infinity

A related concept is the idea of limits "at" infinity, which answers the question "what happens to $f(x)$ when x gets very big?" We can formally define this in terms of ϵ.

Definition 1.107. Let f be a function defined for (a, ∞) for some number a. We write

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

to indicate that when x is large enough, the values of $f(x)$ get arbitrarily close to L. Formally, if for every $\epsilon > 0$ there is a $M > 0$ so that if $x > M$ then $|f(x) - L| < \epsilon$.

We can write similar definitions for $\lim_{x\to-\infty} f(x)$ and $\lim_{x\to\pm\infty} f(x)$, and talk about when these limits are themselves $\pm\infty$. But here we'll skip over the formal definition and simply think informally.

In principle, we want to do the same thing we did for finite limits. But instead of having zeros on the top and bottom of a fraction, we often have infinities as well. So we want to "cancel" an infinity from the top and the bottom of the fraction. We usually do this by dividing the top and bottom by x. Then we can use the following crucial fact:

Fact 1.108. $\lim_{x\to\pm\infty} \frac{1}{x} = 0$.

This combined with tools we already have is enough to do pretty much any calculuation here.

Example 1.109. If we want to calculuate $\lim_{x\to+\infty} \frac{1}{\sqrt{x}}$, we see that

$$\lim_{x\to+\infty} \frac{1}{\sqrt{x}} = \sqrt{\lim_{x\to+\infty} \frac{1}{x}} = \sqrt{0} = 0.$$

Example 1.110. What is $\lim_{x\to+\infty} \frac{x}{x^2+1}$?

This problem illustrates the primary technique we'll use to solve infinite limits problems. It's difficult to deal with problems that have variables in the numerator and denominator, so we want to get rid of at least one. Thus we will divide out by xs on the top and the bottom until one has none left:

$$\lim_{x\to+\infty} \frac{x}{x^2+1} = \lim_{x\to+\infty} \frac{x/x}{x^2/x + 1/x} = \lim_{x\to+\infty} \frac{1}{x + \frac{1}{x}} = \lim_{x\to+\infty} \frac{1}{x} = 0.$$

Example 1.111. Some more examples of this technique:

$$\lim_{x\to-\infty} \frac{x}{x+1} = \lim_{x\to-\infty} \frac{1}{1+\frac{1}{x}} = \lim_{x\to-\infty} \frac{1}{1} = 1.$$
$$\lim_{x\to-\infty} \frac{x}{3x+1} = \lim_{x\to-\infty} \frac{1}{3+\frac{1}{x}} = \frac{1}{3}.$$

Example 1.112. What is $\lim_{x\to+\infty} \frac{x^{3/2}}{\sqrt{9x^3+1}}$? This one is a bit tricky. We want to divide the top and bottom by $x^{3/2}$. Then we can pull the factor *inside* the square root sign.

$$\lim_{x\to+\infty} \frac{x^{3/2}}{\sqrt{9x^3+1}} = \lim_{x\to+\infty} \frac{1}{\sqrt{9+1/x^{3/2}}} = \frac{1}{\sqrt{9+0}} = \frac{1}{3}.$$

Example 1.113. Sometimes it's a bit harder to see how this works. For instance, what is $\lim_{x\to+\infty} \frac{x}{\sqrt{x^2+1}}$? It's not obvious, but we use the same technique:

$$\lim_{x\to+\infty} \frac{x}{\sqrt{x^2+1}} = \lim_{x\to+\infty} \frac{x/x}{\sqrt{x^2+1}/x} = \lim_{x\to+\infty} \frac{1}{\sqrt{x^2/x^2+1/x^2}} = \lim_{x\to+\infty} \frac{1}{\sqrt{1+\frac{1}{x^2}}} = 1.$$

Example 1.114. What is $\lim_{x\to-\infty} \frac{x}{\sqrt{x^2+1}}$?

We can do the same thing, but we have to be *very careful*. Remember that if $x < 0$ then $\sqrt{x^2} \neq x$! Instead, $x = -\sqrt{x^2}$. Thus we have

$$\lim_{x\to-\infty} \frac{x}{\sqrt{x^2+1}} = \lim_{x\to-\infty} \frac{1}{\sqrt{x^2+1}/x} = \lim_{x\to-\infty} \frac{1}{\sqrt{x^2+1}/(-\sqrt{x^2})} = \lim_{x\to-\infty} \frac{1}{-\sqrt{1+\frac{1}{x^2}}} = -1.$$

When we encounter new functions, one of the ways we will often want to characterize them is by computing their limits at $\pm\infty$. Sometimes these limits do not exist.

Example 1.115. $\lim_{x\to+\infty} \sin(x)$ does not exist, since the function oscillates rather than settling down to one limit value.

$\lim_{x\to+\infty} x\sin(x)$ also does not exist; this function oscillates more and more wildly as x increases.

But $\lim_{x\to+\infty} \frac{1}{x}\sin(x)$ does in fact exist. We can prove this with the squeeze theorem: we can see that $\frac{-1}{x} \leq \frac{1}{x}\sin(x) \leq \frac{1}{x}$, and we know that $\lim_{x\to+\infty} \frac{-1}{x} = \lim_{x\to+\infty} \frac{1}{x} = 0$. So by the Squeeze Theorem, $\lim_{x\to+\infty} \frac{1}{x}\sin(x) = 0$.

Another technique that will also often appear in these limits is combining a sum or difference into one fraction. If we have a sum of two terms that both have infinite limits, we need to combine or factor them into one term to see what is happening.

Example 1.116. What is $\lim_{x\to-\infty} x - x^3$?

Each term goes to $-\infty$, so this is a difference of infinities and thus indeterminate. But we can factor: $\lim_{x\to-\infty} x(1-x^2)$. The first term goes to $-\infty$ and the second term also goes to $-\infty$, so we expect that their product will go to $+\infty$. Thus $\lim_{x\to-\infty} x - x^3 = +\infty$.

To be precise, I should compute:

$$\lim_{x\to-\infty} x - x^3 = \lim_{x\to-\infty} \frac{x-x^3}{1} = \lim_{x\to-\infty} \frac{1/x^2 - 1}{1/x^3}.$$

We see the limit of the top is -1 and the limit of the bottom is 0, so the limit of the whole is $\pm\infty$. In fact the bottom will always be negative (since $x \to -\infty$), and thus the limit is $+\infty$.

Example 1.117. What is $\lim_{x\to+\infty} \sqrt{x^2+1} - x$?

We might want to try to use limit laws here, but we would get $+\infty - +\infty$ which is not defined (and is one of the classic indeterminate forms). Instead we need to combine our expressions into one big fraction.

$$\begin{aligned}\lim_{x\to+\infty} \sqrt{x^2+1} - x &= \lim_{x\to+\infty} \left(\sqrt{x^2+1} - x\right) \frac{\sqrt{x^2+1}+x}{\sqrt{x^2+1}+x} \\ &= \lim_{x\to+\infty} \frac{(x^2+1) - x^2}{\sqrt{x^2+1}+x} = \lim_{x\to+\infty} \frac{1}{\sqrt{x^2+1}+x}.\end{aligned}$$

The numerator is 1 and the denominator approaches $+\infty$ so the limit is 0. This tells us that as x increases, x and $\sqrt{x^2+1}$ get as close together as we wish.

You may have noticed the appearance of our old friend, multiplication by the conjugate. We will often use that technique in this sort of problem.

Example 1.118. What is $\lim_{x\to+\infty} \sqrt{x^2+x+1} - x$?

$$\begin{aligned}\lim_{x\to+\infty} \sqrt{x^2+x+1} - x &= \lim_{x\to+\infty} \left(\sqrt{x^2+x+1} - x\right) \frac{\sqrt{x^2+x+1}+x}{\sqrt{x^2+x+1}+x} \\ &= \lim_{x\to+\infty} \frac{x^2+x+1-x^2}{\sqrt{x^2+x+1}+x} = \lim_{x\to+\infty} \frac{x+1}{\sqrt{x^2+x+1}+x} \\ &= \lim_{x\to+\infty} \frac{1+1/x}{\sqrt{1+1/x+1/x^2}+1} = \frac{1}{2}.\end{aligned}$$

2 Derivatives

2.1 Linear Approximation

In the last section we talked about continuous functions as functions that we could approximate. We know that $\sqrt{5}$ is about 2, and 3.1^3 is about 27. In this section we want to be a bit more precise than that. Most of you told me not only that $\sqrt{5}$ is "about 2", but it's a bit *more* than 2. We want to find a way to estimate that bit more.

We need to use a more complicated formula. But we want to keep the amount of complexity under control. So we want to use a simple function to approximate $f(x)$. The simplest possible function is a constant function; and that's exactly what we used last section. (3.1^3 is about 27, and 3.01^3 is about 27, and 3.2^3 is about 27.) If a is a fixed number then $f(a)$ is a constant, and thus $f(x) \approx f(a)$ approximates f with a constant function.

The next most complex function, as we usually think of it, is a linear function. So we want to approximate f with a linear function. There are a few ways we can write the equation for a line, depending on what information we already know:

$$
\begin{aligned}
y &= mx + b && \text{Slope-Intercept Formula} \\
y - y_0 &= m(x - x_0) && \text{Point-Slope Formula} \\
y - y_0 &= \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) && \text{Two Points Formula}
\end{aligned}
$$

The most common and popular is the slope-intercept formula, which is great for *computing* things; but to write down the equation, you need to know the slope m, and also the y-intercept b. For our approximations we won't generally know this.

The two points formula also isn't terribly useful for us. We know one point: since we're approximating a function f near a, we know it goes through the point $(a, f(a))$. But if we knew the value at other points, we wouldn't need to approximate! (The approximation $f(x) - f(a) \approx \frac{f(x)-f(a)}{(x-a)}(x - a)$ is true, but is kind of vacuous and tautological; it doesn't actually help us).

But the point-slope formula can get us somewhere. We already have a point, so we just need to find the slope. We'll see how to do that soon, but for now we'll just give the slope a name: if we're taking a linear approximation to a function $f(x)$ near a point a, then we will denote the slope $f'(a)$. This tells us, essentially, how much we care about the distance between x and a. When this is small, then $f(x)$ is close to $f(a)$; when $f'(a)$ is large, then $f(x)$ moves away from $f(a)$ pretty quickly.

The equation for our linear approximation is

$$f(x) \approx f'(a)(x-a) + f(a) \tag{1}$$

This is the most important formula in the entire course; essentially everything we do for the next two months will refer back to this approximation in some way.

Example 2.1. We earlier said that $\sqrt{5} \approx \sqrt{4} = 2$. We can see that in fact $\sqrt{5}$ should be a little bigger than 2. But how much better?

A linear approximation would tell us that $\sqrt{5} \approx 2 + f'(2)(5-4)$. That is, we know that $\sqrt{5}$ is a bit bigger than two—and it's a bit bigger by the amount of this mysterious $f'(2)$ slope. We'll see how to compute this later, but for right now I'll tell you that $f'(2) = \frac{1}{4}$. Then we get that $\sqrt{5} \approx 2 + \frac{1}{4}(5-4) = 9/4 = 2.25$.

From this we can make other estimates. For instance, we have that $\sqrt{4.5} \approx 2+\frac{1}{4}(4.5-4) = 17/8$, and $\sqrt{6} \approx 2 + \frac{1}{4}(6-4) = 5/2$.

We can go in the other direction as well. We estimate that $\sqrt{3} \approx 2 + \frac{1}{4}(3-4) = 7/4$. And $\sqrt{2} \approx 2 + \frac{1}{4}(2-4) = 3/2$.

But notice: this gives us $\sqrt{1} \approx 2 + \frac{1}{4}(1-4) = 5/4$, which we know is wrong. And $\sqrt{9} \approx 2+\frac{1}{4}(9-4) = 13/4$, which is also wrong. For that matter, we get $\sqrt{100} \approx 2+\frac{1}{4}(100-4) = 26$, which is really wrong. What's going on here?

A linear approximation is good when x is close to $a = 2$. As x gets further away from a, then our estimate for $f(x)$ gets further from $f(a)$; but in general we would also expect our estimate to get further from the correct answer. These techniques work best when x is very close to a.

(We're not yet ready to be precise about what "very close" means here).

Example 2.2. We've dressed this up in fancy language, but we engage in this sort of reasoning all the time. Suppose you are driving at 30 miles per hour. After an hour, you expect to have gone about thirty miles. After six minutes, you expect to have gone about three.

This is just a linear approximation. If $f(t)$ is our position as a function of time, our approximation is that we're moving 30 miles per hour, or half a mile per minute. Then we have $f(t) \approx 0 + \frac{1}{2}(t-0)$, and if we plug in $t = 6$ we have $f(6) \approx 0 + \frac{1}{2}(6-0) = 3$.

2.2 The Derivative

We understand that we want to do linear approximation now. But without a way to actually find the slope $f'(a)$, it isn't terribly helpful.

So let's look at our formula from equation (3) again. We want to understand $f'(a)$, so we'll solve the equation for that:

$$\begin{aligned} f(x) &\approx f'(a)(x-a) + f(a) \\ f(x) - f(a) &\approx f'(a)(x-a) \\ \frac{f(x)-f(a)}{x-a} &\approx f'(a). \end{aligned}$$

Thus we get a new formula. This formula should also make sense to us. The slope $f'(a)$ tells us how different $f(x)$ is from $f(a)$, based on how x is different from a. This new, rearranged formula tells us that $f'(a)$ approximates the ratio of the change in $f(x)$ to the change in x, which we sometimes write as $\frac{\Delta f}{\Delta x}$. Thus it should tell us how much a change in the input value affects the output value—which is exactly the question we need to answer to write a linear approximation.

But we've also seen this formula somewhere else. In the two points formula for a line, the slope is $\frac{y_1-y_0}{x_1-x_0}$. if $y_1 = f(x_1) = f(x)$ and $y_0 = f(x_0) = f(a)$, then this is just the approximation we have for $f'(a)$. Thus we're saying that $f'(a)$ is approximately the slope of the line through the point $(a, f(a))$ that we know, and the point $(x, f(x))$ that we want. We'll explore this angle more in lab.

On its own, this still isn't helpful: we have an approximate formula for $f'(a)$, but it requires us to already know $f(x)$, which is what we started out wanting to compute. But one more step makes this actually useful.

Definition 2.3. Let f be a function defined near and at a point a. We say the *derivative* of f at a is

$$f'(a) = \lim_{x \to a} \frac{f(x)-f(a)}{x-a} = \lim_{h \to 0} \frac{f(a+h)-f(a)}{h}.$$

The second formula is just a change of variables from the first, setting $h = x - a$. It's not substantively any different, but it's sometimes easier to compute with.

We will also sometimes write $\frac{df}{dx}(a)$ for the derivative of f at a. The is called "Leibniz notation", as opposed to the "Newtonian notation" of $f'(a)$.

Thus the derivative is given by taking our approximate formula for $f'(a)$, and taking the limit as x and a get closer together. Our linear approximation is better when x and a are closer; so as x approaches a, the approximation becomes perfect, and we get an exact equation.

Remark 2.4. Note that we need *two* pieces of information here. You hand me a function f *and* a point a, and I tell you the derivative of f at a. We'll adopt different perspectives from time to time later on in the course.

Example 2.5.

1. Let $f(x) = x^2 + 1$. Then
$$f'(2) = \lim_{h \to 0} \frac{f(2+h) - f(2)}{h} = \lim_{h \to 0} \frac{(2+h)^2 + 1 - 2^2 - 1}{h} = \lim_{h \to 0} \frac{4h + h^2}{h} = 4,$$
and more generally, for any number a we have
$$f'(a) = \lim_{h \to 0} \frac{(a+h)^2 - a^2}{h} = \lim_{h \to 0} \frac{2ah + h^2}{h} = 2a.$$

2. Let $f(x) = x^3$, and let's find the derivative at a point a. Then
$$\begin{aligned} f'(a) &= \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a} \frac{x^3 - a^3}{x - a} \\ &= \lim_{x \to a} \frac{(x-a)(x^2 + ax + a^2)}{x - a} = \lim_{x \to a} x^2 + ax + a^2 = 3a^2. \end{aligned}$$
Notice that it wasn't obvious that we could factor $x^3 - a^3$ this way. We could notice this by noticing that plugging in a gives us zero; in general, if plugging a into a polynomial gives zero, we can always factor out a $(x - a)$ term. In this case, though, it might have been easier to just start with the limit as $h \to 0$, in which case the problem would have essentially solved itself.

3. Let $f(x) = \sqrt{x}$. Then given a number a, we have
$$f'(a) = \lim_{h \to 0} \frac{\sqrt{a+h} - \sqrt{a}}{h} = \lim_{h \to 0} \frac{(a+h) - a}{h(\sqrt{a+h} + \sqrt{a})} = \lim_{h \to 0} \frac{1}{\sqrt{a+h} + \sqrt{a}} = \frac{1}{2\sqrt{a}}$$
Note that f is defined at 0, and we have $f(0) = 0$. But by this computation we have $f'(0) = \frac{1}{2 \cdot 0}$ which is undefined. This isn't an artifact of the way we computed it; the limit in fact does not exist. Further, this isn't just becasue 0 is on the edge of the domain of f, as we shall see:

4. Let $g(x) = \sqrt[3]{x}$. Then we can compute $g'(0)$ and we get
$$g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h} = \lim_{h \to 0} \frac{\sqrt[3]{h}}{h} = \lim_{h \to 0} \frac{1}{\sqrt[3]{h^2}} = +\infty.$$

The cube root function g has no defined derivative at 0, even though the function is defined there. This brings us to a discussion of ways for a function to fail to be differentiable at a point. (There's always the catchall category of "the limit just doesn't exist," which we won't really discuss because there's not much to say about it).

Example 2.6. 1. Our first example of $g(x) = \sqrt[3]{x}$ is not differentiable at 0, and the limit

$$g'(0) = \lim_{h \to 0} \frac{g(h) - g(0)}{h} = +\infty.$$

Graphically, the line tangent to g at 0 is completely vertical; the function is "increasing infinitely fast" at 0.

2. Any function that is not continuous at a point cannot be differentiable at that point. In particular, if f is differentiable at a, then

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

converges. But the bottom goes to zero, so the top must also go to zero, and we have

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

which is precisely waht it means to be continuous.

Conceptually, if the function isn't continuous, it isn't changing smoothly and so doesn't have a "speed" of change. Graphically, a function that has a disconnect in it doesn't have a clear tangent line.

An example here is the Heaviside function $H(x)$. We have

$$\lim_{h \to 0^+} \frac{H(h) - H(0)}{h} = \lim_{h \to 0^+} \frac{0}{h} = 0$$

but

$$\lim_{h \to 0^-} \frac{H(h) - H(0)}{h} = \lim_{h \to 0^-} \frac{-1}{h} = +\infty.$$

Since the one-sided limits aren't equal, the limit does not exist.

3. Any function with a sharp corner at a point doesn't have a well-defined rate of change at that point; the change is instantaneous. For instance, if we let $a(x) = |x|$ be the absolute value function, then

$$a'(x) = \lim_{h \to 0} \frac{a(x+h) - a(x)}{h}.$$

To study piecewise functions we usually break them up and study each piece separately. If $x > 0$, then $a(x) = x$ and $a(x+h) = x+h$ for small h. We have

$$a'(x) = \lim_{h \to 0} \frac{x+h-x}{h} = \lim_{h \to 0} 1 = 1.$$

Conversely, if $x < 0$ then $a(x) = -x$ and $a(x+h) = -x-h$, and

$$a'(x) = \lim_{h \to 0} \frac{-x-h+x}{h} = \lim_{h \to 0} -1 = -1.$$

But if $x = 0$ then the left and right limits don't agree again: the right limit is 1 and the left limit is -1, so the limit does not exist. Thus we have

$$a'(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ \text{undefined} & x = 0. \end{cases}$$

4. Sometimes a function has a "cusp" at a point. This is a point where the tangent line is vertical, but depending on the side from which you approach, you can get a tangent line that goes up incredibly fast or one that goes down incredibly fast.

   Consider the funtion $f(x) = \sqrt[3]{x^2}$. We have

$$f'(0) = \lim_{h \to 0} \frac{\sqrt[3]{h^2} - \sqrt[3]{0}}{h} = \lim_{h \to 0} \frac{h^{2/3}}{h} = \lim_{h \to 0} \frac{1}{\sqrt[3]{h}} = \pm\infty.$$

   This is different from the $\sqrt[3]{x}$ example because the limit is $\pm\infty$ rather than just $+\infty$.

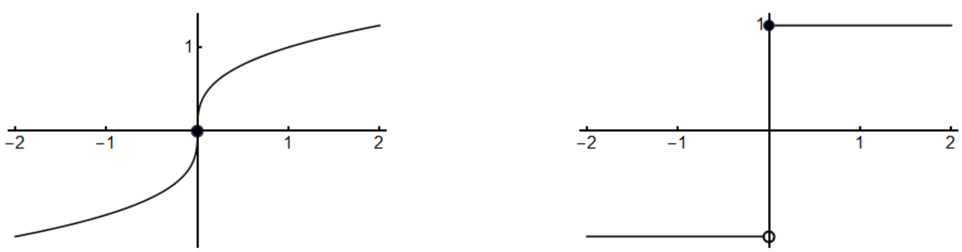


Figure 2.1: A vertical tangent line and a discontinuous function

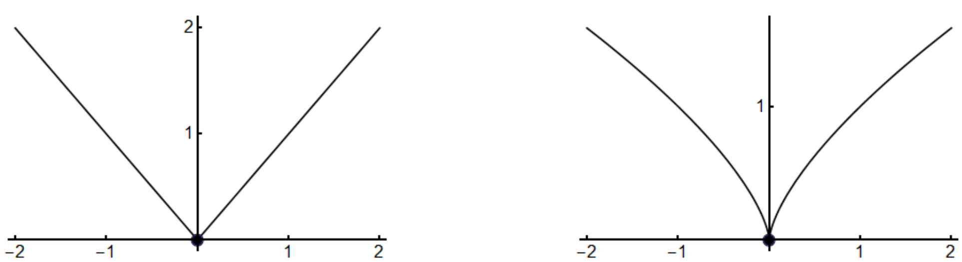


Figure 2.2: A corner and a cusp

Example 2.7. Let $f(x) = \sqrt{x^2 - 4}$. What is $f'(x)$? Where is f differentiable?

$$\begin{aligned}
f'(x) &= \lim_{h \to 0} \frac{\sqrt{(x+h)^2 - 4} - \sqrt{x^2 - 4}}{h} \\
&= \lim_{h \to 0} \frac{(x+h)^2 - 4 - (x^2 - 4)}{h(\sqrt{(x+h)^2 - 4} + \sqrt{x^2 - 4})} \\
&= \lim_{h \to 0} \frac{2xh + h^2}{h(\sqrt{(x+h)^2 - 4} + \sqrt{x^2 - 4})} \\
&= \lim_{h \to 0} \frac{2x + h}{(\sqrt{(x+h)^2 - 4} + \sqrt{x^2 - 4})} \\
&= \frac{2x}{2\sqrt{x^2 + 4}} = \frac{x}{\sqrt{x^2 - 4}}.
\end{aligned}$$

Thus we see that f is differentiable on $(-\infty, -2) \cup (2, +\infty)$.

Our computation of the derivative of $|\cdot|$, and of several other functions, looks a lot like a function itself. Taking the derivative of a function f in fact gives us a new function f': the rule of this function is that given a number a, we compute the derivative of f at a and return that as our output. Thus f' is a function and we can study it the way we did earlier functions.

Definition 2.8. The *derivative of a function* f is the function that takes in an input x and outputs

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

Example 2.9.

1. If $f(x) = x^2 + 1$, we computed that $f'(x) = 2x$. The domain of f is all reals, and so is the domain of $f'(x)$.

2. If $g(x) = \sqrt{x}$ then $g'(x) = \frac{1}{2\sqrt{x}}$. The domain of g is all reals ≥ 0, and the domain of g' is all reals > 0.

3. We saw above that if $a(x) = |x|$, then

$$a'(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ \text{undefined} & x = 0 \end{cases} = \frac{|x|}{x}.$$

   The domain of a is all reals and the domain of a' is all reals except 0.

Further, since f' is a function we can ask about the derivative of f' at a point a.

Definition 2.10. Let f be a function which is differentiable at and near a point a. The *second derivative of* f *at* a is the derivative of the function $f'(x)$ at a, which is

$$f''(a) = \lim_{h \to 0} \frac{f'(a+h) - f'(x)}{h} = \frac{d^2 f}{dx^2}(a).$$

This is again a limit and may or may not exist.

Remark 2.11. The Leibniz notation for a second derivative is $\frac{d^2f}{dx^2}$ and not $\frac{df^2}{dx^2}$. Conceptually, you can think of $\frac{d}{dx}$ as a function whose input is the function f and whose output is the derivative function f'. The second derivative results from applying this function twice.

Example 2.12. What is the second derivative of $f(x) = x^3$ at $a = 2$?

$$f'(x) = \lim_{h \to 0} \frac{(x+h)^3 - x^3}{h} = \lim_{h \to 0} \frac{3x^2h + 3h^2 + h^3}{h} = \lim_{h \to 0} 3x^2 + 3h + h^2 = 3x^2.$$

$$\begin{aligned} f''(2) &= \lim_{h \to 0} \frac{f'(2+h) - f'(2)}{h} = \lim_{h \to 0} \frac{3(2+h)^2 - 3 \cdot 2^2}{h} = \lim_{h \to 0} \frac{3(4 + 4h + h^2) - 12}{h} \\ &= \lim_{h \to 0} \frac{12h + 3h^2}{h} = \lim_{h \to 0} 12 + 3h = 12. \end{aligned}$$

We won't say much more about the second derivative now, but we'll discuss it extensively in section 3.

2.3 Computing Derivatives

By now we're getting pretty tired of computing those examples over and over. In this section we'll come up with some techniques to make computation of derivatives easier.

1. If c is a constant and $f(x) = c$ then $f'(x) = 0$.

   $$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{c - c}{h} = \lim_{h \to 0} 0 = 0.$$

   Conceptually, a constant function never changes, so the rate of change is 0.

   Geometrically, a constant function is a horizontal line; thus we think of the slope everywhere as being 0.

   Example 2.13. $(3^{3^{3^3}})' = 0$.

2. If $f(x) = x$, then $f'(x) = 1$.

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{x + h - x}{h} = \lim_{h \to 0} 1 = 1.$$

   Conceptually, if we have the "identity" function, then whenever we change the input then the output should change by exactly the same amount. Thus the rate of change is 1.

   Geometrically, this is a line with slope 1.

3. If c is a constant and g is a function and $f(x) = c \cdot g(x)$, then $f'(x) = c(g'(x))$.

$$f'(x) = \lim_{h \to 0} \frac{cg(x+h) - cg(x)}{h} = c \cdot \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} = c \cdot g'(x).$$

   Conceptually, if changing x by a bit changes $g(x)$ by a certain amount, then it will change $2g(x)$ by twice that amount–multiplying by a scalar should just change the rate of change by the same amount everywhere.

   Geometrically, multiplying by a constant is just stretching vertically–and all the slopes will be stretched by that same amount.

   Example 2.14. If $f(x) = 5x$ then $f'(x) = (5 \cdot x)' = 5 \cdot x' = 5$.

4. If f and g are functions then $(f + g)'(x) = f'(x) + g'(x)$.

$$\begin{aligned} (f+g)'(x) &= \lim_{h \to 0} \frac{f(x+h) + g(x+h) - f(x) - g(x)}{h} \\ &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} + \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} = f'(x) + g'(x). \end{aligned}$$

   Conceptually, if changing the input by a bit changes f by a certain amount and g by a different amount, then it changes $f + g$ by the sum of those two amounts–figure out how much it changes each part and then add them together to find out how much it changes the whole.

   Geometrically, if we add two functions together it's just like stacking them on top of one another, so the slope at any point will be the sum of the slopes.

   Example 2.15. Let $f(x) = 3x - 7$. Then $f'(x) = (3x)' - 7' = 3(x') - 0 = 3$.

   This rule is really important but so far we can't do much with it–we don't have quite enough rules yet.

5. (Power Rule) If $f(x) = x^n$ where n is a positive integer, then $f'(x) = nx^{n-1}$. In fact, if $g(x) = x^r$ and r is any real number, then $g'(x) = rx^{r-1}$. We'll only prove this for integers, using the difference-of-nth-powers rule.

$$f'(x) = \lim_{z \to x} \frac{z^n - x^n}{z - x} = \lim_{z \to x} \frac{(z-x)(z^{n-1} + z^{n-2}x + \cdots + zx^{n-2} + x^{n-1})}{z - x}$$
$$= \lim_{z \to x} z^{n-1} + z^{n-2}x + \cdots + zx^{n-2} + x^{n-1} = x^{n-1} + \cdots + x^{n-1} = nx^{n-1}.$$

Now that we have this, we can compute all sorts of derivatives.

Example 2.16.
- $(x^2 + 1)' = 2x + 0 = 2x$.
- $(\sqrt{x})' = (x^{1/2})' = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}}$.
- $(\sqrt[3]{x})' = (x^{1/3})' = \frac{1}{3}x^{-2/3} = \frac{1}{3\sqrt[3]{x^2}}$.
- $(3\sqrt{x} + x^5 - 7)' = \frac{3}{2\sqrt{x}} + 5x^4 + 0$.

6. (Product Rule) If f and g are functions then $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$.

   Conceptually, we sort of know this already; if we add a bit on to f and a bit on to g, then we get $(f + f_h)(g + g_h) = fg + fg_h + gf_h + g_h f_h$, and in the limit we can treat $g_h f_h$ as being zero. So this is the same as multiplying the bit we add to g with f, and multiplying the bit we add to f with g, and then adding the two.

   Example 2.17. $((3x-2)(x-1))' = (3x^2 - 5x + 2)' = 6x - 5$.

   Alternatively, $((3x-2)(x-1)' = (3x-2)'(x-1) + (3x-2)(x-1)' = 3 \cdot (x-1) + 1 \cdot (3x-2) = 6x - 5$.

   This rule isn't terribly important as long as we're only working with rational functions. Once we include anything else, like trig functions, it is critical.

   Remark 2.18. We can get the power rule from the product rule instead of trying to get it directly.

7. (Quotient Rule): If f and g are functions then

$$(f/g)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}.$$

$$\begin{aligned}
(f/g)'(x) &= \lim_{h\to 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x+h)}{g(x+h)g(x)h} \\
&= \lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x) + f(x)g(x) - f(x)g(x+h)}{g(x+h)g(x)h} \\
&= \lim_{h\to 0} \frac{1}{g(x+h)g(x)} \left(\lim_{h\to 0} \frac{f(x+h)g(x) - f(x)g(x)}{h} + \lim_{h\to 0} \frac{f(x)g(x) - f(x)g(x+h)}{h} \right) \\
&= \frac{1}{g(x)^2} \left(g(x) \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} - f(x) \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} \right) \\
&= \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}
\end{aligned}$$

Example 2.19.

- $\left(\frac{x-1}{x^3}\right)' = (x^{-2} - x^{-3})' = -2x^{-3} + 3x^{-4}$.

  Alternatively,

  $$\left(\frac{x-1}{x^3}\right)' = \frac{(x-1)'x^3 - (x-1)3x^2}{x^6} = \frac{x^3 - 3x^3 + 3x^2}{x^6} = -2x^{-3} + 3x^{-4}.$$

- $\left(\frac{2+3x}{3-5x}\right)' = \frac{(2+3x)'(3-5x) - (2+3x)(3-5x)'}{(3-5x)^2} = \frac{9 - 15x + 10 + 15x}{(3-5x)^2} = \frac{19}{(3-5x)^2}$

2.4 Trigonometric derivatives

We cannot neglect the trigonometric functions—no matter how much we might wish to on occasion. All of the rules for trigonometric derivatives rely on what are known as the *angle addition formulas*:

$$\sin(a+b) = \sin(a)\cos(b) + \cos(a)\sin(b) \qquad \cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b).$$

Note: you probably won't ever need to know these formulas again in this class. But I will need them for another page or so of these notes.

Using this we can compute

1.
$$\begin{aligned}
(\sin(x))' &= \lim_{h\to 0}\frac{\sin(x+h)-\sin(x)}{h} = \lim_{h\to 0}\frac{\sin(x)\cos(h)+\sin(h)\cos(x)-\sin(x)}{h}\\
&= \left(\lim_{h\to 0}\frac{\sin(h)\cos(x)}{h}\right) + \left(\lim_{h\to 0}\frac{\sin(x)(\cos(h)-1)}{h}\right)\\
&= \cos(x)\lim_{h\to 0}\frac{\sin h}{h} + \sin(x)\lim_{h\to 0}\frac{\cos h - 1}{h}\\
&= \cos(x) + \sin(x)\lim_{h\to 0}\frac{\cos^2(h)-1}{h(\cos(h)+1)}\\
&= \cos(x) - \sin(x)\lim_{h\to 0}\frac{\sin^2(h)}{h(\cos(h)+1)}\\
&= \cos(x) - \sin(x)\left(\lim_{h\to 0}\frac{\sin(h)}{\cos(h)+1}\right)\left(\lim_{h\to 0}\frac{\sin h}{h}\right)\\
&= \cos(x) - \sin(x)\cdot 0\cdot 1 = \cos(x).
\end{aligned}$$

2. A similar argument shows that $(\cos(x))' = -\sin(x)$.

Further using the product and quotient rules, we observe that

- $$(\tan(x))' = \left(\frac{\sin x}{\cos x}\right)' \frac{\cos^2(x)+\sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)} = \sec^2(x)$$
- $$(\cot(x))' = \left(\frac{\cos x}{\sin x}\right)' = \frac{-\sin^2(x)-\cos^2(x)}{\sin^2(x)} = \frac{-1}{\sin^2(x)} = -\csc^2(x)$$
- $$(\sec(x))' = \left(\frac{1}{\cos x}\right)' = \frac{0+\sin x}{\cos^2(x)} = \frac{\sin x}{\cos x}\cdot\frac{1}{\cos x} = \sec(x)\tan(x)$$
- $$(\csc(x))' = \left(\frac{1}{\sin x}\right)' = \frac{0-\cos(x)}{\sin^2(x)} = \frac{-\cos x}{\sin x}\cdot\frac{1}{\sin x} = -\csc(x)\cot(x).$$

Remember that as long as you know the derivatives of sin and cos you can always compute these four derivatives whenever you need them.

Example 2.20. 1. If $f(t) = 3\sin t + \cos t$, then $f'(t) = 3\cos t - \sin t$.

2. Find the tangent line to $y = 6\cos x$ at $(\pi/3, 3)$.

   We see that $y' = -6\sin x$, and thus when $x = \pi/3$ we have $y' = -3\sqrt{3}$. Recalling that the equation of our line is $y = m(x - x_0) + f(x_0)$, we have the equation $y = -3\sqrt{3}(x - \pi/3) + 3$.

3. IF $g(\theta) = \theta \sin\theta + \frac{\cos\theta}{\theta}$, then
$$g'(\theta) = (\sin\theta + \theta\cos\theta) + \frac{-\theta\sin\theta - \cos\theta}{\theta^2}.$$

4. If $h(x) = \frac{x}{2-\tan x}$, then
$$h'(x) = \frac{(2-\tan x) + x\sec^2 x}{(2-\tan x)^2}.$$

5. We can also compute second derivatives. $\sin'' x = -\sin x$. $\cos'' x = -\cos x$.
$$\tan'' x = (\sec x \sec x)' = \sec x \tan x \sec x + \sec x \tan x \sec x = 2\sec^2 x \tan x.$$

2.5 The Chain Rule

To start with an example, suppose $g(x) = (\sin x)^2$. Then
$$g'(x) = ((\sin x)(\sin x))' = \cos x \sin x + \cos x \sin x = 2 \sin x \cos x.$$
Remembering that $(x^2)' = 2x$, we notice that this looks suggestive. It also leads us to ask what happens when we build up functions by composition, that is, plugging one function into another, as we have here.

If we want to freely build complex functions from simple ones, we need to be able to combine them in chains. Remember that we define the function $f \circ g$ by $(f \circ g)(x) = f(g(x))$; we take our input x, plug it into g, and then take the output $g(x)$ and plug it into f.

We can see how this is useful in two different ways. First, as we saw earlier, it lets us build up functions.

1. $(x+1)^2 = (f \circ g)(x)$ where $g(x) = x + 1$ and $f(x) = x^2$.

2. $(x^2+1)^2 = (f \circ g)(x)$ where $g(x) = x^2 + 1$ and $f(x) = x^2$.

3. $\sin^2(x) = (f \circ g)(x)$ where $g(x) = \sin x$ and $f(x) = x^2$.

Second, sometimes composition of functions really is the best way to describe what's going on, especially when you have a "causal chain" where one process causes a second which causes a third. For instance, suppose you're driving up a mountain at 2 km/hr, and the temperature drops 6.5° C per kilometer of altitude. You can think about your temperature as a function of your height, which is itself a function of the time; then the numbers I gave you are the rates of change, or derivatives, of each function.

It's not that hard to convince yourself that you'll get colder by about 13° C per hour. Does this work in general?

Proposition 2.21 (Chain Rule)**.** *Suppose f and g are functions, such that g is differentiable at a and f is differentiable at $g(a)$. Then $(f \circ g)'(a) = f'(g(a)) \cdot g'(a)$.*

Proof.

$$\begin{aligned}
(f \circ g)'(a) &= \lim_{h \to 0} \frac{(f \circ g)(a+h) - (f \circ g)(a)}{h} \\
&= \lim_{h \to 0} \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} \cdot \frac{g(a+h) - g(a)}{h} \\
&= \left(\lim_{h \to 0} \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} \right) \left(\lim_{h \to 0} \frac{g(a+h) - g(a)}{h} \right) \\
&= f'(g(a)) \cdot g'(a).
\end{aligned}$$

□

Remark 2.22.

1. When we write $f'(g(x))$, we mean the function f' evaluated at the point $g(x)$, or in other words, the derivative of f at the point $g(x)$.

2. It can be helpful as a way of remembering the chain rule that

$$\frac{d(f \circ g)}{dx} = \frac{d(f \circ g)}{dg} \cdot \frac{dg}{x}.$$

Don't take this too seriously as actively meaning anything, since it only sort of does, but it's quite helpful for the memory.

Example 2.23.

1. $(x+1)^2 = (f \circ g)(x)$ where $g(x) = x + 1$ and $f(x) = x^2$. Then $f'(x) = 2x$ and $g'(x) = 1$, so

$$(f \circ g)'(x) = f'(g(x)) \cdot g'(x) = 2(g(x)) \cdot 1 = 2(x+1) \cdot 1 = 2x + 2.$$

Sanity check:

$$(f \circ g)'(x) = (x^2 + 2x + 1)' = 2x + 2.$$

2. $(x^2+1)^2 = (f \circ g)(x)$ where $g(x) = x^2 + 1$ and $f(x) = x^2$. Then $f' = 2x$, $g' = 2x$, and

$$(f \circ g)'(x) = f'(g(x)) \cdot g'(x) = 2(g(x)) \cdot 2x = 2(x^2+1) \cdot 2x = 4x^3 + 4x.$$

Sanity check:

$$(f \circ g)'(x) = (x^4 + 2x^2 + 1)' = 4x^3 + 4x.$$

3. $\sin^2(x) = (f \circ g)(x)$ where $g(x) = \sin x$ and $f(x) = x^2$. Then $f'(x) = 2x$, $g'(x) = \cos x$, and we have

$$(f \circ g)'(x) = 2(g(x)) \cdot \cos x = 2(\sin x) \cos x.$$

4. $\cos(3x) = (f \circ g)(x)$ where $f(x) = \cos(x)$ and $g(x) = 3x$. Then $f'(x) = -\sin(x)$ and $g'(x) = 3$ and
$$(f \circ g)'(x) = -\sin(3x) \cdot 3.$$

5. $\sin(x^2) = (f \circ g)(x)$ where $f(x) = \sin(x)$ and $g(x) = x^2$. Then $f'(x) = \cos x, g'(x) = 2x$, and
$$(f \circ g)'(x) = \cos(g(x)) \cdot 2x = 2x\cos(x^2).$$

6. If $f(x)$ is any function, then we can write $(f(x))^r$ as $(g \circ f)(x)$ where $g(x) = x^r$. Then
$$\frac{d}{dx}(f(x)^r = (g \circ f)'(x) = r(f(x))^{r-1} \cdot f'(x).$$

7. The derivative of $\sec(5x)$ is $\sec(5x)\tan(5x)5$.

8. What is the derivative of $\frac{1}{\sqrt[3]{x^4-12x+1}}$? We can view this as $(x^4 - 12x + 1)^{-1/3}$, and using the chain rule, we have
$$\frac{d}{dx}\frac{1}{\sqrt[3]{x^4 - 12x + 1}} = \frac{-1}{3}(x^4 - 12x + 1)^{-4/3} \cdot (4x^3 - 12).$$

9. What is the derivative of $\sec^2(x)$? By the chain rule this is $2 \cdot \sec(x) \cdot \sec'(x) = 2\sec(x) \cdot \sec(x)\tan(x) = 2\sec^2(x)\tan(x)$.

10. What is the derivative of $\sec^4(x)$? We get $4\sec^3(x)\sec'(x) = 4\sec^3(x)\sec(x)\tan(x) = 4\sec^4(x)\tan(x)$.

11. Sometimes we have to nest the chain rule. What is the derivative of $\sqrt{x^3 + \sqrt{x^2+1}}$? We can pull this apart slowly.
$$\begin{aligned}\frac{d}{dx}\sqrt{x^3 + \sqrt{x^2+1}} &= \frac{1}{2}(x^3 + \sqrt{x^2+1})^{-1/2} \cdot \left(\frac{d}{dx}\left(x^3 + \sqrt{x^2+1}\right)\right) \\ &= \frac{1}{2\sqrt{x^3 + \sqrt{x^2+1}}}\left(3x^2 + \frac{1}{2}(x^2+1)^{-1/2} \cdot \left(\frac{d}{dx}x^2+1\right)\right) \\ &= \frac{3x^2 + \frac{2x}{2\sqrt{x^2+1}}}{2\sqrt{x^3 + \sqrt{x^2+1}}}\end{aligned}$$

As we have just seen the chain rule can stack, or chain together. As functions get more complicated we will have to use multiple applications of the product rule, quotient rule, and chain rule to pull our derivative apart.

Example 2.24. Find

$$\frac{d}{dx}\sec(x^2+\sqrt{x^3+1}).$$

$$\begin{aligned}\frac{d}{dx}\sec(x^2+\sqrt{x^3+1}) &= \sec(x^2+\sqrt{x^3+1})\cdot\tan(x^2+\sqrt{x^3+1})\cdot(x^2+\sqrt{x^3+1})' \\ &= \sec(x^2+\sqrt{x^3+1})\cdot\tan(x^2+\sqrt{x^3+1})\cdot(2x+\frac{1}{2}(x^3+1)^{-1/2}\cdot 3x^2)\end{aligned}$$

Example 2.25. Find

$$\frac{d}{dx}\frac{\sin(x^2)+\sin^2(x)}{x^2+1}$$

$$\begin{aligned}\frac{d}{dx}\frac{\sin(x^2)+\sin^2(x)}{x^2+1} &= \frac{(\sin(x^2)+\sin^2(x))'(x^2+1)-2x(\sin(x^2)+\sin^2(x))}{(x^2+1)^2} \\ &= \frac{(\cos(x^2)\cdot 2x+2\sin(x)\cos(x))(x^2+1)-2x(\sin(x^2)+\sin^2(x))}{(x^2+1)^2}.\end{aligned}$$

We can keep going with increasingly complicated problems, basically until we get bored. These are really good practice for making sure you understand how the rules fit together.

Example 2.26. Find

$$\frac{d}{dx}\sqrt{\frac{\sqrt{x}+1}{(\cos x+1)^2}}$$

$$\begin{aligned}\frac{d}{dx}\sqrt{\frac{\sqrt{x}+1}{(\cos x+1)^2}} &= \frac{1}{2}\left(\frac{\sqrt{x}+1}{(\cos x+1)^2}\right)^{-1/2}\cdot\left(\frac{\sqrt{x}+1}{(\cos x+1)^2}\right)' \\ &= \frac{1}{2}\left(\frac{\sqrt{x}+1}{(\cos x+1)^2}\right)^{-1/2}\cdot\frac{\frac{1}{2}x^{-1/2}(\cos x+1)^2-2(\cos x+1)(-\sin x)(\sqrt{x}+1)}{(\cos x+1)^4}\end{aligned}$$

Example 2.27. Calculate

$$\frac{d}{dx}\left(\frac{\sin^2\left(\frac{x^2+1}{\sqrt{x-1}}\right)+\sqrt{x^3-2}}{\cos(\sqrt{x^2+1}+1)-\tan(x^4+3)}\right)^{5/3}$$

2.6 Tangent Lines and Linear Approximation

In section 2.1 we defined the derivative in terms of approximation. We took an *algebraic* approach where we wanted to approximate a function with a line, and found a number $f'(a)$ that made the line $y=f'(a)(x-a)+f(a)$ approximate the function f as well as possible.

In this section we want to return to this idea, now that we know how to compute derivatives. We'll also look at how to interpret this same idea in a *geometric* way, where we can understand what the derivative means when we look at the graph of a function. In the next section 2.7 we'll separately take a *physical* perspective, where we see how Now we want to approach the derivative a third way, from a physical perspective. We'll see that the derivative represents the rate at which some quantity changes.

We know that if we have a function $f(x)$ and know what it looks like at a point a, we can use the derivative to give a linear approximation

$$f(x) \approx f(a) + f'(a)(x-a).$$

Last lab, we drew secant lines, which are lines that intersect the graph of a function in (at least) two points; we may recall that by rearranging some information, we can write

$$f(x) = \frac{f(x)-f(a)}{x-a}(x-a) + f(a).$$

As x approaches a, this becomes closer to being a tangent line, and the slope term becomes closer to $f'(a)$. Thus we can get a decent approximation, if x and a are close, by replacing this difference quotient with the derivative:

$$f(x) \approx f'(a)(x-a) + f(a).$$

Example 2.28. We can find an estimate of 2.1^5.

We take $f(x) = x^5$ and $a = 2$. Then $f'(x) = 5x^4$, so we have $f(2) = 32, f'(2) = 80$, and

$$f(2.1) \approx 80(2.1-2) + 32 = 40.$$

The exact answer is 40.841. We can see our original estimate as the constant term in our tangent line approximation.

What if we approximate $(2.5)^5$ using $a = 2$. Approximate 3^5 using $a = 2$. Are your approximations getting better or worse? Why? What does this tell you about what counts as "close" to 2?

We have

$$(2.5)^5 \approx 80 \cdot (2.5-2) + 32 = 72$$
$$3^5 \approx 80 \cdot (3-2) + 32 = 112.$$

The true answers are 97.6563 and 243. Unlike in part (a), these estimates are not especially good. This is because 3 is actually not very close to 2—especially proportionately. Of course, it's not that hard to compute 3^5 directly.

These methods are best when $x - a$ is very small relative to everything else. We often use them in the real world for $x - a < .1$ or so.

Example 2.29. Let's approximate $\sqrt[3]{28}$ and $\sqrt[4]{82}$.

We take $a = 27$ and $a = 81$ respectively.

$$\sqrt[3]{28} \approx \frac{1}{3}(27)^{-2/3}(28-27) + 3 = \frac{1}{27} + 3 \approx 3.03704$$
$$\sqrt[4]{82} \approx \frac{1}{4}(81)^{-3/4}(82-81) + 3 = \frac{1}{108} + 3 \approx 3.00926.$$

The true answers are approximately 3.03659 and 3.00922 respectively.

Now we'll approximate 28^3 and 82^4 using the same base points

We have

$$28^3 \approx 3(27)^2(28-27) + 27^3 = 21870$$
$$82^4 \approx 4(81)^3(82-81) + 81^4 = 45172485$$

In contrast the true answers are 21952 and 45172485.

These approximations aren't *terrible* but they aren't very good either. Since the derivative is changing quickly here (the second derivatives are $6 \cdot 27$ and $12 \cdot 81^2$ respectively), the approximation won't be very good.

Example 2.30. If you take $a = 0$ and $f(x) = x^{10}$, use a tangent line to approximate $f(2)$. What happens and why? What if you instead approximate with $a = 1$?

We have $f'(x) = 10x^9$, so we have $f'(0) = 0$, and thus

$$f(2) \approx 0(2-0) + 0 = 0.$$

If we take $a = 1$, we have

$$f(2) \approx 10(2-1) + 1 = 11.$$

The true answer is 1024, which is far away from both of those. In essence, the derivative is changing so quickly that the tangent line approximation is not very good over those distances.

We can also interpret all of this geometrically. Recall that there are a few different ways to write an equation for a line:

$$y = mx + b \qquad \text{Slope-Intercept Formula}$$
$$y - y_0 = m(x - x_0) \qquad \text{Point-Slope Formula}$$
$$y - y_0 = \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) \qquad \text{Two Points Formula}$$

In order to write down the equation of a line, we need two pieces of information: either we need two points, or we need one point and also the slope. So far, we've used information about $f(a)$ to find the slope $f'(a)$.

But what if we know the values of $f(a)$ and also $f(b)$? Then we can use the two-points formula to write the equation of a line through those two points:

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

And this is *almost* the linear approximation formula, since $f'(a) \approx \frac{f(b)-f(a)}{b-a}$. As b gets closer to a, this will get closer and closer to being the linear approximation formula.

So what does this mean geometrically? First we should talk about two types of lines from a geometric perspective.

Definition 2.31. A line that touches a curve at one point without crossing it is *tangent* to the curve at that point, and we call such a line a *tangent line* (from Latin *tangere* "to touch".)

A line crossing a curve in two points is called a *secant* line. (from Latin *secare* "to cut").

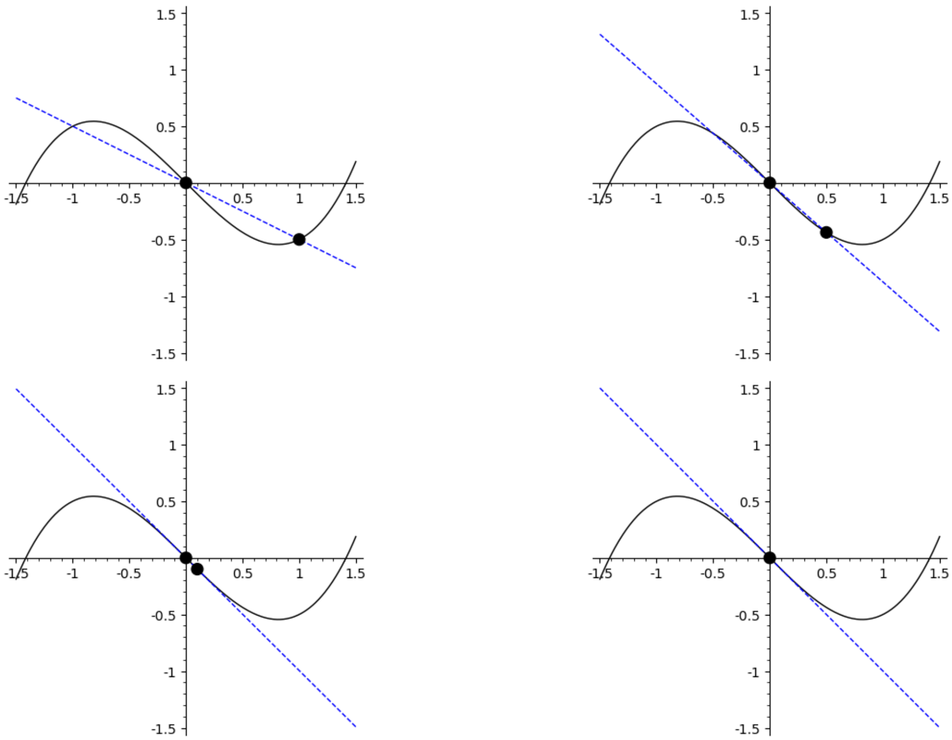
Just as the tangent of an angle is the length of a (specific) tangent line segment, the secant of an angle is the length of a (specific) secant line segment.

Given numbers a and b, a line through $(a, f(a))$ and $(b, f(b))$ is a secant line. As b gets closer to a, then the two points the secant line goes through get closer together. When we take the limit, our line "goes through the same point twice". Thus it only touches the curve at one point—so it is a tangent line. Thus we see that the linear approximation to a function at a point a is the line tangent at that point a.

And geometrically this should make sense. The tangent line touches the function graph at one point, and is going in the "same direction" as the graph at that point. Thus it's the line that looks most like the point. So it *should* be the line that best approximates that function.

Example 2.32. Let $f(x) = \frac{x^3}{2} - x$. We can draw secant lines through the points $(0, f(0))$ and $(b, f(b))$, and see what happens as b gets closer to a. Below, we see the lines for $b = 1, 1/2, 1/10$, and then finally the tangent line given by the linear approximation formula.

We can see that each of the first three lines passes through two points, but as the points get closer and closer together, the secant lines better approximate the tangent line we see in the fourth picture.



Sometimes rather than just approximating a specific number, we want to keep track of the formula we're using for our function.

Example 2.33. Find a formula to approximate $\sin(x)$ when x is "small". (This is the revenge of the Small Angle Approximation). Find a formula to approximate $\cos(x)$ when x is small. What's unusual about this second formula?

We take $a = 0$. Then since $\sin'(x) = \cos(x)$ and so $\sin'(0) = \cos(0) = 1$, we have

$$\sin(x) \approx 1(x - 0) + 0 = x.$$

If we take a small number, we see that $\sin(.05) \approx .05$, where the true answer is about .04998. So this is pretty good.

Similarly, $\cos'(x) = -\sin(x)$ so $\cos'(0) = 0$. Then

$$\cos(x) \approx 0(x - 0) + 1 = 1.$$

This is actually a constant! The line that fits $\cos(x)$ best near 0 is just the horizontal line $y = 1$.

We can calculate, e.g., that $\cos(.05) \approx 1$, where the true answer is about .9986. Though this is also pretty good.

Example 2.34. Let's find a formula to approximate $f(x) = x^3 + 3x^2 + 5x + 1$ near $a = 0$. What do you notice? Why does that happen?

We have $f(0) = 1$ and $f'(x) = 3x^2 + 6x + 5$ so $f'(0) = 5$. Thus

$$f(x) \approx 1 + 5x.$$

This is exactly what you get if you take the original polynomial and cut off all the terms of degree higher than 1.

This makes sense, because we're looking for the closest we can get to f without using terms of degree higher than 1.

Example 2.35. Let's find a formula to linearly approximate $f(x) = \dfrac{1}{1-x}$ near $x = 0$.

We compute that $f'(x) = (1-x)^{-2} = \frac{1}{(1-x)^2}$. Then

$$f(x) \approx 1 + x.$$

This is a special case of what's known as the geometric series formula.

Can we linearly approximate $f(x) = 1/x$ near 0?

No. We see that f is undefined at 0. More importantly, $f'(x) = -1/x^2$ is also undefined at zero. So there's no linear approximation.

Can we linearly approximate $f(x) = 1/x$ near 1?

Yes! We have $f(1) = 1$ and $f'(1) = -(1)^{-2} = -1$ so

$$f(x) \approx 1 - (x-1) = 2 - x.$$

Finally, we have one more approximation that is often quite useful, for exponents.

Example 2.36. As a warmup, let's approximate $(1.01)^{10}$.

Our function is $f(x) = x^{10}$ and our $a = 1$. So $f(a) = 1$ and $f'(a) = 10a^9 = 10$. Then we have

$$f(1.01) \approx 10(1.01 - 1) + 1 = 1.1.$$

The true answer is about 1.10462.

Now let's approximate $(1.01)^\alpha$ where $\alpha \neq 0$ is some constant.

We have $f(x) = x^\alpha$, so $f'(x) = \alpha x^{\alpha-1}$. We again have $f(1) = 1$ and $f'(1) = \alpha(1)^{\alpha-1} = \alpha$, so

$$f(1.01) \approx \alpha(1.01 - 1) + 1 = 1 + \alpha/100.$$

Now let's get the fully general useful formula: approximate $(1+x)^\alpha$ where x is "small" and $\alpha \neq 0$ is a constant. (This rule is the called the "binomial approximation" and is often useful in physics).

We still take $f(x) = x^\alpha$ and $a = 1$. But we compute

$$f(1+x) \approx 1 + \alpha(1+x-1) = 1 + \alpha x.$$

This formula is used constantly in physics and other applications.

It is probably more helpful in the long run to think about $f(x) = (1+x)^\alpha$, though. Then we have $f'(x) = \alpha(1+x)^{\alpha-1}$, and we get

$$f(x) \approx 1 + \alpha x.$$

2.7 Speed and Rates of Change

We often start by considering the idea of *speed*. Speed is defined to be distance covered divided by time spent; that is, $v = \frac{\Delta x}{\Delta t}$. In particular, if your position at time t is given by the function $p(t)$, then your average speed between time t_0 and time t_1 is

$$v = \frac{p(t_1) - p(t_0)}{t_1 - t_0}.$$

This formula should look familiar. It is the slope of a line through the points $(t_0, p(t_0))$ and $(t_1, p(t_1))$. It is *not* the derivative of p, because we didn't take a limit. It is instead a "difference quotient", which is really a fancy way of saying the slope of a line.

Example 2.37. For example, on Earth dropped objects fall about $p(t) = 5t^2$ meters after t seconds. The average speed between time $t = 1$ and time $t = 2$ is

$$v = \frac{p(2) - p(1)}{2-1} = \frac{20-5}{1} = 15\text{m/s}$$

and the average speed between time $t = 3$ and time $t = 1$ is

$$v = \frac{p(3) - f(1)}{3-1} = \frac{45-5}{3-1} = 20\text{m/s}.$$

It's useful here to look at the units. We know that the result is a speed, so comes out in m/s. But how do we know we get those units? We have to think a bit about what the function p is actually doing.

The function p gives us position as a function of time. Thus the *inputs* to p are given in seconds, and the *outputs* are given in meters. So it's not really fully correct to say that $p(t) = 5t^2$; that would suggest that $p(1\text{s}) = 5(1\text{s})^2 = 5\text{s}^2$. But your position isn't described in square seconds!

Instead, we would write something like $p(t\text{seconds}) = 5t^2\text{m}$. The function takes in seconds as inputs, and gives meters as outputs. Thus our last calculation properly should have been

$$v = \frac{p(3\text{s}) - f(1\text{s})}{3\text{s} - 1\text{s}} = \frac{45\text{m} - 5\text{m}}{3\text{s} - 1\text{s}} = 20\text{m/s}.$$

We see that the numerator—which is made up of the outputs of p—has units of meters, while the denominator, which is made up of the inputs of p, has units of seconds. So the entire fraction has units of m/s, which is what it should be.

We can give a more general formula. What's the average speed between time $t_0 = 1$ and time $t_1 = t$? We have

$$v = \frac{p(t\text{s}) - p(1\text{s})}{t\text{s} - 1\text{s}} = \frac{5t^2\text{m} - 5\text{m}}{t\text{s} - 1\text{s}} = 5(t+1)\frac{t-1}{t-1}\text{m/s}.$$

As long as $t \neq 1$, this gives us a formula for average speed between time t and time 1: the average speed is $5(t+1)$m/s. But what if we want to know the speed "at" the time $t = 1$?

On some level, this question doesn't make any sense. Speed is defined as the change in distance divided by the change in time; if time doesn't change, and distance doesn't change, then this doesn't really mean anything. But we'd like it to mean something, so we take a limit instead.

Thus we can define your *instantaneous speed* or *speed at time* t_0 to be

$$\lim_{t_1 \to t_0} \frac{p(t_1) - p(t_0)}{t_1 - t_0} = \lim_{h \to t_0} \frac{p(t_0 + h) - p(t_0)}{h}.$$

Notice that since the function p has input in seconds and output in meters, the instantaneous speed will be in m/s, as it should be. But also notice that this formula is just the definition of the derivative of p.

Thus from the previous example, we can see that the instantaneous speed at time $t_0 = 1$ is

$$v(1\text{s}) = p'(1\text{s}) = \lim_{t \to 1} 5(t+1)\frac{t-1}{t-1}\text{m/s} = 10\text{m/s}.$$

Alternatively, we know that $p(t) = 5t^2$, so by our derivative rules we know that $p'(t) = 10t$ and thus $p'(1) = 10$. Once we add units, we have $p'(t\text{s}) = 10t\text{m/s}$ and thus $p'(1\text{s}) = 10\text{m/s}$.

Thus the derivative of a function has different units from the original function. Since the derivative is given by a formula with output in the numerator and input in the denominator, the derivative will have the units of the output per units of input.

We can take this one step further and look at the derivative of p'. The function p' takes in a time and outputs a speed; its derivative will be

$$p''(t_0\text{s}) = \lim_{t \to t_0} \frac{p'(t\text{s}) - p'(t_0\text{s})}{t\text{s} - t_0\text{s}}.$$

The units of the denominator are still seconds; but the units of the top are m/s, so the second derivative takes in seconds and outputs meters per second *per second*, or m/s^2. This makes sense: the second derivative is the change in the first derivative, so p'' tells us how quickly the speed is changing. So it tells us how many meters per second your speed changes each second. This is otherwise known as "acceleration".

Once we have the speed of a particle in terms of its derivative, we can apply it to do the sort of things we've already been doing. So for instance, we can ask how far a dropped object will have fallen after 2.2 seconds. We could calculate this exactly, but we can also approximate:

$$p(2.2\text{s}) \approx p(2\text{s}) + p'(2\text{s})(2.2\text{s} - 2\text{s}) = 20\text{m} + 10\text{m/s}(.2\text{s}) = 22\text{m}.$$

The derivative occurs in lots of other situations that aren't just physical speed. One common place they show up is in economics.

Example 2.38 (Debt and Deficit)**.** A lot of discussions of economics and public policy address the deficit and the debt. The "deficit" and the "debt" are easy to confuse but importantly different, in a way that maps cleanly to the idea of a derivative.

A "deficit" is the amount of money that is currently owed; it is measured in dollars (or euro or yen or some other currency). The current US national deficit is approximately \$22 trillion.

A "deficit" is the rate at which the debt is increasing. So the national deficit is currently about \$1 trillion. This means we expect the debt next year to be about \$1 trillion bigger than the debt this year.

Mathematically we can define a function $D(t)$ which takes in the year and outputs the number of dollars owed. Then the annual deficit is

$$\frac{D((t+1)\text{y}) - D(t\text{y})}{1\text{y}}.$$

This isn't a derivative, since there's no limit; this is a *difference quotient* that measures a discrete change in debt over a discrete time. It's analogous to average speed, not instantaneous speed.

But we could imagine asking how the deficit is changing from month to month, or from week to week, or from hour to hour. We can take a limit as the time between $t + h$ and t goes to zero, and then the deficit would be the derivative of debt. The function $D'(t)$ will take in years, and output dollars per year.

What about the second derivative? The function D'' will take in years, and output the yearly change in the deficit, measured in dollars per year per year. When people talk about whether the deficit is going up or down, they are looking at the second derivative of the debt.

Both of these examples have one very important trait in common. The position function $p(t)$ and the debt function $D(t)$ output different types of things with different units, but they both take *time* as an input. But it's easy for a function to take inputs other than time, and these functions are often physically important and meaningful.

Example 2.39 (Slope)**.** We've already seen one example of this, in lab 4 when we studied tangent lines. If I'm thinking about the graph of a function, then the input to the function is a horizontal position, measured in inches (or some other unit of distance). And the output is a vertical position, also measured in inches. So $f(x)$ takes in inches and outputs inches.

The derivative $f'(x)$ will still take in inches. But if we compute the derivative $f'(x) = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$, then the denominator is in inches and the numerator is also in inches. This makes the derivative technically unitelss—but in reality, it is measured in inches per inch.

And this has a clear physical interpretation! We already know the derivative gives us the slope of the tangent line, and the slope measures how many units the graph goes up for each unit it goes over. Thus, it measures inches of horizontal position per inch of vertical position.

The second derivative $f''(x)$ will take in inches and output 1/inch, which is really inches per inch per inch. It tells us how much the slope, measured in inches per inch, changes if we move one inch horizontally.

Example 2.40 (Price Elasticity of Demand)**.** Another common economics question is to see how the demand for a product relates to its price. We can define a function $Q(p)$ that takes in a price in dollars, and outputs the quantity of items that will be bought. So if $Q(p) = 10000 - 10p$, this means that if the price is \$100 then people will buy $Q(100) = 10000 - 1000 = 9000$ widgets.

What's the derivative here? The function $Q'(p)$ takes in a price in dollars and outputs a number of widgets per dollar. It tells you how the quantity demanded changes in response to changes in the price. Thus we see that since $Q'(p) = -10$, we expect to sell ten fewer widgets for each dollar we raise the price.

(Economists call this the Price Elasticity of Demand: "elasticity" is how quickly one thing responds to changes in another thing. So any time the term "elasticity" shows up in economics, there's a derivative inolved somewhere).

What if instead we had the function $Q(p) = 10000 - 5p^2$? Now we see that changing the price doesn't have a huge effect if the price is already small, but it has a dramatic effect if the price is big. We compute that $Q'(p) = -10p$. This means that increasing the price by one dollar will decrease the quantity demanded by ten widgets for every dollar of the price.

Thus if the current price is \$10, we expect raising the price to \$11 to reduce sales by about a hundred widgets. If the current price is \$30 then raising the price will lose us nine hundred widgets in sales.

Example 2.41 (Ohm's Law)**.** In physics and electrical engineering, Ohm's Law tells us that current is equal to voltage over resistance, or $I = V/R$. (Here current is generally measured in amperes, voltage in volts, and resistance in, essentially, volts per amp).

The default assumption in most physics problems is that resistance is constant, a property of whatever material you're putting current through. So we have the function $I(V) = \frac{1}{R}V$, which is a linear function and simple to work with.

But this is just an approximation! Most materials will actually have their resistance change as the voltage applied to them changes, so the equation above is just a linear approximation to the actual relationship between current and voltage. This means that the slope $\frac{1}{R}$ is really a derivative.

An incandescent lightbulb works by running a current through a metal wire until it heats up. But as the heat of the wire increases, the resistance goes up. Thus the graph of current as a function of voltage is curving down; the higher the voltage, the less extra current you get from adding another volt. This means that the derivative $\frac{dI}{dV}$ is large when V is small, but small when V is large.

A diode is a material that does the opposite. Resistance is high when the voltage is low, but past some transition point the resistance drops and becomes very low. This means that the derivative is large when V is small, and then small when V is large. The graph of I as a function of V will curve up.

In practice. engineers mostly don't want to worry about the whole curve. If they know about what voltage their devices will experience, they don't need to worry what happens in other places. So they take the local linear approximation, call that "the resistance", and use the equation $I = I_0 + \frac{1}{R}(V - V_0)$. And this is just the linear approximation equation we've been using all class.

2.8 Implicit Differentiation

We defined a function as a rule, that takes some input and gives some output. Usually we give you the rule explicitly, as when we say $y = x^2 - 1$. But sometimes you only know facts about the rule, such as $y^2 + x^2 = 1$ (which describes the unit circle). Sometimes these facts will describe one function uniquely, and sometimes they won't. (This comes up a lot in solving actual problems in physics and economics and other fields).

Regardless of where we get an equation like this, we know that both sides are equal, so the derivatives of both sides are equal. So using the chain rule and thinking of y as a function of x, we can simply take derivatives of both sides, and then do some algebra to find y'.

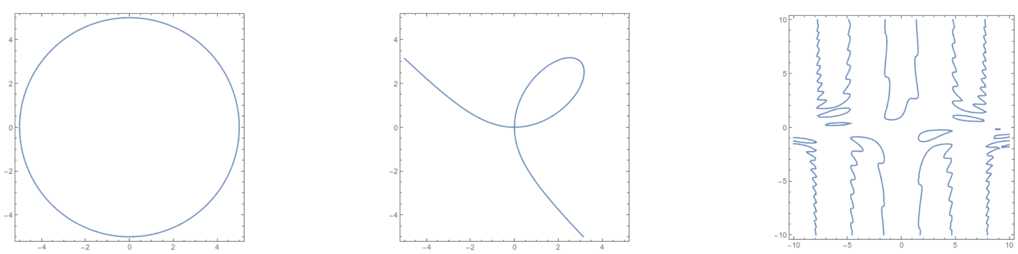


Figure 2.3: Left: The circle $x^2 + y^2 = 25$. Center: the folium of Descartes $x^3 + y^3 = 6xy$. Right: $y\cos(x) = 1 + \sin(xy)$

If we want to find tangent lines for these curves, we can use implicit differentiation. Essentially, we take the derivative of both sides of the equation, treating y as a function of x and applying the chain rule.

Example 2.42.

- If $x^2 + y^2 = 25$, then

$$\begin{aligned} \frac{d}{dx}\left(x^2 + y^2\right) &= \frac{d}{dx}(25) \\ 2x + 2y\frac{dy}{dx} &= 0 \\ \frac{dy}{dx} &= \frac{-x}{y}. \end{aligned}$$

  Thus at the point, say, $(3,4)$ (check that this is on the circle!), we have that $\frac{dy}{dx}(3,4) = \frac{-3}{4} = -3/4$. Thus the equation of the line tangent to the circle at $(3,4)$ is $y - 4 = -\frac{3}{4}(x-3)$.

- If $x^3 + y^3 = 6xy$, then

$$\begin{aligned}
\frac{d}{dx}\left(x^3 + y^3\right) &= \frac{d}{dx}(6xy) \\
3x^2 + 3y^2\frac{dy}{dx} &= 6\left(y + x\frac{dy}{dx}\right) \\
(3y^2 - 6x)\frac{dy}{dx} &= 6y - 3x^2 \\
\frac{dy}{dx} &= \frac{6y - 3x^2}{3y^2 - 6x}.
\end{aligned}$$

  At the point $(0, 0)$ this doesn't actually give us a useful answer; if you look at the picture in Figure 2.3, you see that there's not a clear tangent line there since the curve crosses itself.

  In contrast, at the point $(3, 3)$ we have that

$$\frac{dy}{dx} = \frac{18 - 27}{27 - 18} = -1$$

  and the equation of the tangent line is $y - 3 = -(x - 3)$.

- If $y\cos(x) = 1 + \sin(xy)$, then

$$\begin{aligned}
\frac{d}{dx}(y\cos(x)) &= \frac{d}{dx}(1 + \sin(xy)) \\
\frac{dy}{dx}\cos(x) - y\sin(x) &= \cos(xy)\left(y + x\frac{dy}{dx}\right) \\
\frac{dy}{dx}(\cos(x) - x\cos(xy)) &= y\cos(xy) + y\sin(x) \\
\frac{dy}{dx} &= \frac{y\cos(xy) + y\sin(x)}{\cos(x) - x\cos(xy)}.
\end{aligned}$$

- If $\sqrt{xy} = 1 + x^2y$, then

$$\begin{aligned}
\frac{d}{dx}\sqrt{xy} &= \frac{d}{dx}\left(1 + x^2y\right) \\
\frac{1}{2}(xy)^{-1/2}\left(y + x\frac{dy}{dx}\right) &= 2xy + x^2\frac{dy}{dx} \\
\frac{dy}{dx}\left(x^2 - \frac{1}{2}x(xy)^{-1/2}\right) &= \frac{1}{2}(xy)^{-1/2}y - 2xy \\
\frac{dy}{dx} &= \frac{\frac{1}{2}(xy)^{-1/2}y - 2xy}{x^2 - \frac{1}{2}x(xy)^{-1/2}}.
\end{aligned}$$

Example 2.43. We can also compute second derivatives implicitly. If $9x^2 + y^2 = 9$ then we have

$$\begin{aligned}
18x + 2y\frac{dy}{dx} &= 0 \\
\frac{dy}{dx} &= -\frac{9x}{y} \\
\frac{d^2y}{dx^2} &= \frac{d}{dx}\left(-\frac{9x}{y}\right) \\
&= -\frac{9y - 9x\frac{dy}{dx}}{y^2} \\
&= -\frac{9y - 9x(-\frac{9x}{y})}{y^2} \\
&= -\frac{9y + \frac{81x^2}{y}}{y^2}
\end{aligned}$$

We see that at the point $(0, 3)$ we have $y' = 0$ and $y'' = -3$. At the point $(\sqrt{5}/3, 2)$, then $y' = -\frac{3\sqrt{5}}{2}$ and $y'' = -\frac{18+\frac{45}{2}}{4}$.

Example 2.44. Find y'' if $x^6 + \sqrt[3]{y} = 1$. Then find the first and second derivatives at the point $(0, 1)$.

$$\begin{aligned}
6x^5 + \frac{1}{3}y^{-2/3}y' &= 0 \\
-18x^5y^{2/3} &= y' \\
-18(5x^4y^{2/3} + \frac{2}{3}x^5y^{-1/3}y') &= y'' \\
-18(5x^4y^{2/3} + \frac{2}{3}x^5y^{-1/3}(-18x^5y^{2/3})) &= y''
\end{aligned}$$

Thus at $(0, 1)$, we have $y' = 0$ and $y'' = 0$. So the tangent line to the curve is horizontal at the point $(0, 1)$.

We can also use implicit differentiation on relationships that apply to functions.

Example 2.45. Suppose we have some function f such that $8f(x) + x^2(f(x))^3 = 24$, and we want to find a linear approximation of f near $f(4) = 1$. (Say we've measured this experimentally and now want to understand or compute with the function). Then we have

$$\begin{aligned}\frac{d}{dx}\left(8f(x) + x^2(f(x))^3\right) &= \frac{d}{dx}24\\ 8f'(x) + 2x(f(x))^3 + 3x^2(f(x))^2 f'(x) &= 0\\ 8f'(4) + 2\cdot 4\cdot 1^3 + 3\cdot 4^2\cdot 1^2 f'(4) &= 0\\ 8f'(4) + 8 + 48f'(4) &= 0\end{aligned}$$

and thus $f'(4) = -1/7$.

This leaves us with a question, though. We know $f(1)$; can we figure out the value of f at other points?

We have a derivative, so we can again compute a linear approximation. We get

$$f(x) \approx f'(4)(x-4) + f(4) = \frac{-1}{7}(x-4) + 1.$$

Thus we compute

$$f(5) \approx \frac{-1}{7}(5-4) + 1 = 1 + \frac{-1}{7} = \frac{6}{7} \approx .857.$$

Checking Mathematica, we see that the actual solution is .879. So we were pretty close.

2.9 Related Rates

Sometimes we have word problems that require us to translate verbal information into equations, and then solve the problem.

Checklist of steps for solving word problems:

1. Draw a picture.

2. Think about what you expect the answer to look like. What is physically plausible?

3. Create notation, choose variable names, and label your picture.

   (a) Write down all the information you were given in the problem.

   (b) Write down the question in your notation.

4. Write down equations that relate the variables you have.

5. Abstractly: "solve the problem." Concretely differentiate your equation.
6. Plug in values and read off the answer.
7. Do a sanity check. Does you answer make sense? Are you running at hundreds of miles an hour, or driving a car twenty gallons per mile to the east?

Example 2.46. Suppose one car drives north at 40 mph, and an hour later another starts driving west from the same place at 60 mph. After a second hour, how quicly is the distance between them increasing?

Write a for the distance the first car has traveled, and b for the distance the second car has traveled. We have that $a = 80, b = 60, a' = 40, b' = 60$. If the distance between the cars is d then after two hours, $d = 100$, and we have

$$\begin{aligned} d^2 &= a^2 + b^2 \\ 2dd' &= 2aa' + 2bb' \\ 2 \cdot 100 \cdot d' &= 2 \cdot 80 \cdot 40 + 2 \cdot 60 \cdot 60 \\ d' &= \frac{3200 + 3600}{100} = 68, \end{aligned}$$

so the distance between the cars is increasing at 68 mph. This seems reasonable because the cars are traveling at 40 mph and 60 mph.

Example 2.47. A twenty foot ladder rests against a wall. The bit on the wall is sliding down at 1 foot per second. How quickly is the bottom end sliding out when the top is 12 feet from the ground?

Let h be the height of the ladder on the wall, and b be the distance of the foot of the ladder from the wall. Then $h = 12, h' = -1$, and $b = \sqrt{400 - 144} = 16$. We have

$$\begin{aligned} h^2 + b^2 &= 400 \\ 2hh' + 2bb' &= 0 \\ 2 \cdot 12 \cdot (-1) + 2 \cdot 16 \cdot b' &= 0 \\ b' &= \frac{24}{32} = 3/4 \end{aligned}$$

so the foot of the ladder is sliding away from the wall at 3/4 ft/s. Again, the direction of the sliding is correct (away from the wall), and the number seems plausible.

Example 2.48. A spherical balloon is inflating at 12 cm^3 per second. How quickly is the radius increasing when the radius is 3 cm?

A sphere has volume $V = \frac{4}{3}\pi r^3$. We have $V' = 12$ and $r = 3$. We compute

$$\begin{aligned} V' &= 4\pi r^2 r' \\ 12 &= 4\pi(3)^2 r' \\ r' &= \frac{1}{3\pi} \end{aligned}$$

So the radius is increasing by $1/3\pi$ cm per second.

Example 2.49. A rectangle is getting longer by one inch per second and wider by two inches per second. When the rectangle is 5 inches long and 7 inches wide, how quickly is the area increasing?

We have $l = 5, w = 7, l' = 1, w' = 2$, and $A = lw$. Taking a derivative gives us $A' = lw' + wl' = 5 \cdot 2 + 7 \cdot 1 = 17$ square inches per second.

Example 2.50. An inverted conical water tank with radius 2m and height 4m is being filled with water at a rate of 2m^3/min. How fast is the water rising when the water is 3 m tall?

Let h be the current height of the water, r the current radius, and V the current volume of water. We know that $h = 3$, and by similar triangles we see that $\frac{h}{r} = \frac{4}{2}$ and thus $r = h/2$. We know that $V' = 2$, and the volume formula for a cone gives us $V = \frac{1}{3}\pi r^2 h$. We compute

$$\begin{aligned} V &= \frac{1}{3}\pi \left(\frac{h}{2}\right)^2 h = \frac{1}{3}\pi \frac{h^3}{4} \\ V' &= \frac{\pi}{4} h^2 h' \\ 2 &= \frac{\pi}{4} 3^2 h' \\ \frac{8}{9\pi} &= h', \end{aligned}$$

so the water level is rising at $\frac{8}{9\pi}$ meters per minute.

Example 2.51. A street light is mounted at the top of a 15-foot-tall pole. A six-foot-tall man walks straight away from the pole at 5 feet per second. How fast is the tip of his shadow moving when he is forty feet from the pole?

Let d be the distance of the man from the pole, and L be the distance from the pole to the tip of his shadow. We have $d' = 5$ and we set up a similar triangles equation.

$$\begin{aligned} \frac{15}{L} &= \frac{6}{L-d} & 6L &= 15L - 15d \\ 9L &= 15d & d &= \frac{3}{5}L \\ d' &= \frac{3}{5}L' & 5 &= \frac{3}{5}L' \end{aligned}$$

and thus the tip of his shadow is moving at $\frac{25}{3}$ feet per second.

Example 2.52. A lighthouse is located three kilometers away from the nearest point P on shore, and its light makes four revolutions per minute. How fast is the beam of light moving along the shoreline 1 kilometer from P?

Let's say the angle of the light away from P is θ, and the distance from P is d. Then we have $d = 1$ and $\theta' = 8\pi$ (in radians per minute). We also have the relationship that $\tan\theta = \frac{d}{3}$.

Taking the derivative gives us $\sec^2(\theta) \cdot \theta' = d'/3$. We need to work out $\sec^2(\theta)$, but looking at our triangle we see that the adjacent side is length 3 and the hypotenuse is length $\sqrt{10}$ (by the Pythagorean theorem), so we have $\sec^2(\theta) = (\sqrt{10}/3)^2 = 10/9$.

Thus we have $d' = 3\sec^2(\theta) \cdot 8\pi = \frac{80\pi}{3}$ kilometers per second.

Example 2.53. A kite is flying 100 feet over the ground, moving horizontally at 8 ft/s. At what rate is the angle between the string and the ground decreasing when 200ft of string is let out?

Call the distance between the kite-holder and the kite d and the angle between the string and the ground θ. When the length of string is 200 then $d = \sqrt{200^2 - 100^2} = 100\sqrt{3}$. We have that $d' = 8$ (since the angle is decreasing, the kite must be getting farther away). And finally we have the relationship $\tan\theta = \frac{100}{d}$ by the definition of tan in terms of triangles. Then we have

$$\begin{aligned}
\tan\theta &= 100d^{-1} \\
\sec^2(\theta)\theta' &= -100d^{-2}d' \\
\theta' &= \frac{-100 \cdot 8\cos^2(\theta)}{d^2}.
\end{aligned}$$

We see that $\cos(\theta) = \frac{100\sqrt{3}}{200} = \sqrt{3}2$, so we have

$$\theta' = \frac{-100 \cdot 8 \cdot 3/4}{(100\sqrt{3})^2} = -\frac{8}{100 \cdot 4} = \frac{-1}{50}.$$

So the angle between the string and the ground is decreasing at a rate of 1/50 per second. (Note: radians are unitless!)

3 Optimization

We'd like to start using calculus to answer questions about functions, other than the question "what can calculus tell us about functions?" One thing we could plausibly ask about the behavior of a function is its extreme values: where is it biggest? Where is it smallest? Where is it big or small relative to nearby points?

3.1 Extreme Values and Critical Points

Definition 3.1. If $f(c) \geq f(x)$ for every x in the domain of f, then $f(c)$ is an *absolute maximum* or *global maximum* for f. We say that f has an absolute maximum at c.

Similarly, if $f(c) \leq f(x)$ for every x in the domain of f, then $f(c)$ is an *absolute minimum* or *global minimum* for f, and f has a global minimum at c.

Absolute maxima and absolute minima are somtimes collectively called *absolute extrema.* ("Extremum" comes from "extreme value," meaning a value that is very big or small or otherwise unusual).

Note that absolute maxima and minima do not necessarily exist: the function $f(x) = x$ has no absolute maxima or minima on the real line, and $\tan x$ defined between $-\pi/2$ and $\pi/2$ has no absolute extrema. Nor are they necessarily unique; if we define $f(x) = c$ for some constant c, then there is an absolute maximum and an absolute minimum at every point–every point outputs both the largest possible value and the smallest possible value.

Theorem 3.2 (Extreme Value Theorem)**.** *If f is continuous on a closed interval $[a, b]$, then f has an absolute maximum $f(c)$ at some point c in the interval $[a, b]$, and an absolute minimum $f(d)$ at some point d in the interval $[a, b]$.*

Note that both the continuity and the closed-ness are important here. Also, this is another "existence theorem": it tells us that a global maximum and a global minimum exist, but not anything about where. We can answer this question and find them, but it will require a bit more setup.

We can also look for places where the graph of our function has a peak or a valley, even if it's not the biggest or smallest possible point:

Definition 3.3. If $f(c) \geq f(x)$ for all x near c, we say that $f(c)$ is a *relative maximum* or a *local maximum* for, and that f has a relative maximum at c.

If $f(c) \leq f(x)$ for all x near c, we say that $f(c)$ is a *relative minimum* or a *local minimum* for f, and that f has a relative minimum at c.

Theorem 3.4 (Fermat's Theorem/Critical Point Theorem)**.** *If f has a local extremum at c, and c is not an endpoint of the domain of f, and $f'(c)$ exists, then $f'(c) = 0$.*

Proof. Intuitive idea: If $f'(c) > 0$ then f is increasing, so $f(c+h) > f(c)$ for some small positive h. If $f'(c) < 0$ then f is decreasing, so $f(c+h) > f(c)$ for some small negative h.

To keep things simple, let's suppose f has a local maximum at c, and $f'(c)$ exists. Since $f(c)$ is a local maximum, we know that $f(c) \geq f(c+h)$ for small h, and thus that $f(c+h) - f(c) \leq 0$.

If we take h to be positive, then we can divide both sides by h and we get

$$\frac{f(c+h) - f(c)}{h} \leq 0$$
$$\lim_{h \to 0^+} \frac{f(c+h) - f(c)}{h} \leq 0.$$

But since $f'(c)$ exists, this limit must be $f'(c)$, so $f'(c) \leq 0$.

If we take h to be negative, then dividing both sides of our inequality by h flips the inequality, and we get

$$\frac{f(c+h) - f(c)}{h} \geq 0$$
$$\lim_{h \to 0^-} \frac{f(c+h) - f(c)}{h} \geq 0.$$

But since $f'(c)$ exists, this limit must be $f'(c)$, so $f'(c) \geq 0$.

But then $f'(c) \geq 0$ and $f'(c) \leq 0$, so $f'(c) = 0$. □

Remark 3.5.

- The converse of this theorem isn't true: you can have points where $f'(c) = 0$ or $f'(c)$ does not exist that are not local extrema.
- Your textbook uses its words slightly differently, and believes that you cannot have a relative extremum at the endpoint of an interval. I think this is poor word choice, but you should be aware of it when reading the textbook.

Definition 3.6. We say that c is a *critical point* of a function f if either $f'(c) = 0$ or $f'(c)$ does not exist.

Then Fermat's theorem says specifically that if f has a local extremum at c, then c is a critical point. (Again, remember that c can be a critical point without being a local extremum).

Example 3.7.
- Let $f(x) = x^3 - x$. Then $f'(x) = 3x^2 - 1$; this is defined everywhere, and $f'(x) = 0$ when $x = \pm\frac{\sqrt{3}}{3}$. So the critical points are $\pm\frac{\sqrt{3}}{3}$.
- If $f(x) = x^2$, then $g'(x) = 2x$ and is 0 when $x = 0$. So the only critical point is 0.
- If $h(x) = \sin(x)$ then $h'(x) = \cos(x)$, which is 0 when $x = (n + 1/2)\pi$ for any integer n. Thus the critical points are $\pi/2, 3\pi/2, 5\pi/2, \dots$.
- If $f(x) = x^3$ then $f'(x) = 3x^2$ which is 0 when x is 0. Thus the only critical point is at 0.
- If $g(x) = |x|$ then
$$g'(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ DNE & x = 0 \end{cases}$$
and thus has a critical point at $x = 0$ since the derivative does not exist there.
- If $f(x) = |x^2 - 4|$ then we know that $|x|$ isn't differentiable at 0, so $f(x)$ won't be differentiable at $x^2 - 4 = 0$ and thus at $x = \pm 2$. We see the derivative of the inside is $2x$, so $f'(x) = \pm 2x = 0$ when $x = 0$, and thus the critical points are $0, \pm 2$.

The obvious next question is "how can we determine whether these critical points are a maximum or a minimum or neither?" This is a bit tricky, so we'll hold off for a bit. First we will identify the absolute extrema of a continuous function on a closed interval.

Remember that if f is continuous on $[a, b]$, it must have an absolute maximum and an absolute minimum. By Fermat's theorem, if the absolute extrema are in the interior they must be at critical points. So we can find the absolute extrema by the following method:

1. List all the critical points.
2. Evaluate f at each critical point, and at a and b.
3. The largest value is the maximum and the smallest is the minimum.

Example 3.8.
- If $f(x) = x^3 - x$, we saw the critical points are $\pm\sqrt{3}/3$. If we want the absolute maximum on $[0, 2]$, we compute that $f(0) = 0, f(2) = 6$, and $f(\sqrt{3}/3) = -2\sqrt{3}9$. Thus the absolute maximum is 6 at 2 and the absolute minimum is $-2\sqrt{3}/9$ at $\sqrt{3}/3$.

- Consider $g(x) = x^3 - 3x^2 + 1$ on $[-1, 4]$. We have $g'(x) = 3x^2 - 6x = 0$ when $x = 0$ or $x = 2$, so the critical points are 0 and 2. We compute $g(-1) = -3, g(0) = 1, g(2) = -3, g(4) = 17$. Thus the absolute maximum is 17 at 4, and the absolute minimum is -3 at -1 and 2.

- Let $h(x) = 2\cos t + \sin(2t)$ on $[0, \pi/2]$. Then $h'(x) = -2\sin(t) + 2\cos(2t) = 0$ when $\sin(t) = \cos(2t)$. On $[0, \pi/2]$ this happens precisely when $x = \pi/6$, so this is the only critical point. We compute $h(0) = 2, h(\pi/2) = 0, h(\pi/6) = 3\sqrt{3}/2$, so the absolute maximum is $3\sqrt{3}/2$ at $\pi/6$ and the absolute minimum is 0 at $\pi/2$.

- Let $f(x) = \frac{x^2+3}{x-1}$ on $[-2, 0]$. Then we see that

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

  does not exist at 1. To test when $f'(x) = 0$ we need only consider the numerator, so we have $0 = x^2 - 2x - 3 = (x-3)(x+1)$ and thus $x = 3$ or $x = -1$. So the critical points are $-1, 1, 3$.

  f is continuous on $[-2, 0]$ and so must have global extrema. To find them we only need to look at the critical points in $[-2, 0]$, and thus only at -1. So we compute $f(0) = -3, f(-1) = -2, f(-2) = -7/3$. Thus the maximum is -2 (at -1) and the minimum is -3 (at 0).

- What about the global extrema of that same function on $[0, 2]$? We already know the critical points, so we need to check $0, 1, 2$. We have $f(0) = -3$ and $f(2) = 7$, but $f(1)$ is not defined. In fact the function is not defined everywhere on $[0, 2]$ and so not continuous; it has an asymptote at $x = 1$ and thus no minimum or maximum.

- Let's find the global extrema of $g(x) = \sqrt[3]{x^3 + 3x^2}$ on the closed interval $[-2, 2]$. This is a continuous function on a closed interval, so by the Extreme Value Theorem it has absolute extrema. We take the derivative, and get

  $$g'(x) = \frac{1}{3}(x^3 + 3x^2)^{-2/3}(3x^2 + 6x) = \frac{3x(x+2)}{3\sqrt[3]{(x^3 + 3x^2)^2}}.$$

  This derivative is zero when $x = 0$ or $x = -2$, and it does't exist when $x = 0$ or $x = -3$.

We'd still like to determine what each critical point is like, but for that we will need more tools.

3.2 The Mean Value Theorem

We begin with a theorem called Rolle's Theorem:

Theorem 3.9 (Rolle)**.** *If f is continuous on $[a, b]$ and differentiable on (a, b), and $f(a) = f(b)$, then there is a point c in (a, b) where $f'(c) = 0$.*

Proof. If f is constant everywhere, then the derivative is 0 everywhere.

By the Extreme Value theorem, f has a global maximum on $[a, b]$. If there is some x in (a, b) with $f(x) > f(a)$, then the maximum is in the interior at some point c, and by Fermat's theorem, since $f'(c)$ must exist, we have $f'(c) = 0$.

If f is not constant, and there is no x with $f(x) > f(a)$, then there is some f with $f(x) < f(a)$. Then f has an absolute minimum in the interior at some point c. By Fermat's theorem $f'(c) = 0$. □

Remark 3.10. We need f to be continuous at the endpoints, but it doesn't have to be differentiable there. Rolle's theorem does guarantee a derivative of zero somewhere in the interior–not just at the endpoints.

Example 3.11. If $f(x)$ represents the height of an object, $f'(x)$ represents its speed. If I throw an object up and wait for it to fall back down to the ground, at some point during the process (at the top of its arc) it's instantaneous velocity will be 0.

Example 3.12. We can prove that $f(x) = x^3 + x - 1$ has exactly one real root.

First we use the Intermediate Value Theorem to show that a root exists at all. f is continuous because it's a polynomial. We see that $f(0) = -1 < 0$ and $f(1) = 1 > 0$, so by the Intermediate Value Theorem there's some a in $(0, 1)$ with $f(a) = 0$. Thus f has at least one real root.

Now suppose $f(b) = 0$ and $b \neq a$. Then f is continuous and differentiable everywhere, and $f(a) = f(b)$, so by Rolle's theorem there's some c in between a and b with $f'(c) = 0$.

But $f'(c) = 3c^2 + 1$, and since $c^2 \geq 0$, we know that $f'(c) \geq 1$ for every c. Thus there's no c with $f'(c) = 0$, so there's no $b \neq a$ with $f(b) = 0$. Thus f has exactly one real root.

Rolle's theorem can be useful, but it's very limited by the need for $f(a) = f(b)$. The Mean Value Theorem lets us lift that restriction.

Theorem 3.13 (Mean Value Theorem)**.** *If f is continuous on $[a, b]$ and differentiable on (a, b), then there's a c in (a, b) with*

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

Proof. We prove this using Rolle's theorem, by writing an altered version of f that satisfies the hypotheses of Rolle's theorem. Define

$$h(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

This is basically just taking $f(x)$ and then subtracting off the line from $(a, f(a))$ to $(b, f(b))$. It's clear that

$$\begin{aligned} h(a) &= f(a) - f(a) - \frac{f(b) - f(a)}{b - a}(a - a) = 0 - \frac{f(b) - f(a)}{b - a}0 = 0 \\ h(b) &= f(b) - f(a) - \frac{f(b) - f(a)}{b - a}(b - a) = (f(b) - f(a)) - (f(b) - f(a)) = 0 \end{aligned}$$

so $h(a) = h(b)$. h is continuous on $[a, b]$ because f is continuous on $[a, b]$, polynomials are continuous, and the sum of two continuous functions is continuous. h is differentiable on (a, b) because f is differentiable on (a, b), polynomials are differentiable, and the sum of two differentiable functions is differentiable.

Thus h satisfies the hypotheses of Rolle's theorem. Then there's some c in (a, b) with $h'(c) = 0$. But

$$\begin{aligned} h'(x) &= f'(x) - \frac{f(b) - f(a)}{b - a}(1 - 0) \\ 0 &= f'(c) - \frac{f(b) - f(a)}{b - a} \\ f'(c) &= \frac{f(b) - f(a)}{b - a} \end{aligned}$$

as we desired. □

Example 3.14. Earlier in the class, we talked about driving to San Diego. That's about 120 miles, so if it takes me two hours to get there, my average speed is 60 mph. That doesn't mean my speed at each point is 60 mph, though; I might go 90 part of the way and then 20 part of the way while I'm stuck in traffic. But the Mean Value Theorem tells me that at some point during that drive the needle on my speedometer pointed at the 60–which makes sense, since it will do that while I'm accelerating up to 90.

Example 3.15. We can also use the mean value theorem to constrain the possible values for a function. For instance, suppose I have a function f, and all I know is that $f(1) = 10$ and $f'(x) \geq 2$ for every x. Then if I want to know about $f(4)$, I can conclude that there is some c in $(1, 4)$, such that:

$$\begin{aligned} f'(c) &= \frac{f(4) - f(1)}{4 - 1} \\ 3f'(c) &= f(4) - 10 \\ f(4) &= 10 + 3f'(c) \geq 10 + 3 \cdot 2 = 16. \end{aligned}$$

Thus $f(4) \geq 16$.

Example 3.16. Suppose $|f'(x)| \leq 2$ for all x, and $f(0) = 7$. What do we know about $f(5)$?

We know that for any x, $-2 \leq f(x) \leq 2$. By the mean value theorem, we have

$$\begin{aligned} f'(c) &= \frac{f(5) - f(0)}{5 - 0} \\ -2 &\leq \frac{f(5) - f(0)}{5 - 0} \leq 2 \\ -10 &\leq f(5) - 7 \leq 10 \\ -3 &\leq f(5) \leq 17. \end{aligned}$$

This corresponds to the intuition that if you're travelling less than 2 miles per hour, you won't get more than ten miles in five hours; and if you start at 7, you'll wind up between -3 and 17.

Example 3.17. Show $f(x) = x^5 + x^3 + x$ has exactly one root.

It's pretty clear that f has a root; we could use the intermediate value theorem, but we can also observe that $f(0) = 0$.

Suppose $f(a) = f(b) = 0$. Then by Rolle's Theorem there is some c with $f'(c) = 0$. But $f'(x) = 5x^4 + 3x^2 + 1 \geq 1$ and thus $f'(c)$ is never zero; so f has at most one root, and thus exactly one root.

More intuitively, $f(x)$ has at most one root because it's always increasing, and so one it gets above zero it can't come back down and hit zero again. Which leads us to discuss the idea of increasing or decreasing functions.

3.3 Increasing or Decreasing Functions and Finding Relative Extrema

We now want to use the Mean Value Theorem to answer our original question, about which critical points are maxima or minima. We start with a defnition:

Definition 3.18. We say that f is *(strictly) increasing* on an interval (a, b) if, whenever x_1 and x_2 are points in (a, b) and $x_2 > x_1$, then $f(x_2) > f(x_1)$.

We say that f is *(strictly) decreasing* on an interval (a, b) if, whenever x_1 and x_2 are points in (a, b) and $x_2 > x_1$, then $f(x_2) < f(x_1)$.

Notice that these definitions make sense if you assume we're moving to the right; an increasing function is one where $f(x)$ increases as x increases.

Proposition 3.19.

- *If $f'(x) = 0$ for all x in (a, b), then f is constant on (a, b).*
- *If $f'(x) > 0$ for all x in (a, b), then f is increasing on (a, b).*
- *If $f'(x) < 0$ for all x in (a, b), then f is decreasing on (a, b).*

Proof. Let x_1, x_2 be two points in (a, b) with $x_2 > x_1$. Then since f is differentiable (and thus continuous) everywhere in (a, b), it is continuous and differentiable everywhere on $[x_1, x_2]$, and by the mean value theorem there is some c with

$$f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$
$$(x_2 - x_1)f'(c) = f(x_2) - f(x_1).$$

- Now, if $f'(x) = 0$ for all x, then $f'(c) = 0$ and thus $f(x_2) - f(x_1) = 0$. This is true for any points x_1 and x_2, and thus f is constant.
- If $f'(x) > 0$ for all x, then $f'(c) > 0$. Since $x_2 - x_1 > 0$, this implies that $f(x_2) - f(x_1) > 0$. This is true for any points $x_1 < x_2$ and thus f is increasing.
- If $f'(x) < 0$ for all x, then $f'(c) < 0$. Since $x_2 - x_1 < 0$, this implies that $f(x_2) - f(x_1) < 0$. This is true for any points $x_1 < x_2$ and thus f is decreasing.

□

Remark 3.20. This theorem doesn't say anything about intervals where f isn't always differentiable. It also doesn't say anything about intervals where f' switches sign in the middle. In practice, we split the domain of our function up into intervals on which exactly one of these things is happening and study each interval separately.

Example 3.21. Let $f(x) = 3x^4 - 4x^3 - 12x^2 + 5$. Where is f increasing or decreasing?

$f'(x) = 12x^3 - 12x^2 - 24x = 12x(x-2)(x+1)$ is 0 when $x = 0, -1, 2$. These three points are the critical points. $f'(x)$ has three factors, and it will be positive when one or all three factors are positive. We make a chart:

	$12x$	$x-2$	$x+1$	$f'(x)$
$x < -1$	$-$	$-$	$-$	$-$
$-1 < x < 0$	$-$	$-$	$+$	$+$
$0 < x < 2$	$+$	$-$	$+$	$-$
$2 < x$	$+$	$+$	$+$	$+$

Thus $f'(x)$ is positive when $-1 < x < 0$ or $2 < x$, so f is increasing on $(-1, 0)$ and on $(2, +\infty)$. $f'(x)$ is negative when $x < -1$ or $0 < x < 2$, so f is increasing on $(-\infty, -1)$ and $(0, 2)$.

Can we use this information about increasing and decreasing functions to say something about relative maxima and minima? In fact, assuming f is continuous at c, if f is increasing to the left of a point c and decreasing to the right of c, then it must have a maximum at c. Similarly, if f is decreasing to the left and increasing to the right, it must have a minimum. If it increases on both sides or decreases on both sides, then c is neither a maximum nor a minimum. Therefore:

Proposition 3.22 (First derivative test for extrema)**.** *If c is a critical point of f and f is continuous at c, then*

- *If f' changes from positive to negative at c then f has a relative maximum at c.*
- *If f' changes from negative to positive at c then f has a relative minimum at c.*
- *If f' "changes" from positive to positive or negative to negative at c then f has neither a relative maximum nor a relative minimum at c.*

Remark 3.23. If f' is continuous, the sign of f' actually only *can* change at a critical point by the intermediate value theorem. So we just have to check the sign of f' at one point in between each critical point.

So what does this say about our previous example? We had three critical points, at $-1, 0, 2$. At -1 we saw that f' changed from negative to positive, so f has a relative minimum $f(-1) = 0$ at -1. Similarly, at 0 f' changed from positive to negative and at 2 f' changed from negative to positive, so f has a relative maximum of $f(0) = 5$ at 0 and a relative minimum of $f(2) = -27$ at 2.

Example 3.24. Let $g(x) = x + \sin(x)$. Then $g'(x) = 1 + \cos(x)$ is zero precisely when $x = (2n+1)\pi$ for some integer n. Since we only need to check the sign of g' at one point between each critical point, we check that $g'(2n\pi) = 1 + \cos(2n\pi) = 2$. Thus g' is positive everywhere except at the critical points, so g is increasing everywhere except at the critical points. Thus g has no relative maxima or minima.

Now let $h(x) = x + 2\sin(x)$. We have $h'(x) = 1 + 2\cos(x) = 0$ when $x = 2n\pi + 4\pi/3$ or $x = 2n\pi + 2\pi/3$. We compute that $h'(0) == 3, h'(\pi) = -1$, and $h'(2\pi) = 3$. Thus h' changes from positive to negative at $2\pi/3$, so this is a relative maximum. h' changes from negative to positive at $4\pi/3$, so this is a relative minimum.

Example 3.25. Let $f(x) = 2x^3 + 3x^2 - 36x$. Then $f'(x) = 6x^2 + 6x - 36 = 6(x^2 + x - 6) = 6(x+3)(x-2)$. The critical points are $-3, 2$. It's not hard to see that f' is positive if $x < -3$, is negative if $-3 < x < 2$, and is positive if $x > 2$. So f is increasing on $(-\infty, -3)$ and $(2, +\infty)$ and is decreasing on $(-3, 2)$.

Therefore f has a local max of $f(-3) = 81$ at -3 and a local min of $f(2) = -44$ at 2.

But we'd like to find relative maxima and minima with even less work, which brings us to the subject of concavity.

3.4 Concavity and the Second Derivative Test

Definition 3.26. We say a function f is *concave upward* on an interval (a, b) if every tangent line to a point in (a, b) lies below the graph of f.

We say a function f is *concave downard* on (a, b) if every tangent line to a point in (a, b) lies above the graph of f.

We say a point c is an *inflection point* for a function f if the graph of f changes from concave up to concave down, or concave down to concave up, at c.

Remark 3.27. Functions that are concave upward are curving up, like a bowl. Functions that are concave downward are curving down, like an umbrella.

Example 3.28. Looking at graphs, we can see:

- x^2 is concave upward everywhere. $-x^2$ is concave downward everywhere.
- x^3 is concave downward when $x < 0$ and is concave upward when $x > 0$.
- $\sqrt[3]{x}$ is concave upward when $x < 0$ and concave downward when $x > 0$.

- $\sin(x)$ is concave downward when $0 < x < \pi$ and concave upward when $\pi < x < 2\pi$.

We see that when a function is concave upward, the slopes of its tangent lines are increasing–which means the derivative is increasing. Similarly, a function is concave downward when its derivative is decreasing. But we just showed that we can determine whether a function is increasing or decreasing by looking at its derivative. So we need to study the derivative of the derivative–the second derivative.

Proposition 3.29 (Concavity Test)**.**
- *If $f''(x) > 0$ for all x in (a, b), then the graph of f is concave upward on (a, b).*
- *If $f''(x) < 0$ for all x in (a, b), then the graph of f is concave downward on (a, b).*

Remark 3.30. It's not necessarily true that f has an inflection point whenever $f''(x) = 0$. But it often is.

Example 3.31.
- $\frac{d}{dx}x^2 = 2x$, so $\frac{d^2}{dx^2}x^2 = 2 > 0$, so x^2 is concave upward everywhere. Similarly, $\frac{d^2}{dx^2} - x^2 = -2, 0$, so $-x^2$ is concave downward everywhere. Neither function has an inflection point.
- $\frac{d^2}{dx^2}x^3 = 6x$ is positive if $x > 0$ and negative if $x < 0$, so the function is concave upward when $x > 0$ and concave downward when $x < 0$. It has an inflection point when $x = 0$.
- $\frac{d^2}{dx^2}\sqrt[3]{x} = \frac{-2}{9\sqrt[3]{x^5}}$ is negative when $x > 0$ and positive when $x < 0$, so the function is concave upward when $x < 0$ and concave downward when $x > 0$. It has an inflection point when $x = 0$.
- $\frac{d^2}{dx^2}\sin(x) = -\sin(x)$, so $\sin(x)$ is concave upwards precisely when it is positive, and concave downwards when it is negative. It has an inflection point at $0, \pi, 2\pi$, and in general at $n\pi$ for any integer n.
- Consider $f(x) = x^4$. $f''(x) = 12x^2$ is positive everywhere except at 0, so the function is concave upwards everywhere except at 0. $f''(0) = 0$, so the second derivative concavity test doesn't tell us anything. But this isn't an inflection point, because the concavity doesn't change on either side–in fact the function is concave at $x = 0$ as well, as you can see from a graph.

Why do we care? Notice that if f is concave upward then the first derivative is increasing; so if $f'(c) = 0$ and f is concave upwards at c, the derivative is changing from negative to positive, and f has a local minimum at c. A similar argument works for local maxima, and thus:

Proposition 3.32 (The Second Derivative Test)**.** *If f'' is continuous near c, then*

- *If $f'(c) = 0$ and $f''(c) > 0$, then f has a local minimum at c.*
- *If $f'(c) = 0$ and $f''(c) < 0$, then f has a local maximum at c.*

Remark 3.33.

- If $f''(c) = 0$ this theorem tells us nothing; almost anything could happen. We can use the increasing/decreasing function test, or we can use the third and fourth derivatives to give us information.
- This rule only works if $f'(c) = 0$; if $f'(c)$ doesn't exist, then $f''(c)$ certainly doesn't exist and this proposition is not helpful.

Example 3.34. Let $f(x) = x^{2/3}(6-x)^{1/3}$. Where does f have relative maxima and minima? Where is it increasing or decreasing?

$$f'(x) = \frac{4-x}{x^{1/3}(6-x)^{2/3}}$$
$$f''(x) = \frac{-8}{x^{4/3}(6-x)^{5/3}}.$$

Then $f'(x) = 0$ when $x = 4$, and $f'(x)$ does not exist when $x = 0$ or $x = 6$, so these are the three critical points. We can again make a table:

	$4-x$	$x^{1/3}$	$(6-x)^{2/3}$	$f'(x)$
$x < 0$	$+$	$-$	$+$	$-$
$0 < x < 4$	$+$	$+$	$+$	$+$
$4 < x < 6$	$-$	$+$	$+$	$-$
$6 < x$	$-$	$+$	$+$	$-$

This tells us that f has a minimum of $f(0) = 0$ at 0 and a maximum of $f(4) = 2^{5/3}$ at 4. It doesn't have a local maximum or minimum at 6.

We can also use the second derivative test at 4 (but not 0 or 6—why?). We see that $f''(4) = \frac{-8}{2^{13/3}} = -2^{-4/3} < 0$ so f has a maximum at 4.

Further looking at $f''(x)$, we see that $x^{4/3} \geq 0$ for all x, and $(6-x)^{5/3} > 0$ when $x < 0$ or $0 < x < 6$, and $(6-x)^{5/3} < 0$ when $x > 6$. Thus $f''(x) < 0$ when $x < 6$ except at 0, and $f''(x) > 0$ when $x > 6$. So the function is concave down for $x < 6$ and concave up for $x > 6$, except at the points 0 and 6 where the derivative doesn't exist. There is a point of inflection at 6. This is enough information to sketch a graph of the function.

3.5 Curve sketching

And now we're ready to approach the task of sketching the graph of a function in an organized way. What follows is a good checklist, though not every point is relevant to every function.

1. Find the domain of the function. If it has holes, what happens near them? Does it go to infinity, or jump, or just skip a point?
2. Find the roots–where does the function hit the x-axis?
3. Find the limits as x goes to $\pm\infty$–what happens to the function "far away" from 0?
4. Compute f' and find the critical points. It can be helpful to evaluate f at the critical points.
5. Find intervals of increase or decrease. Identify local maxima and minima.
6. Compute f'' if you haven't already. Determine where the function is concave, and find inflection points.
7. Use all this information to sketch a graph of the function.

Example 3.35. Let $f(x) = x(x-4)^3 = x^4 - 12x^3 + 48x^2 - 64x$. Then:

1. The function is a polynomial, so its domain is all real numbers.
2. The function has roots at 0 and 4.
3. $\lim_{x\to+\infty} f(x) = \lim_{x\to-\infty} f(x) = +\infty$.
4. $f'(x) = (x-4)^3 + 3x(x-4)^2 = (x-4)^2(4x-4) = 4(x-1)(x-4)^2$. So $f'(x) = 0$ when $x = 1$ or $x = 4$. These are the critical points. $f(1) = -27$ and $f(0) = 0$.
5. Looking at our factorization, it's clear that $f'(x) < 0$ when $x < 1$ and $f'(x) > 0$ when $x > 1$, except $f'(x) = 0$ when $x = 4$. So f is decreasing when $x < 1$ and is increasing when $x > 1$ except at 4. Thus f has a minimum of -27 at 1.
6. $f''(x) = (x-4)^2 + 2(x-1)(x-4) = (x-4)(3x-6) = 3(x-2)(x-4)$. We see that $f''(x) > 0$ is $x < 2$ or $x > 4$, and $f''(x) < 0$ if $2 < x < 4$. Thus f is concave up on $(-\infty, 2)$ and $(4, +\infty)$, is concave down on $(2, 4)$, and has inflection points at 2 and 4.

Example 3.36. Let $g(x) = x\tan(x)$. Then

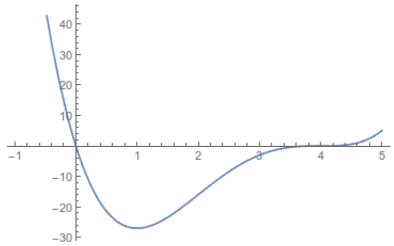


Figure 3.1: The graph of $f(x) = x(x-4)^3$

1. The domain of g is real numbers except $n\pi + \pi/2$. For simplicity we'll just look at x between $-\pi/2$ and $\pi/2$. $\lim_{x\to-\pi/2^+} g(x) = +\infty$ and $\lim_{x\to\pi/2^-} g(x) = +\infty$.

2. The function is 0 when $x = 0$ (and when $x = n\pi$ if we look farther out).

3. This isn't applicable since we're not looking out to $\pm\infty$.

4. $g'(x) = \tan(x) + x\sec^2(x) = \frac{\sin(x)\cos(x)+x}{\cos^2(x)}$. It's not hard to see that when $-\pi/2 < x < 0$ then $g'(x) < 0$, and when $0 < x < \pi/2$ then $g'(x) > 0$, and $g'(0) = 0$. So the only critical point is at 0.

5. And we saw that g is decreasing on $(-\pi/2, 0)$ and increasing on $(0, \pi/2)$. Thus g has a local minimum at 0. $g(0) = 0$.

6. $g''(x) = \sec^2(x) + \sec^2(x) + 2x\sec(x)\sec(x)\tan(x) = 2\sec^2(x)(1 + x\tan(x))$. $x\tan x \geq 0$ on $(-\pi/2, \pi/2)$, so the function is concave up everywhere.

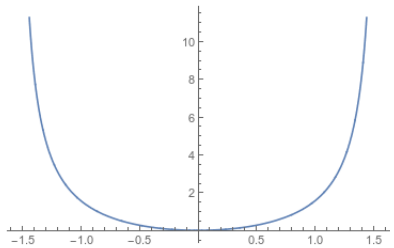


Figure 3.2: The graph of $g(x) = x\tan(x)$

Example 3.37. Let $h(x) = \frac{x+2}{x-1}$.

1. The domain of h is all real numbers except 1. We see that $\lim_{x\to1^-} h(x) = -\infty$ and $\lim_{x\to1^+} h(x) = +\infty$.

2. The function has a root at $x = -2$.

3. We have $\lim_{x\to+\infty} h(x) = \lim_{x\to-\infty} h(x) = 1$. (We can use L'Hôpital's rule or divide the top and bottom by x).

4. We have $h'(x) = \frac{(x-1)-(x+2)}{(x-1)^2} = -3(x-1)^{-2}$. This has no roots and fails to exist when $x = 1$. Thus there are no "real" critical points.

5. We make a chart for increase and decrease:

	-3	$(x-1)^{-2}$	$h'(x)$
$x < 1$	$-$	$+$	$-$
$1 < x$	$-$	$+$	$-$

   Thus h is decreasing everywhere. It has no local maxima or minima.

6. $h''(x) = 6(x-1)^{-3}$ is positive when $x > 1$ and negative when $x < 1$, so it is concave down on the left, and concave up on the right.

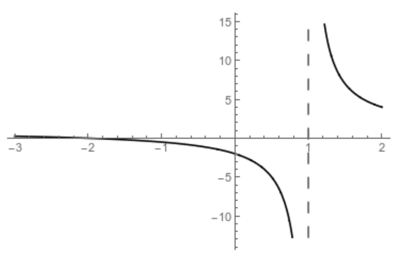


Figure 3.3: The graph of $h(x) = \frac{x+2}{x-1}$

Example 3.38.

- $f(x) = x^5 - 4x^3 + 4x + 7$
- $\frac{x^2-1}{x^2-4}$
- $\ln(x^2 - 3x + 2)$
- $\ln(1 + x^2) - x$
- Just picture: $\sin(x)\sin(1.1x)$ from -100 to 100.

3.6 Optimization

Through most of this section we've been finding the minimum and maximum values of functions purely to understand the functions. But the techniques used to maximize a function are extremely useful in finding optimum inputs to real world processes.

In other words, we're going to do more word problems.

Example 3.39. Suppose we have 2400 feet of fencing and we'd like to build a rectangular fence that encloses the most possible area. How can we do this?

If we have a rectangular fence, then one side will have a length L and another will have a width W. We know that the area $A = W \cdot L$ and that $2W + 2L = 2400$. So we can write $W = 1200 - L$ and see that $A = L(1200 - L)$. We'd like to maximize area.

We observe that our L has to be between 0 and 1200, so we're maximizing the function A on the closed interval $[0, 1200]$. By the extreme value theorem there must be some absolute maximum.

$A' = 1200 - 2L$. We see that the only critical point is $L = 600$. $A(0) = A(1200) = 0$ and $A(600) = 600^2 = 360,000$. $A(600)$ is the largest of these values, and so is the absolute max.

But what if we build the fence against a river, so we only need to build three sides? Then $A = W \cdot L$ but $W + 2L = 2400$, and thus $W = 2400 - 2L$. Then we have $A = L(2400 - 2L)$. A is still a function of L defined on $[0, 1200]$, and we compute $A' = 2400 - 4L$ and the only critical point is $L = 600$, again. $A(0) = A(1200) = 0$, and $A(600) = 600 \cdot 1200 = 720,000$. This last is the largest of the values, and the absolute max.

Example 3.40. Suppose we want to construct a cylindrical can that holds one liter of liquid, and we want to use the least possible metal to construct the can—and thus build the can with the least possible surface area. We have $A = 2\pi r^2 + 2\pi rh$.

To eliminate the h, we note that the can holds one liter or 1000 cm^3, and thus $\pi r^2 h = 1000$ and $h = \frac{1000}{\pi r^2}$. (We also could have written it as one cubic decimeter, but nobody ever works in decimeters). Thus we have $A = 2\pi r^2 + \frac{2000}{r}$.

$A' = 4\pi r - \frac{2000}{r^2} = \frac{4\pi r^3 - 2000}{r^2} = 0$ when $\pi r^3 = 500$, or when $r = \sqrt[3]{500/\pi}$. So this is the only critical point. Our function A has domain $(0, +\infty)$ so we can't use the extreme value theorem here. But we can see that A' is negative when $r < \sqrt[3]{500/\pi}$ and positive when $r > \sqrt[3]{500/\pi}$, so that must be a global minimum.

(Alternatively: $A'' = 4\pi + \frac{4000}{r^3}$ is always positive, so A is concave upwards everywhere, and has a unique minimum at its critical point).

But now what if the curved material for the sides costs more than the flat material for the ends, and we want to minimize cost? Say the material for the sides costs twice as much as material for the base. Then we have $C = 2\pi r^2 + \frac{4000}{r}$, and $C' = 4\pi r - \frac{4000}{r^2} = 0$ when $\pi r^3 = 1000$, when $r = 10/\sqrt[3]{\pi}$. This is the only critical point, and a similar argument to before shows it must be a global minimum.

We can break down our approach to these problems just as we did for related rates.

1. Draw a picture of the setup.

2. Create notation. Give names to all the quantities involved in the problem. Write down any equations that relate them.

3. Express the quantity you want to maximize or minimize as a function of the other quantities in the problem. Rewrite it so it's a function of a single variable.

4. Take the derivative and find the critical points.

5. Determine the absolute maximum or minimum.

6. Do a sanity check! Does your answer make sense?

Example 3.41. If we have 1200 cm^2 of cardboard to make a box with a square base and an open top, what is the largest possible volume of the box?

Well, we know that the total surface area of the box is $A = 1200$, and we also know that if the height of the box is h and the length of one of the base sides is b, then the area is $A = b^2 + 4bh$. So we can write $h = \frac{1200-b^2}{4b}$. We also know that the volume of the box is $V = b^2h$, so we have

$$\begin{aligned} V &= b^2 h = b^2 \frac{1200 - b^2}{4b} \\ &= 300b - b^3/4 \\ V' &= 300 - 3b^2/4 \\ 300 &= 3b^2/4 \\ 400 &= b^2 \\ 20 &= b \end{aligned}$$

so the only critical point occurs at 20. We see that $V(20) = 400 \cdot 10 = 4000$, so this is the largest possible volume of the box. (We can see that this is the absolute maximum via the Extreme Value Theorem, and observing that $V(0) = V(\sqrt{1200}) = 0$.

Example 3.42. Suppose a man wishes to cross a 20 m river and reach a house on the other side that is 48m downstream. The man can walk at 5 m/s or swim at 3 m/s. What is the optimal path for him to take to reach the house?

The man will swim for some point on the bank of the river, and then walk the other way. Let b be a number in $[0, 48]$ representing how far he travels towards the house. Then he travels $\sqrt{400 + b^2}$ meters in the river, at a speed of 3 m/s, and thus spends $\frac{1}{3}\sqrt{400 + b^2}$ seconds in the river. He then spends $(48 - b)/5$ seconds walking.

So total time spent is

$$\begin{aligned}
T &= \frac{\sqrt{400+b^2}}{3} + \frac{48-b}{5} \\
T' &= \frac{b}{3\sqrt{400+b^2}} - \frac{1}{5} \\
\frac{1}{5} &= \frac{b}{3\sqrt{400+b^2}} \\
3\sqrt{400+b^2} &= 5b \\
3600 + 9b^2 &= 25b^2 \\
225 &= b^2 \\
15 &= b
\end{aligned}$$

so we have a critical point at $b = 15$. On this path we have $T = 25/3 + 33/5 = (125+99)/15 = 224/15 \approx 14.9$ seconds.

What about the two other paths? If we head straight to the house, we travel $\sqrt{48^2 + 20^2} = 52$ meters at a speed of 3 m/s, for a total time of 17.3 seconds. If instead we head straight across the river to begin walking as soon as possible, we travel 20 m at 3 m/2 and then 48 m at 5 m/s, for a total time of $20/3 + 48/5 = (100 + 144)/15 = 244/15 \approx 16.3$ seconds. So the shortest path has us swim 25 m and deposits us 33 m from the house.

Example 3.43. A piece of wire 10 m long is going to be cut into two pieces. WE will fold one piece into a square and the other into an equilateral triangle. What is the largest joint area we can enclose? What is the smallest?

Let L be the length of the wire bent into a triangle (so that $10 - L$ is the length of the wire bent into a square). Then the area of the square is $A_1 = (10 - L)^2/16$. The area of the triangle is $bh/2$; the length of the base is $L/3$ and the height of the triangle is $\sin(\pi/3) \cdot L/3 = (1/2) \cdot (\sqrt{3}/2) \cdot L/3 = \sqrt{3}L/12$. So the area of the triangle is $A_2 = (1/2)(L/3)(\sqrt{3}L/6) = L^2\sqrt{3}/36$. Then we have

$$\begin{aligned}
A &= A_1 + A_2 = (100 - 20L + L^2)/16 + L^2\sqrt{3}/36 \\
A' &= -5/4 + L/8 + L\sqrt{3}/18 \\
5/4 &= L/8 + L\sqrt{3}/18 \\
90 &= 9L + 4\sqrt{3}L \\
L &= 90/(9 + 4\sqrt{3})
\end{aligned}$$

This is the only critical point. At that point,

$$A \approx 1.2 + 1.5 = 2.7.$$

But checking the endpoints, if we use all the wire for the square, we have area $A = 100/16 = 6.25$ and if we use all the wire for the triangle we have $A = 100\sqrt{3}/36 \approx 4.8$. So we get the biggest area when we use all the wire for the square, and the smallest if we use $90/(9+4\sqrt{3})$ m of wire for the triangle.

4 Approximation

This section is a bit of an interlude; it'll be a short bridge between section 3 on optimization, and section 5 on integration.

In this section we want to talk a bit more about the idea of approximation. We introduced this in section 1.3, when we talked about continuous approximation: if $x \approx a$, we can estimate $f(x) \approx f(a)$. We refined this a bit in section 2.1 and 2.6. The derivative allows us to estiamte that $f(x) \approx f(a) + f'(a)(x - a)$. But can we do even better?

4.1 Quadratic Approximation

In this class we've spent a lot of time on *linear approximation*: we can approximate a function with its tangent line, which is the linear function most similar to our starting function. This simplifies a lot of things, but is only an approximation.

$$f(x) \approx f(a) + f'(a)(x - a). \tag{2}$$

How good this approximation is depends on two things. The first is the distance $|x - a|$; the approximation is better when your goal point x is close to your starting point a. There are other techniques (like Fourier series) that don't have this limitation, but we won't discuss them in this course.

The other is the speed at which the derivative changes. If the derivative is constant, your function is just a line and the "approximation" is perfect. But the faster the derivative changes, the faster the function deviates from the line.

Thus we might try to get a better approximation using the second derivative, which tells us how quickly the derivative is changing. So how can we do this?

We're looking for some function $g(x)$ so that

$$f(x) \approx f(a) + f'(a)(x - a) + g(a)(x - a)^2.$$

(We want the linear approximation to be the same as (4), and we want the third derivative to be zero, so the only thing that can change at all is the degree two term). Taking derivatives of both sides gives us

$$\begin{aligned} f'(x) &\approx f'(a) + 2g(a)(x - a) \\ f''(x) &\approx 2g(a). \end{aligned}$$

Thus we set $g(a) = f''(x)/2$, and we get the equation

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2. \tag{3}$$

This is the *parabola* that best approximates our function near a.

Example 4.1. Let's again ask our old question: what is $\sqrt{5}$?

We use the function $f(x) = \sqrt{x}$ and we compute $f'(x) = \frac{1}{2\sqrt{x}}$ and $f''(x) = \frac{-1}{4\sqrt{x^3}}$. Then we have

$$\begin{aligned}
f'(4) &= \frac{1}{4} \\
f''(4) &= \frac{-1}{32} \\
f(x) &\approx f(4) + f'(4)(x-4) + \frac{f''(4)}{2}(x-4)^2 \\
&= 2 + \frac{1}{4}(x-4) - \frac{1}{64}(x-4)^2 \\
f(5) &\approx 2 + \frac{1}{4} - \frac{1}{64} = 2 + \frac{15}{64} \approx 2.23483.
\end{aligned}$$

We see we've slightly overcorrected: rather than being .014 too big, we're now .0012 too small.

Example 4.2. Compute the quadratic approximations of $\sin(x)$ and $\cos(x)$ centered at zero. Estimate $\sin(.01)$ and $\cos(.01)$? How does this relate to the Small Angle Approximation?

$$\begin{aligned}
\sin'(x) &= \cos(x) \\
\sin'(0) &= 1 \\
\sin''(x) &= -\sin(x) \\
\sin''(0) &= 0 \\
\sin(x) &\approx 0 + 1(x-0) + \frac{0}{2}(x-0)^2 = x \\
\sin(.01) &\approx .01.
\end{aligned}$$

Recall the small angle approximation told us that $\sin(x) \approx x$. Here we see that this is not just a linear approximation, but in fact also the quadratic approximation; the reason the small angle approximation worked so well is that it was correct ot second order.

$$
\begin{aligned}
\cos'(x) &= -\sin(x) \\
\cos'(0) &= 0 \\
\cos''(x) &= -\cos(x) \\
\cos''(0) &= -1 \\
\cos(x) &\approx 1 + 0(x-0) - 1(x-0)^2 = 1 - \frac{x^2}{2} \\
\cos(.01) &\approx .99995.
\end{aligned}
$$

Example 4.3. Let $g(x) = x^4 - 3x^3 + 4x^2 + 4x - 2$. Compute the quadratic approximations at $a = 0$ and at $a = -2$. Compare them to $g(x)$. Estimate $g(-1.97)$.

$$
\begin{aligned}
g(0) &= -2 \\
g'(x) &= 4x^3 - 9x^2 + 8x + 4 \\
g'(0) &= 4 \\
g''(x) &= 12x^2 - 18x + 8 \\
g''(0) &= 8 \\
g(x) &\approx -2 + 4(x-0) + \frac{8}{x}x^2 = 4x^2 + 4x - 2.
\end{aligned}
$$

Notice that this is just the lower-degree terms of our original polynomial!

$$
\begin{aligned}
g(-2) &= 16 + 24 + 16 - 8 - 2 = 46 \\
g'(x) &= 4x^3 - 9x^2 + 8x + 4 \\
g'(-2) &= -32 - 24 - 16 + 4 = -80 \\
g''(x) &= 12x^2 - 18x + 8 \\
g''(-2) &= 48 + 36 + 8 = 92 \\
g(x) &\approx 46 - 80(x+2) + 46(x+2)^2 \\
f(-1.97) &\approx 46 - 80(.03) + 46(.009) = 43.6414.
\end{aligned}
$$

However, if we take $h(x) = 4x^2 + 4x - 2$ and approximate near -2, we get

$$\begin{aligned}
h(-2) &= 6 \\
h'(x) &= 8x + 4 \\
h'(-2) &= -12 \\
h''(x) &= 8 \\
h''(-2) &= 8 \\
h(x) &\approx 6 - 12(x+2) + 4(x+2)^2 = 6 - 12x - 24 + 4x^2 + 16x + 16 \\
&= 4x^2 + 4x - 2 = h(x).
\end{aligned}$$

No matter where we center our approximation, the best quadratic approximation to our parabola is our original parabola.

Example 4.4. Now let's estimate 1.01^{25} using a quadratic approximation. We use the function $f(x) = (1+x)^{25}$, and center our approximation at $x = 0$. (Equivalently we could consider $g(x) = x^{25}$ and center our approximation at $x = 1$; the way I set it up is a bit more common).

We take $f'(x) = 25(1+x)^{24}$ so $f'(0) = 25$, and $f''(x) = 25 \cdot 24(1+x)^{23}$ so $f''(0) = 25 \cdot 24 = 600$. Then we have

$$\begin{aligned}
f(x) &\approx 1 + 25(x-0) + \frac{600}{2}(x-0)^2 = 1 + 25x + 300x^2 \\
1.01^{25} = f(.01) &\approx 1 + 25 \cdot .01 + 300 \cdot .0001 = 1 + .25 + .03 = 1.28.
\end{aligned}$$

Since $1.01^{25} \approx 1.28243$ this is pretty good.

What if we move a bit farther? If we want to estimate 1.04^{25} we get

$$1.04^{25} = f(.04) \approx 1 + 25 \cdot .04 + 300 \cdot .0016 = 1 + 1 + .48 = 2.48$$

while $1.04^{25} \approx 2.66584$. We've lost fidelity because our move away is bigger.

But while .4 is still much smaller than 1, this estimate is much worse than our estimate of $\sqrt{5}$ from earlier. Why is this much worse? Linear are bad for two reasons: either because x and a are far apart, or because the second derivative is large. Here we've taken care of the second derivative, but we haven't taken care of everything. Our quadratic approximations will be bad when the *third* derivative is large.

Finally, let's use this to estimate 2^{25}. We get

$$2^{25} = f(1) \approx 1 + 25 \cdot 1 + 300 \cdot 1^2 = 326.$$

But $2^{25} = 33,554,432$, so this is very far off. We see here even more problems with the largeness of the higher derivatives.

4.1.1 Cubics and Beyond: Taylor Series

We can carry this logic further. We can work out that if we want to match the first *three* derivatives and get a cubic approximation, we get the formula

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f'''(a)}{3 \cdot 2}(x-a)^3.$$

More generally, we can get a degree-n polynomial approximation, called the *Taylor polynomial of degree n*, with the formula

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f'''(a)}{3 \cdot 2}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

If a function is infinitely differentiable, we can take an infinite sum here and get the *Taylor series*:

$$T_f(x,a) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \ldots.$$

Most functions we're interested in are equal to their own Taylor series. (Not all functions are, though!) In particular, we can work out the following formulas:

$$\begin{aligned} \sin(x) &= x - \frac{x^3}{6} + \frac{x^5}{120} - \frac{x^7}{7!} + \ldots \\ \cos(x) &= 1 - \frac{x^2}{2} + \frac{x^4}{24} - \frac{x^6}{720} + \ldots \\ e^x &= 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \ldots. \end{aligned}$$

Taylor series are extremely important in nay sort of computational or advanced math, and you will talk about them a lot more if you take Calculus II.

However, in practice, just like we rarely use third or fourth derivatives, we rarely use approximations of degree higher than two. If the quadratic approximation doesn't pick up whatever you need to think about, we will do something else entirely.

4.2 Iterative Approximation: Newton's Method

In section 2.6 we saw that there were two things that make a linear approximation work better or worse. The first was the size of the second derivative; in section 4.1 we leveraged the second derivative to improve our approximations.

To keep things simple, we'll assume that we want to solve $f(x) = 0$. (If not, we can just subtract our number y from both sides of the equation). If we know the value of f and of f' at a point x_0, then recall that by linear approximation we estimate that $f(x_1) =$

$f(x_0) + f'(x_0)(x_1 - x_0)$. Since we want $f(x_1) = 0$, we set $f(x_1) = 0$ and solve this equation for x_1, and get

$$x_1 = x_0 - (f(x_0)/f'(x_0)).$$

In many conditions, we will get the result that x_1 is closer to being a root of f than x_0 is.

We can repeat this process to find x_2, x_3, etc., and ideally each will be a better estimate than the previous estimate was. A good rule of thumb for when to stop: if you want five decimal places of accuracy, you can stop when the nth step and the $n+1$st step agree to five decimal places.

This method does have limitations. First, we have to start with a guess x_1 for our root x. Second, if $f'(x_1)$ is very close to zero, Newton's method will work poorly if it works at all, and we might have to pick a better guess. But it can be very useful for finding approximate solutions to equations.

Example 4.5. Let's approximate the square root of 5, one more time. First, we need to turn this into finding a solution to an equation. We want to solve the equation $x^2 = 5$, which we can rewrite as $f(x) = x^2 - 5 = 0$. We compute $f'(x) = 2x$.

We need to pick a starting estimate, which should probably be $x_0 = 2$. Then we have $f(x_0) = -1$, and $f'(x_0) = 4$. So we get

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 2 - \frac{-1}{4} = 9/4 = 2.25.$$

You might notice that this is exactly what we got by doing a simple linear approximation. So what did we get from this new method? Now we can *iterate*.

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 9/4 - \frac{81/16 - 5}{9/2} = 161/72 \approx 2.23611$$
$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = 161/72 - \frac{1/5184}{161/36} = \frac{51851}{23184} \approx 2.23607$$

Checking with a computer tells us that $\sqrt{5} \approx 2.23607$, so we're now correct to five decimal places.

5 Integration

5.1 The Area Problem

For the next couple months, we will primarily be occupied by the question of *area*.

What is area? This actually gets a little fuzzy. We know how to compute the area of a rectangle: base times height. From that fact, and drawing a quick picture, we know the area of triangle: $\frac{1}{2}bh$, since it's half a rectangle.

We also know the area of a circle. But how? What about an ellipse? Or something funny-looking and squiggly? What does "area" mean, exactly, in these cases?

To measure the area of a shape, we can try filling it up with small squares or rectangles—we know how to measure those. (Similar principle: if you need to measure the length of something curved, run a string along it, straighten it out, measure the string. We'll actually return to this in the second half of the class).

We're going to make our lives easier, and assume our shape has one straight side. (This isn't as strict a condition as it seems; we can always cut our shape in half. We'll talk a lot more about that later on). In fact, let's look at shapes that are given by graphs of functions.

Definition 5.1. A function is a rule that gives one output for each input. The graph of a function is all the points $(x, f(x))$. In particular, the graph passes the *vertical line test*: if you draw a vertical line through the graph of a function, it intersects it at most one time.

We want to find the area of the shape "under" the graph. For right now we'll assume the function is always positive, so we get an actual area of an actual shape. (We'll relax that assumption very soon).

When we were trying to get areas earlier, we used a lot of rectangles. We can fill this area with rectangles in a bunch of different ways. But one particular way turns out to work very well, which is to have a bunch of tall skinny rectangles.

So what's the area of these rectangles? If a rectangle goes from a to b, then its width is $b - a$. How tall is it? That depends on where we put the top. There are a few things we can do, but the easiest is to make one of the top corners lie exactly on the graph. If we pick the right corner, then the width is $(b - a)f(b)$.

Example 5.2. Let's find the area under the curve $y = x^2$, between 0 and 1. If we use just one rectangle, with width 1, then we get either 0 or 1. This is true, but not super helpful.

Let's try two rectangles. They each are $\frac{1}{2}$ wide. If we line up the right-hand corners, then the area of thefirst one is $\frac{1}{2} \cdot \frac{1}{2}^2 = \frac{1}{8}$, and the area of the second one is $\frac{1}{2} \cdot 1^2 = \frac{1}{2}$. We get a total area of $\frac{5}{8}$.

What if we used the left-hand corners instead? Then the first rectangle is $\frac{1}{2} \cdot 0^2 = 0$ and the second is $\frac{1}{2} \cdot \frac{1}{2}^2 = \frac{1}{8}$. So the "true" area is somewhere between $\frac{1}{8}$ and $\frac{5}{8}$.

Let's get skinnier. If we use four rectangles, then with the right-hand point, we get

$$A_R \approx \frac{1}{4} \cdot \frac{1}{4}^2 + \frac{1}{4} \cdot \frac{1}{2}^2 + \frac{1}{4} \cdot \frac{3}{4}^2 + \frac{1}{4} \cdot 1^2 = \frac{1}{64} + \frac{1}{16} + \frac{9}{64} + \frac{1}{4} = \frac{30}{64} = \frac{15}{32},$$

and if we line up the left-hand point instead, we get

$$A_L \approx \frac{1}{4} \cdot 0^2 + \frac{1}{4} \cdot \frac{1}{4}^2 + \frac{1}{4} \cdot \frac{1}{2}^2 + \frac{1}{4} \cdot \frac{3}{4}^2 = 0 + \frac{1}{64} + \frac{1}{16} + \frac{9}{64} = \frac{14}{64} = \frac{7}{32}.$$

So the "true" area is between $\frac{7}{32}$ and $\frac{15}{32}$.

Notice that as we draw more rectangles, these numbers are getting closer. If we use 8 rectangles, we see the area is between $\frac{35}{128}$ and $\frac{51}{128}$, and if we use 64 we find that the area is between .326 and .341.

You can probably guess what happens as the number of rectangles gets very big, but let's work it out. If we have n rectangles, then each one has width $1/n$, and if we use the right-hand approximation then each rectangle has height $\left(\frac{i}{n}\right)^2$. So we have

$$\begin{aligned} R_n &= \frac{1}{n} \cdot \frac{1}{n}^2 + \frac{1}{n} \cdot \frac{2}{n}^2 + \cdots + \frac{1}{n} \cdot \frac{n}{n}^2 \\ &= \frac{1}{n^3}\left(1^2 + 2^2 + \cdots + n^2\right) \\ &= \frac{1}{n^3} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6n^2}. \end{aligned}$$

(We had to use a "sum of squares" formula to get to the third line; feel free to check it on your own, but don't worry about it too much. It's in Appendix E of your textbook if you want to check it out).

What happens to R_n as n gets large? From what we learned about limits in section 1.5, we can compute that this limit is $\frac{1}{3}$.

We can generalize this process to define exactly what we mean by the area under a curve.

Definition 5.3. We define the area under a curve to be the limit of the sums of the areas of these rectangles. We write

$$A = \lim_{n \to +\infty} R_n = \lim_{n \to +\infty} \left(f(x_1)\Delta x + f(x_2)\Delta x + \cdots + f(x_n)\Delta x \right).$$

Here n is the number of rectangles, and Δx is the width of each rectangle. Thus $\Delta x = \frac{L}{n}$ where L is the length of our shape.

Example 5.4. Estimate the area under the curve of $f(x) = 2x$ between $x = 1$ and $x = 4$, using three rectangles and using six rectangles. Try using both right endpoints and left endpoints. Is it what you expected?

$$\begin{aligned} R_3 &= \frac{3}{3}(4+6+8) = 18. \\ L_3 &= \frac{3}{3}(2+4+6) = 12. \\ R_6 &= \frac{3}{6}(3+4+5+6+7+8) = 16.5. \\ L_6 &= \frac{3}{6}(2+3+4+5+6+7) = 13.5. \end{aligned}$$

What if the number of rectangles goes to infinity? We have

$$\begin{aligned} R_n &= \frac{3}{n}f(1+3/n) + \frac{3}{n}f(1+2\cdot 3/n) + \cdots + \frac{3}{n}f(1+n\cdot 3/n) \\ &= \frac{3}{n}\left(2+2\frac{3}{n}+2+4\frac{3}{n}+\cdots+2+2n\frac{3}{n}\right) \\ &= \frac{3}{n}(2+\cdots+2) + \frac{3}{n}\left(2\frac{3}{n}+4\frac{3}{n}+\cdots+2n\frac{3}{n}\right) \\ &= 6 + \frac{18}{n^2}(1+2+\cdots+n) \\ &= 6 + \frac{18}{n^2}\frac{n(n+1)}{2} = 6 + 9\frac{n+1}{n}. \end{aligned}$$

We check that this formula still works for 3 and 6. Then we take the limit:

$$\lim_{n\to+\infty} R_n = \lim_{n\to+\infty} 6 + 9\frac{n+1}{n} = 6 + 9\lim_{n\to+\infty}\frac{1+\frac{1}{n}}{1} = 15.$$

This makes sense, since using the area formula for triangles we get an area of 15. (It's a 4×8 triangle minus a 1×2 triangle).

5.2 Riemann Sums and The Definite Integral

5.2.1 A brief note on summation notation

For the next couple weeks we'll be writing a lot of sums, and we'd like to have notation to talk about this.

We write $\sum_{i=1}^n a_i$ for $a_1 + a_2 + \cdots + a_n$ to be the sum of a bunch of things. We can index the sums other ways—and in particular, sometimes it's helpful to start from 0 instead of from 1.

You'll learn a lot more about sums in Calculus 2, but for right now, here are a few useful facts:

- $\sum_{i=1}^n c = nc$.
- $\sum_{1=1}^n ca_i = c\sum_{i=1}^n a_i$.
- $\sum_{i=1}^n (a_i \pm b_i) = (\sum_{i=1}^n a_i) \pm (\sum_{i=1}^n b_i)$.
- $\sum_{i=1}^n i = \frac{n(n+1)}{2}$.
- $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$.
- $\sum_{i=1}^n i^3 = \left(\frac{n(n+1)}{2}\right)^2$.

5.2.2 Signed Area

Last class we talked about finding the area under a curve. But a lot of functions are sometimes negative. We want a formalism that lets us keep track of this.

Definition 5.5. The *signed area* under a graph is the area below the graph but above the x-axis, minus the area below the x-axis and above the graph.

You can think of this as the "net area". If a rectangle with a positive height has a positive area, then a rectangle with a negative height has a negative area.

5.2.3 Back to Riemann Sums

Suppose f is a function defined on a closed interval $[a, b]$. We divide $[a, b]$ into n smaller subintervals by picking points $a = x_0 < x_1 < \cdots < x_n = b$. We get a collection of subintervals $[x_0, x_1], [x_1, x_2], \ldots, [x_{n-1}, x_n]$, which we call a *partition* P of $[a, b]$. We will also sometimes use Δx_i to refer to the length $x_i - x_{i-1}$ of the ith subinterval in our partition.

For each subinterval, we can pick a *sample point* x_i^* in the interval. We could use the left endpoints or the right endpoints, as we did last class, or we could pick others; for most of our purposes in this class it doesn't really matter. (In lab next week we'll talk about what to do when it does matter).

Definition 5.6. The *Riemann sum* associated to a partion P and a function f on an interval $[a, b]$ is given by

$$R(P, f) = \sum_{i=1}^n f(x_i^*)\Delta x_i = f(x_1^*)\Delta x_1 + f(x_2^*)\Delta x_2 + \cdots + f(x_n^*)\Delta x_n.$$

This gives an approximation to the signed area under the graph of f.

We can think about taking the limit as our partition gets very small—as we use more and more rectangles and the width of each gets close to 0. We define

Definition 5.7. If f is a function defined on $[a, b]$, the *definite integral of f from a to b* is

$$\int_a^b f(x)\,dx = \lim_{P \to 0} R(P, f) = \lim_{\max \Delta x_i \to 0} \sum_{i=1}^n f(x_i^*)\Delta x_i,$$

if the limit exists. If the limit exists, we say f is *integrable* on $[a, b]$. (otherwise, f is not integrable).

We say a is the *lower limit* of the integral, b is the *upper limit*, and $f(x)$ is the *integrand*.

Remark 5.8. It's important to note that while there are xs inside or "under" the integral sign, after the integral is computed there are no xs left. The x is a "dummy variable" or a "parameter." We'd get the exact same answer if we calculated $\int_a^b f(t)\,dt$ or $\int_a^b f(\text{🦇})\,d\text{🦇}$ or $\int_a^b f(thisisavariable)\,dthisisavariable$.

In our definition, we took the limit over "all" partitions. This is hard to work with in practice, since there are a lot of partitions. (There are infinitely many partitions of $[0, 1]$, for instance, where $x_1 = .99999$. These are in fact partitions but they aren't incredibly helpful).

But if a function is integrable, we can always do our calculations using any collection of partitions that gets small. In particular there's one nice partition we will often use:

Theorem 5.9. *If f is integrable on $[a, b]$, then*

$$\int_a^b f(x)\,dx = \lim_{n \to +\infty} \sum_{i=1}^n f(x_i)\Delta x$$

where $\Delta x = \frac{b-a}{n}$ and $x_i = a + i\Delta x$. That is,

$$\int_a^b f(x)\,dx = \lim_{n \to +\infty} \sum_{i=1}^n f\left(a + (b-a)\frac{i}{n}\right)\frac{b-a}{n}.$$

In some sense, the dx corresponds to the Δx and the $f(x)$ corresponds to the $f(x_i^*)$. This can be made rigorous, but probably won't be in this course.

Example 5.10.

$$
\begin{aligned}
\int_3^5 x^2\,dx &= \lim_{n\to+\infty} \sum_{i=1}^n \left(3+\frac{2i}{n}\right)^2 \frac{2}{n} \\
&= \lim_{n\to+\infty} \sum_{i=1}^n \left(9+\frac{12i}{n}+\frac{4i^2}{n^2}\right)\frac{2}{n} \\
&= \lim_{n\to+\infty} \sum_{i=1}^n \frac{18}{n}+\frac{24i}{n^2}+\frac{8i^2}{n^3} \\
&= \lim_{n\to+\infty} \left(\sum_{i=1}^n \frac{18}{n}+\sum_{i=1}^n\frac{24i}{n^2}+\sum_{i=1}^n\frac{8i^2}{n^3}\right) \\
&= \lim_{n\to+\infty} \left(\frac{18}{n}\sum_{i=1}^n 1+\frac{24}{n^2}\sum_{i=1}^n i+\frac{8}{n^3}\sum_{i=1}^n i^2\right) \\
&= \lim_{n\to+\infty} \left(\frac{18}{n}\cdot n+\frac{24}{n^2}\cdot\frac{n(n+1)}{2}+\frac{8}{n^3}\cdot\frac{(n)(n+1)(2n+1)}{6}\right) \\
&= \lim_{n\to+\infty} \left(18+12\frac{n(n+1)}{n^2}+\frac{4}{3}\cdot\frac{n(n+1)(2n+1)}{n^3}\right) \\
&= 18+12+\frac{8}{3}=\frac{98}{3}\approx 32.7.
\end{aligned}
$$

Proposition 5.11 (Properties of the Integral)**.** *The following equations are true whenever they make sense, for real numbers a, b, c and functions f, g.*

- $\int_a^b c\,dx = c(b-a)$.
- $\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx$.
- $\int_a^b (f(x)\pm g(x))\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$.
- $\int_a^b cf(x)\,dx = c\int_z^b f(x)\,dx$.
- $\int_a^c f(x)\,dx + \int_c^b f(x)\,dx = \int_a^b f(x)\,dx$.

Remark 5.12. These properties are derivable from the corresponding properties of sums.

Remark 5.13. Note that while addition and scalar multiplication behave nicely, we didn't make any statements about multiplication or division, because integrals don't actually behave nicely with respect to multiplication. (We call operations like this "linear," and we study them in Math 2184).

In Calculus 2, you will return to the idea of "the integral of the product of two functions" when you study integration by parts. But we won't quite get to that in this course.

Example 5.14. Compute $\int_1^0 2+3x^2+4x^3\,dx$.

By these integral properties, we know that

$$\begin{aligned}\int_1^0 2+3x^2+4x^3\,dx &= -\int_0^1 2+3x^2+4x^3\,dx \\ &= -\int_0^1 2 - \int_0^1 3x^2 - \int_0^1 4x^3\,dx \\ &= -\int_0^1 2 - 3\int_0^1 x^2 - 4\int_0^1 x^3\,dx \\ &= -2-3(1/3)-4(1/4) = 4.\end{aligned}$$

Example 5.15. If $\int_1^5 f(x)\,dx = 3$ and $\int_3^5 f(x)\,dx = 2$, then

$$\int_1^3 f(x)\,dx = 1 = \int_1^5 f(x)\,dx + \int_5^3 f(x)\,dx = \int_1^5 f(x)\,dx - \int_3^5 f(x)\,dx = 3-2=1.$$

Proposition 5.16 (Comparison Propreties of the Integral)**.** *These properties only work when $a<b$. If we have a case where $a>b$ then we can always rewrite the integral before using them.*

- *If $f(x)\geq 0$ for $a\leq x\leq b$ then $\int_a^b f(x)\,dx \geq 0$.*
- *If $m\leq f(x)\leq M$ for $a\leq x\leq b$ then $m(b-a)\leq \int_a^b f(x)\,dx \leq M(b-a)$.*
- *If $f(x)\geq g(x)$ for $a\leq x\leq b$ then $\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx$.*

Example 5.17. We've used these implicitly before, when e.g. we said that $0\leq \int_0^1 x^2 \leq 1$.

Referencing our earlier example, we know that $9\leq x^2\leq 25$ on $[3,5]$, so we have $18\leq \int_3^5 x^2\,dx \leq 50$. Indeed, we calculated that $\int_3^5 x^2\,dx \approx 33$.

Suppose we want to know about $\int_0^\pi \sin(x)\,dx$. We know that $0\leq \sin(x)\leq 1$ on $[0,\pi]$, so we see that $0\leq \int_0^\pi \sin(x)\,dx \leq \pi$. (In fact, the integral is equal to 2, but we don't yet have the tools to calculate that).

5.3 The Fundamental Theorem of Calculus Part 1

From this perspective, the definite integral $\int_a^b f(t)\,dt$ is always a number, as long as f is integrable.(Technically the integral is a function from the set of integrable functions to the set of real numbers, but we don't need to worry about that in this class). In fact the integral is just "the area of a shape I just described," so it should always be a number. If I asked you for the area of a shape you shouldn't ever tell me $y=x^2$, for instance.

But we can use the integral to define a function (in the same way that we can have the function "input a number x and return the area of a square with side length x"—that is, $f(x) = x^2$). In particular, we want to consider functions of the form

$$F(x) = \int_a^x f(t)\,dt \tag{4}$$

where a is some fixed constant, and x is a variable. So our function is "put in a number x, and output the number $\int_a^x f(t)\,dt$, which is the area of some shape, determined by x."

Now that we have a function, there are a bunch of questions we can ask about it. What is its domain? Is it continuous? Is it differentiable?

The domain of $F(x) = \int_a^x f(t)\,dt$ is all x so that f is integrable on $[a, x]$; this answer isn't terribly satisfying, since it boils down to "The domain of F is the domain of F." It's not possible to do better without knowing something about f. But if we impose a fairly mild condition, we can say a bit more:

Theorem 5.18. *If f is continuous on $[a, b]$, or if it is continuous except for finitely many jump discontinuities, then f is integrable on $[a, b]$.*

Sketch of proof. If f has finitely many jump discontinuities, we can pick our partition to chop it up into a finite collection of continuous functions. So we just have to worry about continuous functions.

For any partition, you can always pick a "biggest" sample point in each interval, and a "smallest." The first will give you an upper bound to the integral, and the second will give you a lower bound. If the function is continuous, we can show that those two sums will always get closer together, and every other possible sum will be between the two; so all possible sums converge to the same integral. □

Example 5.19. $f(x) = x^n$ is integrable, as is $|x|$ and $\sqrt[n]{x}$ on any interval on which it is defined. The Heaviside (step) function is integrable. $1/x$ is not integrable on $[0, 1]$. The characteristic function of the rationals is not integrable (At least, not until grad school, when they change the definitions on you).

We can see a bit more. It's not too hard to show that F is continuous on its domain. Geometrically, changing x a little bit will change $F(x)$ by about the height of the function times the change in input; if the change in input is small, the change in output will also be small. Algebraically:

$$\begin{aligned}
\lim_{x\to b} F(x) - F(b) &= \lim_{x\to b} \int_a^x f(t)\,dt - \int_a^b f(t)\,dt \\
&= \lim_{x\to b} \int_a^x f(t)\,dt + \int_b^a f(t)\,dt \\
&= \lim_{x\to b} \int_b^x f(t)\,dt.
\end{aligned}$$

If x and b are close enough we can always find m, M such that $m \leq f(t) \leq m$ on $[x, b]$, so we get

$$\begin{aligned}
\lim_{x\to b} m(x-b) \leq \lim_{x\to b} &\int_b^x f(t)\,dt \leq \lim_{x\to b} M(x-b) \\
0 \leq \lim_{x\to b} &\int_b^x f(t)\,dt \leq 0 \\
0 = \lim_{x\to b} &\int_b^x f(t)\,dt.
\end{aligned}$$

The question of differentiability is a little trickier, but significantly more important. Intuitively and geometrically, we can simply look at pictures and ask how much the area under a curve changes if we widen our x-values a bit. After drawing some pictures we conclude that the area should change by "about" the height of the curve on one end.

We can in fact prove this fact. It's important enough for us to give it a silly name:

Theorem 5.20 (The Fundamental Theorem of Calculus, Part 1). *Suppose f is continuous on $[a, b]$, and set*

$$F(x) = \int_a^x f(t)\,dt.$$

Then $\frac{d}{dx}F(x) = f(x)$ for $a < x < b$.

Remark 5.21. As we'll discuss shortly, this theorem is the key to calculuating integrals. Note that it only applies to continuous functions. But if we have a function that's continuous in pieces, we can just split it up into separate integrals, and we see it has the correct derivative on each piece.

Proof. We want to capture our geometric intuitions. Recall that by definition, we have

$$\begin{aligned}
F'(x) &= \lim_{h\to 0} \frac{\int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt}{h} \\
&= \frac{1}{h} \int_x^{x+h} f(t)\,dt.
\end{aligned}$$

(This calculuation should look similar to the one above for continuity.) Let's assume for now that $h > 0$. By the extreme value theorem, f has an absolute minimum m and an absolute maximum M on $[x, x+h]$, and further we can write $f(u) = m$ and $f(v) = M$ for u, v in $[x, x+h]$. Then

$$f(u)h \leq \int_x^{x+h} f(t)\,dt \leq f(v)h$$
$$f(u) \leq \frac{1}{h}\int_x^{x+h} f(t)\,dt \leq f(v).$$

As $h \to 0$, the numbers u and v must get closer together, and in fact closer to x, and so by continuity $\lim_{h\to 0} f(u) = \lim_{h\to 0} f(v) = f(x)$. So we have $F'(x) = \lim_{h\to 0} \frac{1}{h}\int_x^{x+h} f(t)\,dt = f(x)$ as desired. ☐

Example 5.22.

- If $F(x) = \int_a^x \sqrt{x^3+1}\,dt$ then $F'(x) = \sqrt{x^3+1}$.
- If $G(x) = \int_a^x \sin(\pi t)\cos(\pi t)\,dt$ then $G'(x) = \sin(\pi x)\cos(\pi x)$.
- If $H(x) = \int_a^{x^3} \sqrt{1+t}\,dt$ then we have to be careful. We can write $H(x) = H_1(x^3)$ where $H_1(x) = \int_a^x \sqrt{1+t}\,dt$. So by the chain rule, we have $H'(x) = \sqrt{1+x^3}\cdot 3x^2$.

5.4 Computing Integrals and the FTC 2

We still haven't quite figured out how to compute integrals without going back to the Riemann sum formulation. But we're almost there!

The Fundamental Theorem of Calculus tells us that $\frac{d}{dx}\int_a^x f(t)\,dt = f(x)$. But it isn't the only function with this property. We can give this a name:

Definition 5.23. If $F'(x) = f(x)$, we call F an *antiderivative* of f.

Example 5.24. $\frac{1}{3}x^3$ is an antiderivative of x^2.

$\sin(x)$ is an antiderivative of $\cos(x)$.

7 is an antiderivative of 0.

So $\int_a^x f(t)\,dt$ is an antiderivative of f. Further, we know a lot about what antiderivatives look like:

Proposition 5.25. *If $F'(x) = G'(x)$ for all x, then $F(x) = G(x) + C$ for some constant C.*

Proof. Differentiation is additive, so $(F-G)'(x) = F'(x) - G'(x) = 0$. But since the derivative is the rate of change, any function with zero derivative is constant. (We proved this in proposition 3.19 in section 3.3, using the Mean Value Theorem.) Thus $(F-G)(x) = C$ for some constant C, and so $F(x) = G(x) + C$. ☐

This proposition incredibly useful, because it means *any* function whose derivative is $f(x)$ is "almost" the same as $\int_a^x f(t)\,dt$. We have some sort of constant hanging around, which we need to get rid of; it turns out that this constant is essentially related to the a, the lower limit of integration.

Theorem 5.26 (Fundamental Theorem of Calculus, Part 2)**.** *Suppose f is continuous on $[a, b]$, and F is any antiderivative of f. Then*

$$\int_a^b f(t)\,dt = F(b) - F(a).$$

Proof. Since $F(x)$ and $\int_a^x f(t)\,dt$ are both antiderivatives of $f(x)$, we know that $F(x) = \int_a^x f(t)\,dt + C$ for some constant C. Then

$$F(b) - F(a) = \int_a^b f(t)\,dt + C - \left(\int_a^a f(t)\,dt + C\right) = \int_a^b f(t)\,dt + C - 0 - C = \int_a^b f(t)\,dt.$$

□

Example 5.27. What is $\int_1^3 3x^2\,dx$?

We can see that $F(x) = x^3$ is an antiderivative of $3x^2$. (It's not the only one, but that's okay.) So $\int_1^3 3x^2\,dx = F(3) - F(1) = 27 - 1 = 26$.

What if we'd picked, say, $G(x) = x^3 + 5$? Then we'd have $\int_1^3 3x^2\,dx = G(3) - G(1) = 32 - 6 = 26$ again.

Example 5.28. What is $\int_{\pi/4}^{3\pi/4} \cos(x)\,dx$?

We see that $\sin(x)$ is an antiderivative for $\cos(x)$. So we have $\int_{\pi/4}^{3\pi/4} \cos(x)\,dx = \sin(3\pi/4) - \sin(\pi/4) = \sqrt{2}/2 - \sqrt{2}/2 = 0$.

5.4.1 Indefinite Integrals

Because antiderivatives are so important, we want a notation for them that is less awkward than having to write the word "antiderivative" over and over. Because they are so closely tied to integrals, we use notation specifically designed to confuse you about what the integral sign means.

Definition 5.29. The *indefinite integral* of a function f, written $\int f(t)\,dt$, is any antiderivative of f. That is, $\int f(t)\,dt$ refers to any function $F(x)$ such that $F'(x) = f(x)$.

The *general form of the indefinite integral* is $\int f(x)\,dx = F(x) + C$. The constant represents the fact that there are many possible antiderivatives of f.

Very Important Note: Remember the difference between the definite and indefinite integrals. The definite integral $\int_a^b f(x)\,dx$ is a number. It is the area of some region under a graph. The indefinite integral $\int f(x)\,dx$ is a collection of functions, which are all antiderivatives of f and are all the same up to a constant. They are related by

$$\int_a^b f(x)\,dx = \int f(x)\,dx\Big|_a^b = F(b) - F(a).$$

In general the notation $|_a^b$ means "the value at b minus the value at a. We will use it a lot while doing integrals.

Example 5.30. We can write $\int x^5\,dx = \frac{1}{6}x^6 + C$, and $\int \sec^2(x)\,dx = \tan(x) + C$.

5.4.2 Antiderivatives, Net Change, and Linear Approximation

We can look at all of what we've done from another perspective, and connect it back to the work we did earlier on linear approximation.

Suppose we have a function F that we want to know about, but we only know about the derivative $F'(x)$. For instance, we may want to know the position of an object but only have measured the speed, or want to know the speed after measuring the acceleration. Or we want to figure out how much money we owe from a record of our annual deficits; we've seen a lot of examples of derivatives.

The example of deficit and debt makes this maybe easy to think of. Suppose you have a deficit of \$3000 one year, \$5000 the second year, and \$2000 the third year. At the end of three years, the debt has increased by \$10,000, which we get by adding the three deficits up.

This works exactly because we have a discrete set of payments, but if we don't have that we can still approximate it. Suppose that $F(t)$ gives the position of a particle at time t, and we know the velocity $F'(t)$. If we also know the starting position $F(0)$, we could estimate $F(4) \approx F(0) + F'(0)(4-0)$, but that might not be very good.

One way we could make this better is to do something like a quadratic approximation, or a Taylor series, but that gets messy. Another option is to do *multiple approximations*. Since the approximation gets worse the further x gets from a, we can try to bring it closer, and approximate in multiple steps.

Thus maybe we have

$$\begin{aligned} F(2) &\approx F(0) + F'(0)(2-0) \\ F(4) &\approx F(2) + F'(2)(4-2) \approx F(0) + F'(0)(2-0) + F'(2)(4-2). \end{aligned}$$

So if we take, say, $F'(t) = 10t$ and $F(0) = 0$, this would give us

$$\begin{aligned} F(2) &\approx 0 + 0(2-0) \\ F(4) &\approx 0 + 20(2) = 40 \end{aligned}$$

which is close-ish but not super close to the true answer of 80 (as we'll see soon).

What if we take more steps? We get

$$\begin{aligned} F(1) &\approx F(0) + F'(0)(1-0) && \approx 0 + 0(1-0) \\ F(2) &\approx F(1) + F'(1)(2-1) && \approx 0 + 10(2-1) = 10 \\ F(3) &\approx F(2) + F'(2)(3-2) && \approx 10 + 20(3-2) = 30 \\ F(4) &\approx F(3) + F'(3)(4-3) && \approx 30 + 30(4-3) = 60. \end{aligned}$$

But what is this last formula, really? It's

$$F(4) \approx F(0) + F'(0)(1-0) + F'(1)(2-1) + F'(2)(3-2) + F'(3)(4-3).$$

If we rearrange this a bit, we just get

$$F(4) - F(0) \approx F'(0)(1-0) + F'(1)(2-1) + F'(2)(3-2) + F'(3)(4-3)$$

and the right-hand side is a sum of terms that look like $F'(x_i)\Delta x_i$. So we have

$$F(4) - F(0) \approx \sum_{i=1}^{n} F'(x_i)\frac{4}{n}.$$

This is just a Riemann sum! And as we take the limit, we get an integral

$$F(4) - F(0) = \lim_{n\to\infty} \sum_{i=1}^{n} F'(x_i)\frac{4}{n} = \int_0^4 F'(x)\,dx.$$

Early on in the class, we saw that if you know the value of F and the derivative of F at 0, then you can use a linear approximation to estimate the value at any point. What we see now is that if you know the derivative of F everywhere, and the value at one point, you can find the value exactly, by taking an infinite collection of very small linear approximations.

Specifically, if you know the derivative, you can figure out the net change of F between any two values; so if you have one value, you can find any value.

Corollary 5.31 (Net Change Theorem)**.** *The integral of a rate of change is the total (net) change.*

$$\int_a^b F'(x)\,dx = F(b) - F(a).$$

Remark 5.32. Note that to find the value of $F(b)$ this way, we need to start by knowing $F(a)$ for some a. If we think of F as just being an antiderivative of F', the starting value is nailing down exactly the constant C.

Remark 5.33. This process of taking a large number of linear approximations is used in the real world a lot. If you have an integral that you *can't* find an exact formula for, this is very useful. It generalizes even more to solving differential equations, which are equations that specify F using a *formula* for $F'(x)$. They are more complicated than simple integrals, and you will see a little of them in calculus 2. But they are also the fundamental underpinning of most mathematical models, in the physical sciences and the social scienes.

5.4.3 Computing Integrals for the Practical Person

We've learned that computing integrals is reducible to finding antiderivatives. Now we're finally ready to practice actually computing integrals. In order to do this, we start by recalling a number of antiderivatives.

I'll list a few in these notes. There is an extensive card listing many of htese rules on page 6 of the reference in the back of Stewart, and a shorter table on page 331 in section 4.4.

- $\int f(x) + g(x)\, dx = \int f(x)\, dx + \int g(x)\, dx$.
- $\int cf(x)\, dx = c \int f(x)\, dx$.
- $\int x^n\, dx = \frac{x^n}{n+1} + C$ if $n \neq -1$.
- $\int \sin(x)\, dx = -\cos(x) + C$.
- $\int \cos(x)\, dx = \sin(x) + C$.
- $\int \sec^2(x)\, dx = \tan(x) + C$.
- $\int \csc^2(x)\, dx = -\cot(x) + C$.
- $\int \sec(x)\tan(x)\, dx = \sec(x) + C$.
- $\int \csc(x)\cot(x)\, dx = -\csc(x) + C$.

Example 5.34.

- What is $\int_1^4 x^2\, dx$? We know that $\int x^2\, dx = \frac{1}{3}x^3 + C$, so $\int_1^4 x^2\, dx = \frac{1}{3}x^3|_1^4 = \frac{1}{3}(64 - 1) = 21$. Note the Cs cancel each other out so it doesn't matter what they are.

- What is $\int_2^3 x + x^3\,dx$? We can work out that $\int x + x^3 = \frac{x^2}{2} + \frac{x^4}{4}$, so

$$\int_2^3 x + x^3\,dx = \frac{x^2}{2} + \frac{x^4}{4}\Big|_2^3 = \frac{9}{2} + \frac{81}{4} - \frac{4}{2} - \frac{16}{4} = \frac{99}{4} - 6 = \frac{75}{4}.$$

- Calculuate $\int_{-1}^2 |x|\,dx$. We don't really have an antiderivative of $|x|$, so the easiest way to approach this is probably to break it up into two distinct integrals.

  If $x \geq 0$ then $|x| = x$, so we have $\int_0^2 |x|\,dx = \int_0^2 x\,dx = \frac{x^2}{2}|_0^2 = 2 - 0 = 2$.

  If $x \leq 0$ then $|x| = -x$ and we have $\int_{-1}^0 |x|\,dx = \int_{-1}^0 -x\,dx = \frac{-x^2}{2}|_{-1}^0 = 0 - \frac{-1}{2} = \frac{1}{2}$.

  Thus $\int_{-1}^2 |x|\,dx = \int_{-1}^0 |x|\,dx + \int_0^2 |x|\,dx = \frac{1}{2} + 2 = \frac{5}{2}$.

- Calculate $\int_0^{\pi/4} \sec(x)\tan(x)\,dx$. At first blush this looks hard, until you remember that $\sec'(x) = \sec(x)\tan(x)$. So we have

$$\int_0^{\pi/4} \sec(x)\tan(x)\,dx = \sec(x)|_0^{\pi/4} = \sec(\pi/4) - \sec(0) = \sqrt{2} - 1.$$

- What if we want $\int_0^{\pi} \sec(x)\tan(x)$? This is a much bigger problem, because $\sec(x)\tan(x)$ is not continuous on $[0, \pi]$. We actually won't be able to do that one without new ideas that we won't develop in this course.

Leading question: can you do $\int 3x^2\sqrt{9 + x^3}\,dx$?

5.5 Integration by Substitution

The Fundamental Theorem of Calculus is a powerful tool for computing integrals. And with functions that are obviously the derivatives of some other function, like x^2 or $\cos(x)$, it's very easy to apply. With more complicated functions it takes a bit more work.

Example 5.35. What is $\int 3x^2\sqrt{9 + x^3}\,dx$?

There are two ways to approach this problem. The first is to notice that you almost have an antiderivative to $\sqrt{9 + x^3}$, because $(9 + x^3)^{3/2}$ has $\frac{3}{2}(9 + x^3)^{1/2} \cdot 3x^2$ as its derivative. The extra $3x^2$ from the chain rule precisely matches up with the extra $3x^2$ from the problem, so we just have to correct for the constant, and we have that $\int 3x^2\sqrt{9 + x^3} = \frac{2}{3}(9 + x^3)^{3/2} + C$.

If that made sense, great. Whenever you can "just see" the antiderivative, you can go for it; the fact that you can check your work by taking a derivative means that you are safe. But for the cases where you can't just see the answer, we'd like to be a little more systematic in our approach.

We know how to take the antiderivative of $\sqrt{x}$. So let's try using a new variable, which we traditionally call u. We write $u = 9 + x^3$ so the thing under the radical is a u. We also notice that $\frac{du}{dx} = 3x^2$; by "abuse of notation" (by which I mean we won't justify it, but just assume it works) we write $du = 3x^2 dx$. Since our original integral was $\int \sqrt{9 + x^3} \cdot 3x^2 dx$, we can rewrite this as $\int \sqrt{u}\, du$, or just $\int u^{1/2}\, du$.

From our integral table, we know that $\int u^{1/2}\, du = \frac{2}{3} u^{3/2} + C$. Now we can replace the u with $9 + x^3$ to get $\int 3x^2 \sqrt{9 + x^3}\, dx = \frac{2}{3}(9 + x^3)^{3/2} + C$.

We can formalize this into a rule:

Proposition 5.36 (The Substitution Rule for Indefinite Integrals)**.** *If $u = g(x)$ is differentiable, and $f(x)$ is continuous on the range of g, then*

$$\int f(g(x))g'(x)\, dx = \int f(u) du.$$

Proof. This follows from the chain rule. Let F be an antiderivative of f; then $(F(g(x)))' = F'(g(x)) \cdot g'(x) = f(g(x)g'(x)$. Thus $F(g(x))$ is an antiderivative of $f(g(x))g'(x)$.

I'd like to give you geometric intuition here, but it's a bit hard to communicate. In essence we're changing to a new coordinate system where the integral is easy, but it's hard to make that observation *useful* until you get to multivariable calculus. For right now, yo ushould probably think of this as a way of keeping track of algebraic manipulations. ☐

How do we use this? Basically, when we see a complicated integral, there are a couple things we can look for. The first is to check whether one part is a derivative of another part, in a way that could reflect a chain rule. The other is to find the most complicated chunk of the expression and replace it with a u, and see how much of our problem that solves.

Choosing the right variable to substitute is a bit of an art; I can't possibly give you a complete set of rules, but I can give you a lot of examples to model off of.

Example 5.37.

- Consider $\int x^2 \sin(x^3 + 3)\, dx$. We can take $u = x^3 + 3$, and then $du = 3x^2\, dx$ so $dx = \frac{du}{3x^2}$. So this becomes $\int \sin(u)/3\, du = -\cos(u)/3 + C = \cos(x^3+3)/3 + C$.
- Consider $\int \sqrt{5x + 2}\, dx$. It makes sense to take $u = 5x + 2$, so $du = 5dx$. Then $\int \sqrt{u}/5\, du = \frac{2}{15} u^{3/2} + C = \frac{2}{15}(5x + 2)^{3/2} + C$.

  Alternatively, we could take $u = \sqrt{5x + 2}$. Then $du = \frac{5}{2\sqrt{5x+2}} dx$ and we get $dx = \frac{2}{5}\sqrt{5x + 2} = \frac{2}{5} u$. So we have $\int \frac{2}{5} u^2\, du = \frac{2}{15} u^3 + C = \frac{2}{15}(5x + 2)^{3/2} + C$.

- For a more complex example, we can look at $\int \sqrt{1+x^2}x^5\,dx$. This doesn't look like it will happen automatically, and indeed it doesn't. But we can still get rid of the complicated bit by taking $u = 1 + x^2$, so $du = 2x\,dx$ or $dx = du/2x$.

  This gives us $\int \sqrt{u}x^4\frac{1}{2}du$, but what do we do with the other x^4 term? Well, if $u = 1+x^2$ that means that $x^2 = u - 1$, so our integral is

$$\begin{aligned}\int \frac{1}{2}\sqrt{u}(u-1)^2\,du &= \int \frac{1}{2}\left(u^{5/2} - 2u^{3/2} + u^{1/2}\right)\,du \\ &= \frac{1}{7}u^{7/2} - \frac{2}{5}u^{5/2} + \frac{1}{3}u^{3/2} + C \\ &= \frac{1}{7}(1+x^2)^{7/2} - \frac{2}{5}(1+x^2)^{5/2} + \frac{1}{3}(1+x^2)^{3/2} + C.\end{aligned}$$

5.5.1 Substitution and Definite Integrals

The above talked about indefinite integrals. When we have a definite integral, we can be more specific. We can use substitution in two ways: one is to do what we did above, where we substitute in a u, then integrate, then switch the us back to xs. But we can avoid switching back at all by changing the limits of integration.

Proposition 5.38 (The Substitution Rule for Definite Integrals). *If g' is continuous on $[a,b]$, and f is continuous on the range of $g(x)$, then*

$$\int_a^b f(g(x)) \cdot g'(x)\,dx = \int_{g(a)}^{g(b)} f(u)\,du.$$

Proof. If F is an antiderivative of f, then the left side is clearly $F(g(b)) - F(g(a))$. But the antiderivative of $f(g(x))g'(x)$ is $F(g(x)$, so the left side is also $F(g(b)) - F(g(a))$. ☐

Example 5.39.

- Find $\int_0^2 \frac{x}{\sqrt{1+2x^2}}\,dx$. We take $u = g(x) = 1 + 2x^2$ so that $du = 4dx$, so $dx = du/4$, and $g(0) = 1, g(2) = 9$. We have

$$\frac{1}{4}\int_1^9 u^{-1/2}\,du = \frac{1}{4}2u^{1/2}|_1^9 = \frac{1}{2}(3-1) = 1.$$

- Find $\int_1^3 \frac{dx}{(1-2x)^2}$. Set $u = g(x) = 1 - 2x$, then $du = -2dx$ and $g(1) = -1, g(3) = -5$. So

$$\int_1^3 \frac{dx}{(1-2x)^2} = \int_{-1}^{-5} \frac{-du}{2u^2} = \left.\frac{1}{2u}\right|_{-1}^{-5} = \frac{1}{-10} - \frac{1}{-2} = \frac{2}{5}.$$

A nice bonus application of this is to look at symmetric functions. Since even and odd functions have nice geometric symmetries, integrals, which are about the area under the curve, should also have nice properties.

Corollary 5.40 (Integrals of Symmetric Functions)**.** *Suppose f is a continuous function on $[-a, a]$. Then*

- *If f is even, then $\int_{-a}^{a} f(t)\,dt = 2\int_0^a f(t)\,dt$.*
- *If f is odd, then $\int_{-a}^{a} f(t)\,dt = 0$.*

Proof. Intuitively this should be plausible; even functions look the same on either side of the y-axis, and so you should get the same area on both sides, while odd functions are the same but upside down, so you should get the opposite area. (Try sketching a picture of sin and cos to see this).

For either integral, notice that $\int_{-a}^{a} f(t)\,dt = \int_{-a}^{0} f(t)\,dt + \int_0^a f(t)\,dt$. Consider the first integral, and use the substitution $u = g(t) = -t$, and thus $-du = -dt$. Then $\int_{-a}^{0} f(t)\,dt = \int_a^0 f(-t)(-dt) = \int_0^a f(-t)\,dt$.

If f is even then $f(-t) = f(t)$, so $\int_{-a}^{0} f(t)\,dt = \int_0^a f(t)\,dt$. If f is odd then $f(-t) = -f(t)$ and thus $\int_{-a}^{0} f(t)\,dt = -\int_0^a f(t)\,dt$. □

Example 5.41.

- $\int_{-3}^{3} x^5 - x^3\,dx = 0$.
- $\int_{-2}^{2} x^6 + 1\,dx = 2\int_0^2 x^6 + 1\,dx = 2(x^7/7 + x)|_0^2 = 2(128/7 + 2) = \frac{284}{7}$.

5.6 A Brief Note on How to Cheat

We've now learned how to compute basic integrals. There are a lot of integrals we haven't yet learned to compute; a prominent example is $\int \frac{1}{x}\,dx$, but there are many. In calculus 2 you will develop many other techniques of integration which allow us to integrate more difficult functions. However, as good mathematicians we're also fundamentally lazy and would prefer to avoid work when we can manage it. There are two common solutions here.

First, the back of your textbook has an extensive integral table, and even more extensive tables can be found online. It often requires minor massaging to get your integral into the form of the table, but for complex integrals the table will be much easier than figuring things out from scratch. (For instance, the table incorporates the results of trig subsitution without making you work through it explicitly).

Second, computers are very good at doing integrals. Wolfram Alpha can often integrate a function for you, as can Mathematica and other computer tools. It's dangerous to become overly reliant on these tools—it's easy to make a mistake if you don't understand what's going on, and sometimes the computer will return the answer in a less useful form. They are very good for automated computations and checking your work, however.

A final cautionary note: there are some functions that don't have a nice closed-form antiderivative. Famously, there's no way to write $\int e^{x^2}\,dx$ in terms of "elementary functions." That doesn't mean there is no antiderivative; the obvious one is $\int_0^x e^{t^2}\,dt$. But while correct, that answer isn't terribly enlightening.

We can't easily compute these definite integrals exactly, but we can approximate them using various approximation techniques (among other things, just computing a finite Riemann sum). We can also use the concept of "infinite series" to handle this sort of situation; those techniques occur towards the end of Calculus 2.

6 Applications of Integrals

6.1 The Average Value of a Function

This is a convenient time to address the concept of "average value." If we have some finite collection of numbers, the average is what we get when we add them up, and divide by the number of numbers:

$$\frac{1}{n}\sum_{i=1}^{n} a_i.$$

A function gives us infinitely many numbers; but integration is in some sense a sensible way to add infinitely many numbers up, and so hopefully to average them.

In particular, if we sample the function at n evenly spaced points, our average is

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i^*) = \frac{1}{b-a}\sum_{i=1}^{n} \frac{b-a}{n} f(x_i^*)$$

which you should recognize as a Riemann sum (times $\frac{1}{b-a}$). If we take the limit—which represents taking the average value after "infinitely many" sample points—we get the following definition:

Definition 6.1. The *average value of a function* f over an interval $[a,b]$ is

$$f_{ave} = \frac{1}{b-a}\int_a^b f(t)dt.$$

Example 6.2. What is the average value of $f(x) = x^2$ on $[0,1]$? We have

$$f_{ave} = \frac{1}{1}\int_0^1 x^2\,dx = \frac{1}{3}.$$

The biggest value is 1, the smallest is 0, and the one in the middle is $\frac{1}{4}$, but the "average" value is $\frac{1}{3}$.

If I have a finite set of numbers and take the average, my average might not be anywhere in the set; for instance, if I roll a six-sided die, the average output will be 3.5, which isn't on the die at all. When I average continuous quantities, however, this can't happen.

Theorem 6.3 (Mean Value Theorem for Integrals)**.** *If f is continuous on $[a,b]$, then there is a number c in $[a,b]$ such that*

$$f(c) = f_{ave} = \frac{1}{b-a}\int_a^b f(t)\,dt.$$

In other words,

$$\int_a^b f(t)\,dt = f(c)(b-a).$$

Proof. This statement, as well as its name, might look familiar. In fact this is just the mean value theorem from differential calculus repackaged. Let $F(x) = \int_a^x f(t)\,dt$. Then F is continuous on $[a, b]$ and differentiable on (a, b), and so by the Mean Value Theorem there is some c such that $F(b) - F(a) = F'(c)(b - a)$.

But by the Fundamental Theorem of Calculus, $F'(c) = f(c)$. And it's easy to see that $F(b) = \int_a^b f(t)\,dt$, and $F(a) = \int_a^a f(t)\,dt = 0$. So we have

$$\int_a^b f(t)\,dt - 0 = f(c)(b - a).$$

□

Remark 6.4. Geometrically, this essentially tells us that there is some rectangle with the same area as the region under the graph of f. In particular, we can take a rectangle with width $b - a$, whose top edge intersects the graph of our function *somewhere*, and whose area is the same as the area of the region under the curve.

6.2 Finding Areas

Recall that we originally constructe the integral to find the area of some shape, in particular of shapes that lie under the graph of some function. We can use the same tools to find the area of a region that is not, properly speaking, the graph of one function.

The simplest (well, second-simplest) case is the case where we want the area of a region that lies in between the graph of two functions. We can approximate area by drawing, as before, a great many skinny rectangles which are approximately the right height to cover our region. If our region lies in between two functions f and g, the combined area of our rectangles is

$$\sum_{i=1}^{n}(x_i - x_{i-1})(f(x_i^*) - g(x_i^*))$$

and as the number of rectangles increases this approximation gets increasingly good. We say the area of the region is

$$A = \lim_{n\to+\infty} \sum_{i=1}^{n}(x_i - x_{i-1})(f(x_i^*) - g(x_i^*)).$$

You may recognize this formula as the integral of the function $f - g$; indeed, if we have a region with x coordinates varying from a to b and y coordinates varying from $g(x)$ to $f(x)$, then its area is $\int_a^b (f(x) - g(x))\,x$.

Remark 6.5. Remember that actual areas are always positive! The integral by itself computes the "signed area"; if you want an actual area you must be careful to make sure you're integrating the correct function.

Example 6.6. Let's start with a trivial example: what's the area of a rectangle with base 3 and height 4? Well, this is $\int_0^4 3\,dx = 3x|_0^4 = 12$, as it should be.

Example 6.7. What is the area of the region between $y = x^3$ and $y = 1/x^2$ between $x = 2$ and $x = 4$?

We have

$$\int_2^4 x^3 - (1/x^2)\,dx = \left(\frac{x^4}{4} + \frac{1}{x}\right)\Big|_2^4 = (64 + 1/4) - (4 + 1/2) = 60 - 1/4 = 239/4.$$

Sometimes (usually!) we need to have a visual idea of what our region looks like before we can set up an appropriate integral.

Example 6.8. What is the area of the region bounded by $y = x$ an $y = x^2$?

After we draw a picture, we see that these two graphs enclose a region between $x = 0$ and $x = 1$, and that in that region, $x \geq x^2$. So we compute the integral

$$\int_0^1 x - x^2\,dx = \left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_0^1 = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

Example 6.9. Compute the total area of the "valley" between two peaks of the sine function.

We see that this area is the area of the region between $y = 1$ and $y = \sin x$ between $\pi/2$ and $5\pi/2$. (There are other ways to set this up, but this way works). So we compute

$$\int_{\pi/2}^{5\pi/2} 1 - \sin x\,dx = x + \cos(x)|_{\pi/2}^{5\pi/2} = (5\pi/2 + 0) - (\pi/2 + 0) = 2\pi.$$

Sometimes you have to break your region up into separate pieces/integrals

Example 6.10. What is the area of the region bounded by $y = x^2, y = 2 - x$, and $y = 0$?

We sketch the region and see that we get a sort of collapsed triangle. We compute

$$\begin{aligned} A &= \int_0^1 x^2\,dx + \int_1^2 (2 - x)\,dx = \frac{x^3}{3}|_0^1 + \left(2x - \frac{x^2}{2}\right)\Big|_1^2 \\ &= \frac{1}{3} - 0 + (4 - 2) - (2 - 1/2) = \frac{5}{6}. \end{aligned}$$

We can also do the same problem another way. Notice that we might as well write $x = \sqrt{y}, x = 2 - y$. So we can just as well integrate with respect to y—that is, draw our rectangles stretching horizontally instead of vertically. We have

$$A = \int_0^1 (2-y) - \sqrt{y}\, dy = \left(2y - \frac{y^2}{2} - \frac{2}{3}y^{3/2}\right)\Big|_0^1 = \left(2 - \frac{1}{2} - \frac{2}{3}\right) - 0 = \frac{5}{6}.$$

As expected, we get the same answer.

Remark 6.11. In general, if you have straight line or point boundaries on opposite sides, you should integrate between them. In general, if you can write something as the difference of two functions one way and not the other way, you should do that.

Example 6.12. What is the area of the region between $y^2 = x + 3$ and $y = x - 3$?

These curves intercept when $y^2 = y + 6$, which happens when $y = 3$ or $y = -2$, and thus at $(6, 3)$ and $(1, -2)$. It's more natural to integrate with respect to y, so we write

$$\begin{aligned} A &= \int_{-2}^3 (y+3) - (y^2 - 3)\, dy = \int_{-2}^3 6 + y - y^2\, dy \\ &= \left(6y + \frac{y^2}{2} - \frac{y^3}{3}\right)\Big|_{-2}^3 = \left(18 + \frac{9}{2} - 9\right) - \left(-12 + 2 + \frac{8}{3}\right) = \frac{27}{2} + 10 - \frac{8}{3} = \frac{125}{6} \end{aligned}$$

Example 6.13. What is the area of the region bounded by $y = x^2 + 1, y = 17 - x^2$, and $y = 1$?

We first draw the region, and see a sort of sideways triangle with a base at $x = 1$ and a point at $(\sqrt{8}, 9)$, with x varying from 1 to 17. We have two options: integrate with respect to x, or with respect to y by writing $x = \sqrt{y-1}$ and $x = \sqrt{17-y}$. The second doesn't involves breaking our region into two integrals, and gives us

$$A = \int_2^9 \sqrt{17-y} - 1\, dy + \int_9^{16} \sqrt{17-y} - 1\, dy,$$

which is doable but pretty ugly.

Instead, if we integrate with respect to x, we get

$$\begin{aligned} A &= \int_1^{\sqrt{8}} (17 - x^2) - (x^2 - 1)\, dx = \int_1^{\sqrt{8}} 18 - 2x^2\, dx \\ &= 18x - \frac{2}{3}x^3\Big|_1^{\sqrt{8}} = 36\sqrt{2} - 32\sqrt{2}/3 - 18 + 2/3 = \frac{76\sqrt{2} - 52}{3}. \end{aligned}$$

6.3 Applications to Physics

Now we should discuss some physical processes that are well-described by integration—which is just a fancy way of saying that integrals let us solve these problems.

6.3.1 Work

In physics, *force* is the product of mass and acceleration; intuitively, force is what causes a mass to accelerate, and the more acceleration/the more massive the object, the more force is required. This is often written $F = m \cdot a$, but in our context it is better to say that the position of an object is given by the function $s(t)$, and then $F = m \cdot \frac{d^2s}{dt^2}$, since acceleration is the second derivative of position.

Remark 6.14. In the SI system, mass is measured in kilograms, and force is measured in newtons, where $N = kg \cdot m/s^2$. In the Imperial system most Americans use, the pound is a unit of force; the unit of mass is the *slug*, and one pound is one slug-foot per second squared. I bring this up primarily because the name "slug" is funny.

Intuitively, moving things aroudn takes work, and moving them faster takes more work. Formally, we say that *work* is force times distance: the amount of force applied to an object, times the distance the object is moved. The SI unit for work is the Newton-meter or *joule*, which is $J = kg \cdot m^2/s^2$. The imperial unit for work is the foot-pound, which is about 1.36 joules.

If you lift a 2 kg object a meter, then you have to exert $2 \cdot 9.8$ newtons of force (since acceleration due to gravity is $9.8 m/s^2$, and thus do 19.6 joules of work. If a 20 pound weight is lifted five feet, than 100 foot-pounds of work are done.

When force is constant, work is easy to calculuate–just multiply the force by the distance. Things become more interesting when the force varies. As usual, we can approximate by chopping the movement up into lots of little pieces, assuming the force is constant on each small piece, and adding them up. That is, if the force at position x is $F(x)$, then when an object moves from a to b the work done is approximately

$$W \approx \sum_{i=1}^{n} F(x_i) \frac{b-a}{n}.$$

This is a Riemann sum, so taking the limit gives an integral: the total work done is

$$\int_a^b F(x)\,dx.$$

Remark 6.15. Unlike most of the geometric integrals we've been doing for the past few weeks, work can be a negative number; this just indicates that the force is in the opposite direction of the motion.

Example 6.16. A particle is controlled by a force field such that the force on it is $x^3 + x$ pounds when it is x feet away from the origin. How much work does it take to move the particle from $x = 2$ to $x = 4$?

$$W = \int_2^4 x^3 + x\,dx = \frac{x^4}{4} + \frac{x^2}{2}|_2^4 = 64 + 8 - 4 - 2 = 66.$$

Example 6.17. A physical law called Hooke's Law says that the force exerted by a string stretched x units beyond its natural length is kx, where k is the "spring constant" and depends on the particular spring.

Suppose a spring is naturally 20 cm and it takes 50 N to stretch it to 30 cm. How much work is needed to stretch the spring from 30cm to 35cm?

We have $50 = k \cdot .1$ and so $k = 500$. Thus the force when the spring is stretched x meters beyond its normal length is kx, and the work done is

$$W = \int_{.1}^{.15} 500x\,dx = 250x^2|_{.1}^{.15} = 3.125J.$$

Example 6.18. A 50 meter cable has a mass of 50kg and hangs from the top of a cliff. How much work does it take to raise the cable up the cliff?

The thing that makes this difficult is that the mass of the remaining rope depends on how much mass we've lifted already. Conceptually, you can think about having to lift the first meter of rope one meter, and the second meter of rope two meters, etc. Each meter of rope masses 1 kg, so this would give us a Riemann sum

$$W \approx \sum_{i=1}^{50} 1 \cdot 9.8 \cdot i$$

Or more generally

$$W \approx \sum_{i=1}^{n} \Delta x \cdot 9.8 \cdot x_i.$$

Taking the limit gives the integral

$$W = \int_0^{50} 9.8x\,dx = 4.9x^2|_0^{50} = 2500 \cdot 4.9 = 12250J.$$

Example 6.19. A tank of water is shaped like an upside-down pyramid. (No, I don't know why people keep building tanks shaped like upside-down pyramids). The pyramid has a base side length of 4m and a height of 12m, and it is filled with water to a depth of 8m. How much work will it take to pump the water out of the top of the tank? (water has a density of 1000kg / m^3).

Again, to figure out our integral we may want to set up the Riemann sum, or at least fake set it up. Let 0 be the point of the pyramid and 12 be the base (at the top). The volume of a small cross-sectional volume is $A(h)\Delta h$, thus the mass is $1000A(h)\delta h$ and the force is $1000A(h)\Delta h \cdot 9.8$. The distance we have to pump the water is $12 - h$, so the total work on each cross-section is $(12 - h)9800A\Delta h$ Newtons.

Now we just have to work out area in terms of height. Using a similar triangles argument, we see that $\frac{s(h)}{h} = \frac{4}{12}$ and thus $s(h) = h/3$, and $A(h) = h^2/9$. We integrate from 0 to 8 becasue we're integrating over the height that contains water. Then we have

$$\int_0^8 (12 - h)9800 \cdot h^2/9 \cdot dy = \frac{9800}{9}\left(4h^3 - \frac{h^4}{4}\right)|_0^8 = \frac{9800}{9}(2048 - 1024 - 0) = \frac{10,035,200}{9}J.$$

6.3.2 Hydrostatic Pressure

Another problem we can handle easily with these tools is the idea of water (or fluid) pressure. If you imagine a flat surface submerged in some fluid with density ρ to a depth of d meters, then the weight of the fluid over it is $A\rho dg$ where A is the area of the surface (and thus $Ad\rho$ is the mass of the fluid) and $g = 9.8$ is acceleration due to gravity. We define the pressure to the be the force divided by the area, and thus $P = \frac{F}{A} = \rho dg$.

(In SI units we measure this in Newtons per square meter, otherwise known as Pascals. In Imperial units there are a number of different units used, including "inches of mercury.")

Fact 6.20. *If an object is submerged in a fluid to a given depth, the pressure exerted by the fluid is the same in all directions.*

This means that fluid pressure is effectively a function of height/depth and nothing else. If the pressure is varying and we want to find the total force acting on a surface, we can effectively add up the pressure on each little patch of a surface to find the total force acting on it.

Example 6.21. A 3 by 3 meter square is submerged in water until it is just covered, edge-first. What is the total force the water exerts on the square?

We want to chop the square into strips that are all at roughly the same depth. If we slice the square into three horizontal strips, then the ith strip is roughly at depth i meters and has width 3, and thus has roughly the force $3 \cdot 1 \cdot i \cdot \rho \cdot g$. Adding up the force on all thirty strips gives

$$F \approx \sum_{i=1}^{3} 3 \cdot 1 \cdot i \cdot \rho \cdot g = \sum_{i=1}^{3} 3 \cdot 1000 \cdot 9.8 \cdot h\Delta h$$

In the limit, we get the following integral:

$$\int_0^3 3 \cdot \rho \cdot g \cdot h \, dh = \int_0^3 3000 \cdot 9.8 \cdot h \, dh = 29400(h^2/2)|_0^3 = 29400 \cdot \frac{9}{2} = 132,300.$$

Example 6.22. A cylindrical drum is lying on its side underwater. The drum has radius of 5 feet and is submerged in 20 feet of water. What is the force exerted on one circular face of the drum?

Let's set 0 to be the center of the circle, so that the equation for the circle is $x^2+y^2 = 25$. Then the width of the object at height y is $2\sqrt{25-y^2}$. The depth at height y is $15 - y$ (which ranges from 10 to 20), and the pressure due to water is 62.5 * depth. So we get the integral

$$F = \int_{-5}^{5} 62.5(15-y)2\sqrt{25-y^2} \, dy = 125 \int_{-5}^{5} 15\sqrt{25-y^2} \, dy - 125 \int_{-5}^{5} y\sqrt{25-y^2} \, dy.$$

The second integral is 0 because $y\sqrt{25-y^2}$ is an odd function. The first integral can be done by setting $y = 5\sin\theta$, but we can also observe that it is the integral of a semicircle of radius 5 and thus is equal to 12.5π. So we have

$$F = 125 \cdot 15 \cdot 12.5\pi = 23437.5 lb.$$

6.3.3 Center of Mass

The center of mass of a two dimensional object is, conceptually, the point it can balance on. It is in some sense the "average" location the region occurs.

If the mass of an object occurs in finitely many points, then the center of mass is the weighted average of their locations, where the weighting is by the mass. So if we have particles of mass m_1, m_2, m_3 at points $(x_1, y_1), (x_2, y_2), (x_3, y_3)$, with total mass m, then the x-coordinate of the center of mass of the system is

$$\overline{x} = \frac{1}{m} \sum_{i=1}^{3} m_i x_i = m_1x_1 + m_2x_2 + m_3x_3$$

and the y-coordinate is

$$\overline{y} = \frac{1}{m} \sum_{i=1}^{3} m_i y_i = m_1y_1 + m_2y_2 + m_3y_3$$

As a vocabulary note, we say that each of these $m_i x_i$ or $m_i y_i$ is a *moment* of the mass, and the sum $\sum_{i=1}^n m_i x_i$ is the *moment of the system* about the origin in the x-axis.

Example 6.23. We have particles of masses $1, 4, 5$ at the poitns $(0,0), (3,2), (4,5)$. Then for the center of mass we have

$$\overline{x} = \frac{1}{10}(1 \cdot 0 + 4 \cdot 3 + 5 \cdot 4) = \frac{32}{10}$$
$$\overline{y} = \frac{1}{10}(1 \cdot 0 + 4 \cdot 2 + 5 \cdot 5) = \frac{33}{10}.$$

We extend this study to calculus. Suppose we have a plate of "uniform density" (i.e. it's all the same material, so bits with the same area wil have the same mass/weight). For concreteness, say the region is given by $a \leq x \leq b$ and $g(x) \leq y \leq f(x)$. We'd like to find the center of mass, the point the plate balances perfectly. We can think about how to make it balance in each direction, so we can find the x-coordinate and the y-coordinate separately.

To find the x coordinate of the center of mass, we add up the mass of each vertical strip, weighted by its x-coordinate, just as we did before. The vertical strip has width dx and height $f(x) - g(x)$. Thus each strip has area $(f(x) - g(x))dx$, and we can assume the density is 1 so that it has mass $(f(x) - g(x))dx$ as well. Thus the moments of mass are $x(f(x) - g(x))dx$, and the x-coordinate of the center of mass is

$$\overline{x} = \frac{1}{A}\int_a^b x(f(x) - g(x))dx.$$

To find the y-coordinate, we could do the same thing with respect to y. But if our region is described in terms of a function of x, then this might be awkward. But we can still add up the moment of each *vertical* strip. The strip at x still has area $(f(x) - g(x))dx$, and the "average" position of the strip is the middle of the strip, which is at $\frac{1}{2}(f(x) + g(x))$. So the moment is $\frac{1}{2}(f(x) - g(x))^2\, dx$ and the y-coordinate is

$$\overline{y} = \frac{1}{A}\int_a^b \frac{1}{2}(f(x)^2 - g(x)^2)\, dx.$$

Example 6.24. Find the center of mass of the region bounded by $y = x^2$ and $y = \sqrt{x}$.

The area is

$$A = \int_0^1 \sqrt{x} - x^2\, dx = \frac{2}{3}x^{3/2} - \frac{x^3}{3}|_0^1 = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}.$$

Then we have

$$\begin{aligned}
\overline{x} &= 3\int_0^1 x(\sqrt{x}-x^2)\,dx = 3\left(\frac{2}{5}x^{5/2}-\frac{x^4}{4}\right)\Big|_0^1 = 3\left(\frac{2}{5}-\frac{1}{4}\right) = \frac{9}{20}.\\
\overline{y} &= 3\int_0^1 \frac{1}{2}(\sqrt{x}^2-(x^2)^2)\,dx = \frac{3}{2}\int_0^1 \left(x-x^4\right)\,dx = \frac{3}{2}\left(\frac{x^2}{2}-\frac{x^5}{5}\right)\Big|_0^1\\
&= \frac{3}{2}\left(\frac{1}{2}-\frac{1}{5}\right) = \frac{9}{20}.
\end{aligned}$$

Example 6.25. Find the center of mass of the semicircle bounded by $y=\sqrt{r^2-x^2}$ and $y=0$ between $x=-r$ and $x=r$.

The area is half the area of a circle, and thus $\frac{1}{2}\pi r^2$. Then we have

$$\begin{aligned}
\overline{x} &= \frac{2}{\pi r^2}\int_{-r}^r x\sqrt{r^2-x^2} = 0 \text{ since } x\sqrt{r^2-x^2} \text{ is odd.}\\
\overline{y} &= \frac{2}{\pi r^2}\int_{-r}^r \frac{1}{2}(\sqrt{r^2-x^2})^2\,dx = \frac{1}{\pi r^2}\int_{-r}^r r^2-x^2\,dx\\
&= \frac{1}{\pi r^2}\left(r^2x-\frac{x^3}{3}\right)\Big|_{-r}^r = \frac{1}{\pi r^2}\left(r^3-\frac{r^3}{3}-\left(-r^3-\frac{r^3}{3}\right)\right) = \frac{1}{\pi r^2}\cdot\frac{4}{3}r^3 = \frac{4r}{3\pi} \approx .42.
\end{aligned}$$

Thus the center of mass is at about $(0,.42)$. The fact that the x coordinate should be 0 is geometrically obvious; the y coordinate is less so.

6.4 Finding Volumes by Cross-Sections

Area is fundamentally length times width, and we computed areas by integrating the length against the width–by which I mean, we wrote the length at a point as a function of the width at that point, and took the integral across the whole width.

Volume is area times height. (Or area times length, depending on your perspective). We will compute volume by finding the area of a cross-section and integrating along the entire length of our shape. Geometrically, the Riemann sum corresponds to slicing our shape into many thin cylinders and adding their areas up.

Remark 6.26. In our terminology, a “cylinder” is any solid that has a flat base and an identical flat top, connected by straight sides at right angles. A traditional circular cylinder qualifies, but so does a rectangular box, and so do stranger shapes.

Definition 6.27. If S is a solid, we say the *cross-sectional area* at a point x is the area of the intersection of our solid with the plane which passes through x and is perpendicular to the x-axis (and thus parallel to the yz plane).

If S is a solid lying between $x = a$ and $x = b$, and $A(x)$ is a function giving the cross-sectional area at x, then we say the *volume* V of S is

$$V = \lim_{\max \Delta x_i \to 0} \sum_{i=1}^{n} A(x_i^*)\Delta x_i = \int_a^b A(x)\, dx.$$

Example 6.28. What is the volume of a cone with height 2 and base radius 4?

We draw a picture. By a similar triangles argument, we see that when we are x distance from the point, the radius is $2x$ and thus the area of the cross-section is $4\pi x^2$. Thus the volume is

$$\int_0^2 4\pi x^2\, dx = \frac{4\pi x^3}{3}|_0^2 = \frac{32}{3}\pi.$$

This matches the formula for the volume of a cone, which is $\frac{1}{3}\pi r^2 h$.

In fact, we can also rederive that formula. If a cone has height h and base radius b, then the radius at x distance from the height is $x\frac{b}{h}$ and the area is $\pi x^2 b^2/h^2$. So the volume of the cylinder is

$$\int_0^h \pi x^2 b^2/h^2\, dx = \pi b^2/h^2 \frac{x^3}{3}|_0^h = \frac{b^2 h\pi}{3}.$$

Example 6.29. What is the volume of a solid with a circular base of radius one, where each cross-section is an equilateral triangle?

Make the circle $x^2 + y^2 = 1$. Then the width of the base of the cross-section at x is $2\sqrt{1-x^2}$. Since $\sin 60° = \sqrt{3}2$, we know the height of each triangle is $\sqrt{3}b/2$, and thus the area of the triangle is $\sqrt{3}(1-x^2)$. Thus the volume is

$$\int_{-1}^1 \sqrt{3}(1-x^2)\, dx = \sqrt{3}x - \frac{\sqrt{3}x^3}{3}|_{-1}^1 = \left(\sqrt{3} - \frac{\sqrt{3}}{3}\right) - \left(-\sqrt{3} - \frac{-\sqrt{3}}{3}\right) = \frac{4\sqrt{3}}{3}.$$

These problems are sometimes known as volumes of "solids of rotation," because this technique is particularly good at solving problems like the following:

Example 6.30. What is the volume of the solid obtained by rotating the region bounded by $y = x^2, x = 5, y = 0$ about the x-axis?

We draw a picture, and see that the region has height x^2 at a point x, and thus the solid has a cross-section which is a circle of radius x^2, and thus an area of $\pi(x^2)^2$. It's clear that x varies from 0 to 5. So

$$V = \int_0^5 \pi x^4\, dx = \frac{\pi x^5}{5}|_0^5 = 5^4\pi - 0 = 625\pi.$$

Example 6.31. What is the volume of the solid obtained by rotatin the region bounded by $y = x^2, y = 25$ with $x \geq 0$ around the y-axis?

As before, we draw a picture. Our region has width $\sqrt{y}$ at a point y, and thus has cross-sectional area πy. Then y varies from 0 to 25, and the volume is

$$V = \int_0^{25} \pi y \, dy = \frac{\pi y^2}{2}|_0^2 5 = \frac{625\pi}{2}.$$

Note that in these problems it's easy to see which way to take our "slices": we want to get the circular cross-sections from the rotation, so we slice accordingly, and integrate along the axis we rotate around.

If our region touches the axis we rotate it around, these problems are straightforward: the cross-sectional area is the height (or width!) of the region squared times π. The problem is trickier if we have a hollow inside. We can still compute the cross-sectional area; it is the area of a *washer*, a circle with a smaller circle cut out of the center.

Remark 6.32. If a washer has outer radius R and inner radius r, then the area is $\pi R^2 - \pi r^2$, the area of the outer circle minus the radius of the inner.

Example 6.33. What is the volume of the solid given by rotating the region bounded by $y = x^2$ and $y = x$ around the x-axis.

At a point x, the cross-section of this solid is a washer. The outer circle has radius x and the inner circle has radius x^2, and thus the area of the cross-section is $\pi x^2 - \pi x^4$. So the volume is

$$V = \int_0^1 (\pi x^2 - \pi x^4) \, dx = \frac{\pi x^3}{3} - \frac{\pi x^5}{5} = \frac{\pi}{3} - \frac{\pi}{5} = \frac{2\pi}{15}.$$

We often find ourselves rotating these regions around lines other than the x- or y-axes. In this case we have to use our geometric intuition a bit more to sort out our cross-sectional areas.

Example 6.34. Rotate the same region about $y = 2$. We draw a picture; we see that we will get a solid whose cross-sections are washers centered at $y = 2$. The outer radius will be $2 - x^2$ and the inner radius will be $2 - x$, so the volume is

$$V = \int_0^1 \pi(2 - x^2)^2 - \pi(2 - x)^2 \, dx = \pi \int_0^1 4 - 4x^2 + x^4 - 4 + 4x - x^2 \, dx = \pi \int_0^1 x^4 - 5x^2 + 4x \, dx$$
$$= \pi \left(\frac{x^5}{5} - \frac{5x^3}{3} + 2x^2 \right) \Big|_0^1 = \pi(1/5 - 5/3 + 2) = \frac{4\pi}{15}.$$

Example 6.35. Find the volume of the solid generated by rotating the region bounded by $y = x$ and $y = \sqrt{x}$ about the line $y = 1$.

We will integrate with respect to x since we rotate about a line parallel to the x-axis. We see that the curves intersect at $x = y = 0$ and $x = y = 1$. Our cross-sections are washers, and we see the outer radius is $1 - x$ and the inner radius is $1 - \sqrt{x}$. So the volume is

$$V = \pi \int_0^1 (1-x)^2 - (1-\sqrt{x})^2 \, dx = \pi \int_0^1 x^2 - 3x + 2\sqrt{x} \, dx$$
$$= \pi \left(\frac{x^3}{3} - \frac{3x^2}{2} + \frac{4}{3} x^{3/2} \right) \Big|_0^1 = \pi \left(\frac{1}{3} - \frac{3}{2} + \frac{4}{3} \right) = \frac{\pi}{6}.$$

6.5 Bonus material: Finding Volumes with Cylindrical Shells

Recall we want to find the volume of the solid obtained by rotating the region bounded by $x = 1, y = 2, y = \ln x$ about the x-axis. Slicing it into washers as before generates a difficult integral, so we will try to slice it a different way, by slicing it into *cylindrical shells*.

A cylindrical shell is what we get when we take a cylinder and remove a slightly smaller cylinder from the inside. If the outer radius is r_2 and the inner radius is r_1, it's not hard to see that the volume of the shell is $\pi r_2^2 h - \pi r_1^2 h = \pi h(r_2^2 - r_1^2)$. Less obviously, we factor $r_2^2 - r_1^2 = (r_2 + r_1)(r_2 - r_1)$ and write that the volume is $2\pi \frac{r_1+r_2}{2} h(r_2 - r_1) \approx 2\pi r h \Delta r$.

In many solids of rotation, we can slice the solid into a collection of cylindrical shells to approximate the volume, where the height of each cylinder is $f(x)$ for some x. We get the formula

$$V \approx \sum_{i=1}^n 2\pi x_i^* f(x_i^*) \Delta x.$$

As before, our approximation gets better as we use more and thinner cylinders, and when we take the limit, we get

$$V = \lim_{\max \Delta x_i \to 0} \sum_{i=1}^n 2\pi x_i^* f(x_i^*) \Delta x = \int_a^b 1\pi x f(x) \, dx,$$

where a is the inner radius of our entire solid, and b is the outer radius of the entire solid. (Note that this formula is essentially the surface area of the cylinder; this isn't an accident).

So for our earlier example, we can slice into cylinders whose height is in the x-direction. We see that

$$V = \int_0^2 2\pi y(e^y - 1) \, dy = 2\pi \left(ye^y - e^y - \frac{y^2}{2} \right) \Big|_0^2 = 2\pi(e^2 - 1).$$

Remark 6.36. Unlike in the method of washers, this time we will typically integrate with respect to x when we rotate around the y-axis, and vice versa.

Example 6.37. Find the volume of the solid obtained by rotating the region bounded by $y = 0$ and $y = x - x^2$ around the line $x = 2$.

Inverting the function $y = x - x^2$ would be a huge pain; so we'd like to integrate with respect to x, and thus use the cylinder method. Note that in this case the radius r is not x, but is $2 - x$. So the volume is

$$V = \int_0^1 2\pi(2-x)(x-x^2)\,dx = 2\pi \int_0^1 2x-3x^2+x^3\,dx = 2\pi \left(\frac{x^4}{4} - x^3 + x^2\right)\Big|_0^1 = 2\pi(1/4-1+1) = \frac{\pi}{2}.$$

Example 6.38. What is the volume of the solid obtained by rotating the region bounded by $y = x^3, y = 0, x = 1$ around the line $x = 1$?

$$V = \int_0^1 2\pi(1-x)x^3\,dx = 2\pi\left(\frac{x^4}{4} - \frac{x^5}{5}\right)\Big|_0^1 = 2\pi\left(\frac{1}{4} - \frac{1}{5}\right) = \frac{\pi}{10}.$$

Example 6.39. What is the volume of the solid obtained by rotating the same region around the line $x = 4$?

$$V = \int_0^1 2\pi(4-x)x^3\,dx = 2\pi\left(x^4 - \frac{x^5}{5}\right)\Big|_0^1 = 2\pi\left(1 - \frac{1}{5}\right) = \frac{8\pi}{5}.$$

Example 6.40. What is the volume of the solid obtained by rotating the region bounded by $xy = 1, x = 0, y = 1, y = 3$ about the x-axis?

We draw a picture, and conclude that to use the method of washers we'd have to break the region up into two pieces. Instead we integrate with respect to y and use cylindrical shells. We have y varying from 1 to 3, and the "height" of each cylinder is $1/y - 0$. So the volume is

$$V = \int_1^3 2\pi y(1/y)\,dy = \int_1^3 2\pi\,dy = 2\pi y|_1^3 = 4\pi.$$

Example 6.41. A word has to be said at this point about finding the volume of a sphere. We can view the sphere as a solid of rotation and find its volume using cross-sections:

$$\begin{aligned} V &= \int_{-r}^{r} \pi(\sqrt{r^2 - x^2})^2\,dx = \pi\int_{-r}^{r} r^2 - x^2\,dx = \pi\left(r^2 x - \frac{x^3}{3}\right)\Big|_{-r}^{r} \\ &= \pi\left(\left(r^3 - r^3/3\right) - \left(-r^3 + r^3/3\right)\right) = 4\pi r^3/3. \end{aligned}$$

But we can actually use another approach, similar in spirit to the method of cylindrical shells. We can look at the sphere as being made up of a collection of spherical shells. Taking inspiration from the cylindrical shells method, we see that the volume of each spherical shell will be "about" the surface area of the sphere times thickness; so we integrate the surface area of a sphere of radius x, as x varies from 0 to r. We get

$$V = \int_0^r 4\pi x^2 \, dx = \frac{4\pi x^3}{3}\Big|_0^r = \frac{4\pi r^3}{3}.$$

We haven't entirely justified our argument, but with more care we certainly could.