MATH 1070: Mathematics of Change, part I (article-1.html)

Calculus Group

Preface: approximations and bounds

This e-textbook is about using math for modeling and coming up with plausible analyses. One of the course goals is number sense. Wikipedia defines this as "an intuitive understanding of numbers, their magnitude, relationships, and how they are affected by operations."

> **Example 0.0.1. Food for thought.** If you have a model for spread of disease where the number of infections doubles every three days, how long can this go on before the model has to change: A few years? A few months? A few weeks? A few days?

If some kind of answer began to form in your mind without your stopping to get out a calculator, then you have some of the ingredients of number sense already: perhaps you understand exponential growth, perhaps you can remember about how many people there are in the country or the world, perhaps you are familiar with powers of two and know how they relate to this problem. It is useful to be able to think this way. It's not important whether you use a calculator to answer any given question, but realistically, how often will you stop in casual conversation and whip out a calculator?

In this course, we'll teach you a number of these ingredients: use of logs, converting to powers of ten, tangent line approximation, Taylor polynomials, pairing off positive and negative summands, approximating integrals with sums and vice versa. Discussions of these will be brief. The point is to use them when you need them, which turns out to be nearly every lesson.

Today we're going to start with the tangent line approximation. You might think this odd because we haven't taught you calculus yet. Calculus is a way of computing the slopes of tangent lines to graphs. But conceptually, understanding the tangent line approximation takes knowledge only of algebra and geometry, not calculus. So, we'll preview the idea now, and in fact several more ideas from the course, and then later see how to use calculus to do these analyses more methodically.

1 Where should I put my ladder?

I am hanging wind chimes on my balcony using a ladder 5 meters long. On the highest safe step, my shoulders will be exactly at the top of the ladder, which I need to be at the height of the balcony rail, 4 meters above the ground. Every time I reposition the ladder I scratch the paint, so I'd rather not move it too many times. I need to get my shoulders within a couple of centimeters of the right height in order to drive a nail into the lintel. Where should I put the base of the ladder? The Pythagorean theorem tells me the it should be 3 meters from the wall; see Figure 1.0.1

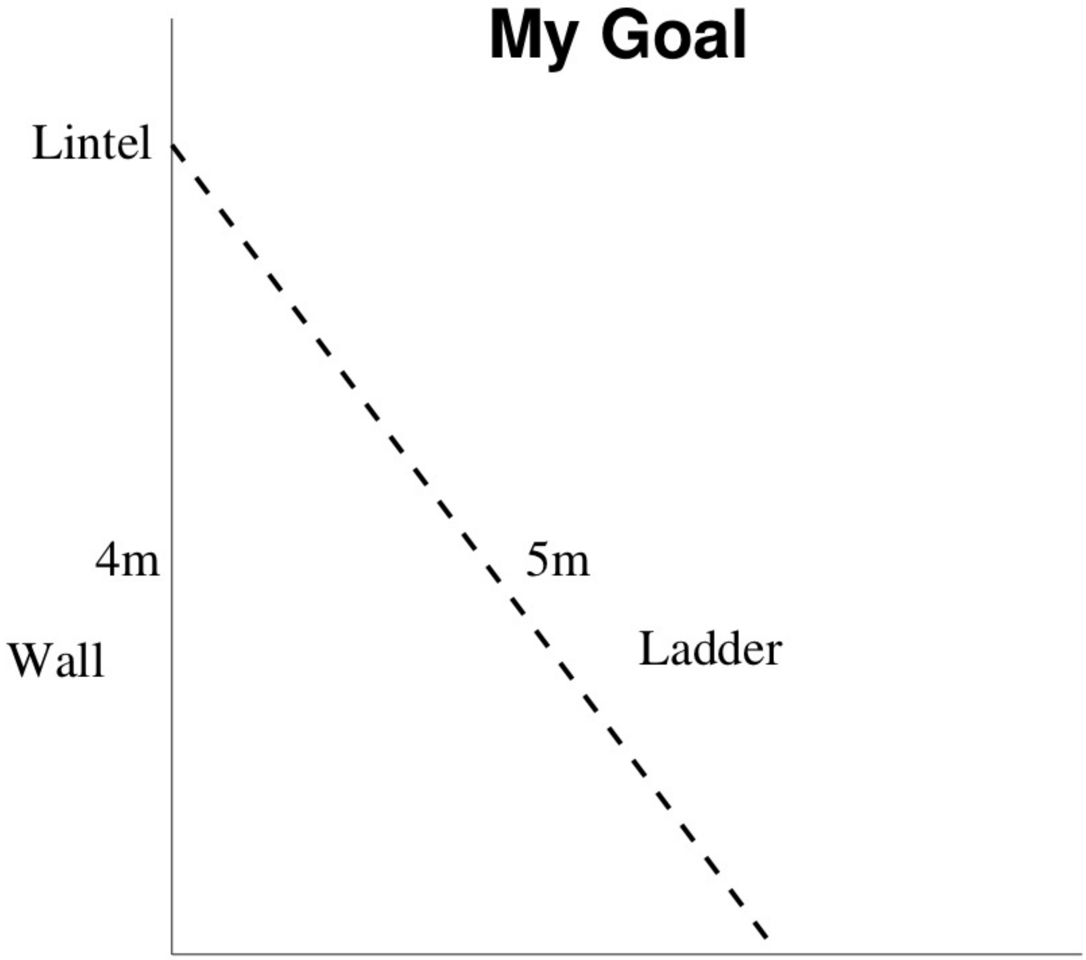


Figure 1.0.1. My Goal

Unfortunately, I didn't measure right, maybe because of the bulge at the bottom of the balcony. I am 20 cm too low. Now what? Solution: Let h be the function representing the height of the ladder as a function of the position of the base, in other words, $h(x)$ is the height of the ladder on the wall (in meters) when the base is x meters from the wall. By the Pythagorean Theorem, $h(x) = \sqrt{25 - x^2}$.

The height I am trying to reach is shown in Figure 1.0.2. which has $x = 3$ and $h(x) = 4$. Instead I hit some other point z with $h(z) = 3.8$. Clearly z is too far from the wall. How far do I need to scoot the ladder toward the wall? As you can see in Figure 1.0.2, due to the balcony and the hedge, it was not feasible to measure either the height of the lintel or the distance I placed the foot of the ladder more accurately.

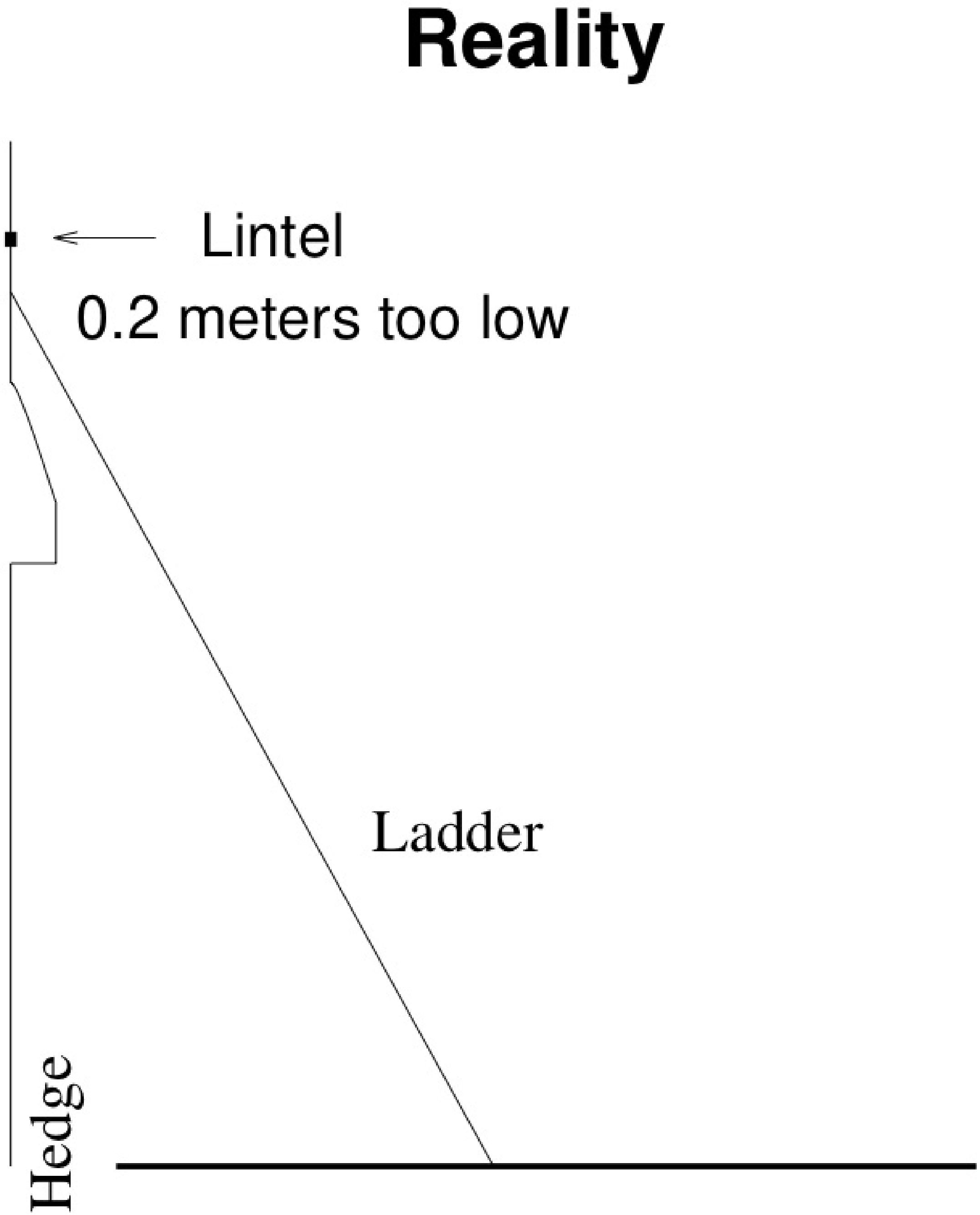


Figure 1.0.2. Reality

Figure 1.0.3 shows the graph of h and a tangent line to the graph of h at the point $(3, 4)$. The tangent line is a very good approximation to the graph near $(3, 4)$. For values of x between perhaps 2.6 and 3.4, the line is still visually indistinguishable form the graph.

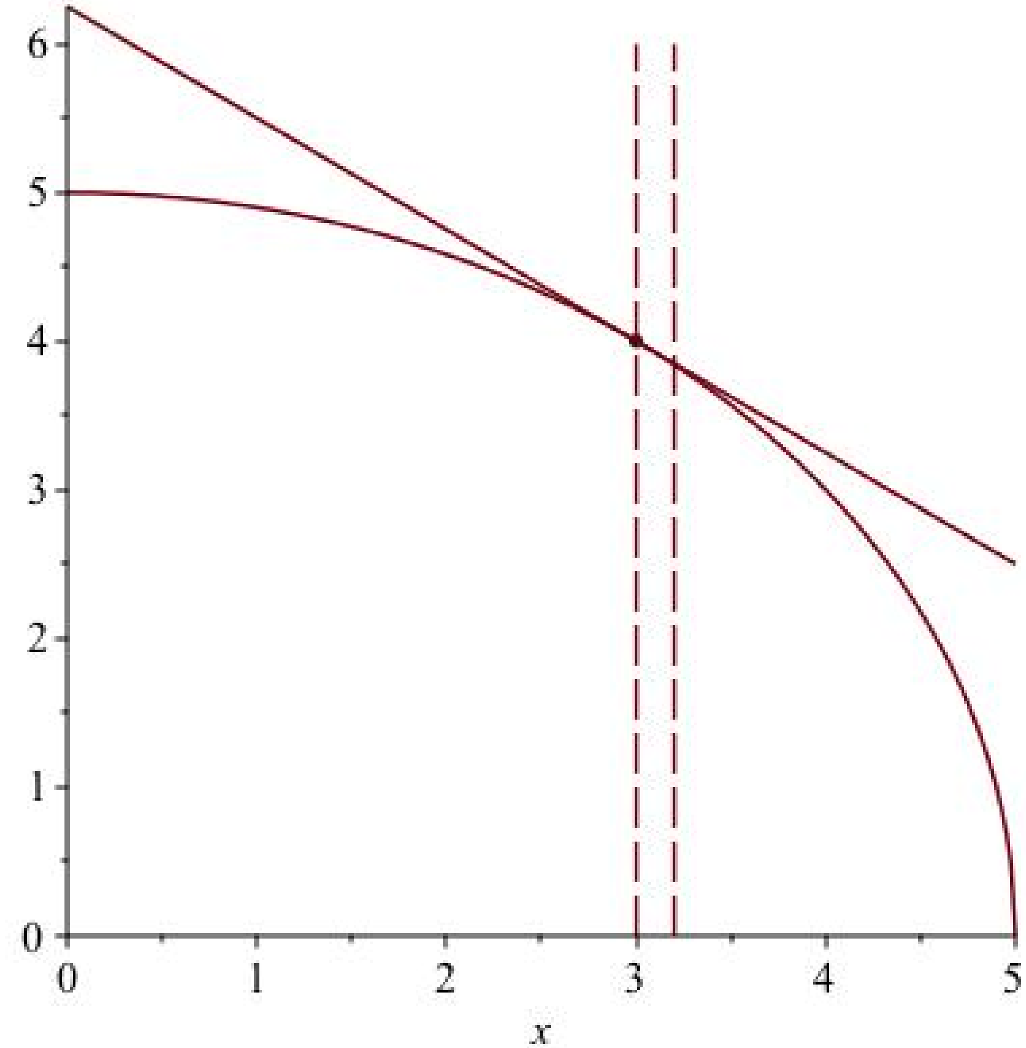


Figure 1.0.3. math.upenn.edu/~ancoop/1070/sec-zero.html#fig-ladder3)

If we know the slope of this line, m, we can write the equation of the line: $(y-4) = m(x-3)$. Because $h(x)$ is very nearly equal to this y (because the curve nearly coincides with the line), we can write $h(x) \approx 4 + m(x-3)$. The wiggly equal sign is not a formal mathematical symbol. Here, it means the two will be close, but has no guarantee of how close, and furthermore, it is only supposed to be close when x is close to 3. This is called an {\bf estimate}. Shortly, we will talk about bounds: estimates that do come with guarantees.

What is the slope of this line? The graph is a quarter-circle. Recall from geometry that any tangent to a circle makes a right angle with the radius. The slope of the radius from $(0,0)$ to $(3,4)$ is $s = 4/3$. The slope of any line making a right angle with this is the negative reciprocal $-1/s = -3/4$. In other words, the slope of the tangent line is $-3/4$.

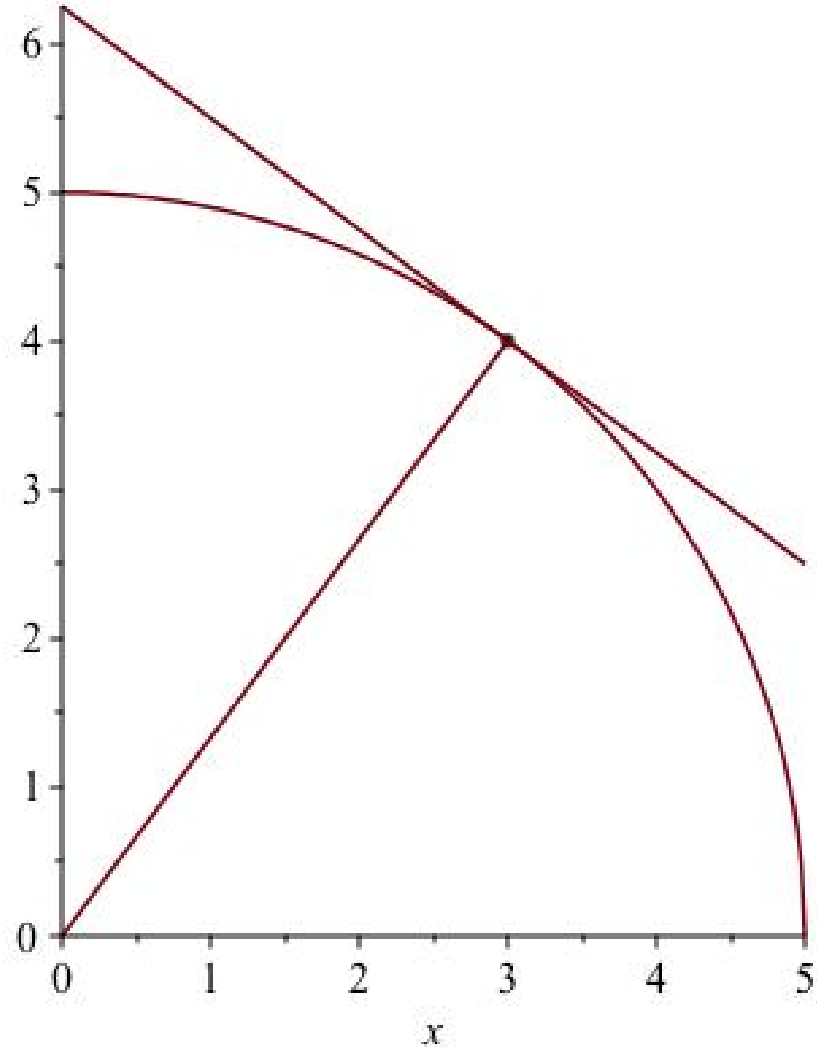


Figure 1.0.4. nath.upenn.edu/~ancoop/1070/sec-zero.html#fig-ladder4)

Checkpoint 1. enn.edu/~ancoop/1070/sec-zero.html#project-1)

The reason we chose this particular example to demonstrate the tangent line approximation is that we could compute the slope with high school geometry. With calculus, we'll be able to find the slope of a tanget line to just about any function we can think of. In fact the word calculus when it was invented meant literally "a method of computing".

Checkpoint 2. enn.edu/~ancoop/1070/sec-zero.html#project-2)

2 Bounding enn.edu/~ancoop/1070/sec-zero.html#subsection-2)

To get an upper bound on $f(x)$ means to find a quantity $U(x)$ that you understand better than $f(x)$ and for which you can prove that $U(x) \geq f(x)$. A lower bound is a quantity $L(x)$ that you understand better than $f(x)$ and that you can prove to satisfy $L(x) \leq f(x)$. If you have both a lower and upper bound, then $f(x)$ is stuck for certain in the interval $[L(x), U(x)]$.

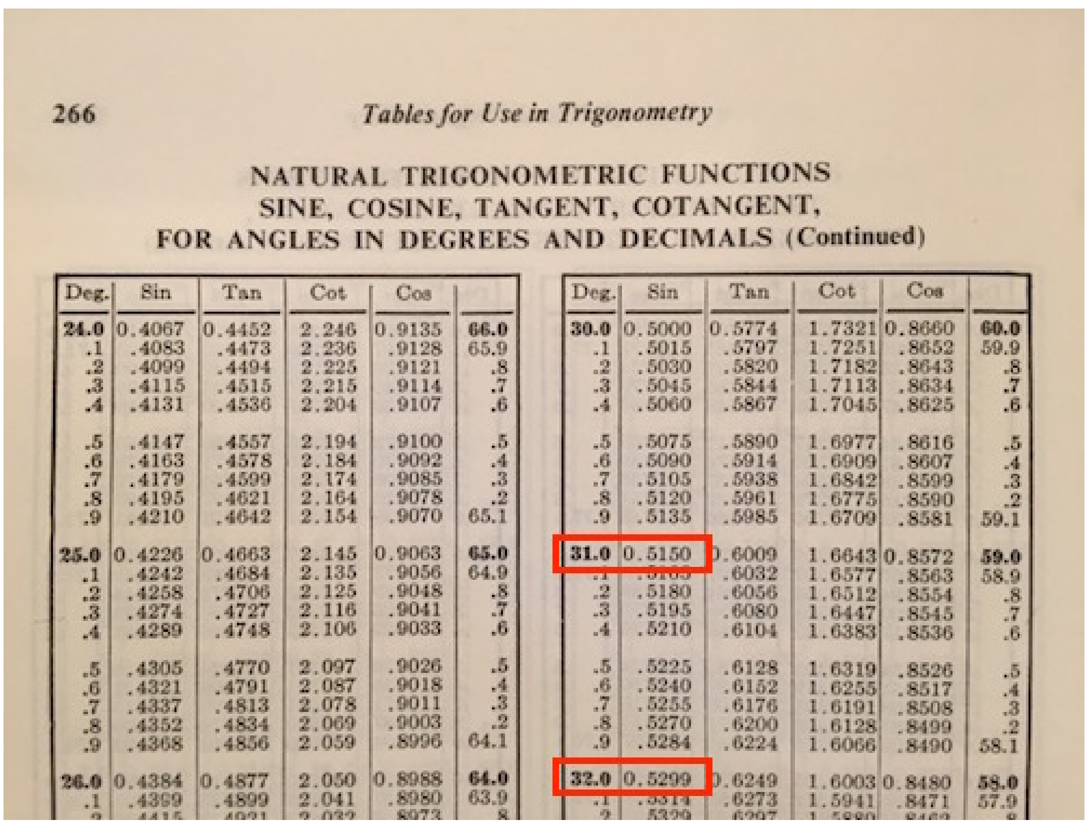
Checkpoint 3. enn.edu/~ancoop/1070/sec-zero.html#project-3)

In a way bounding is harder than estimating because there is no one correct bound (there's no one correct estimate either, but we usually a

While estimating produces statements that are not mathematically well-defined, bounding produces inequalities with precise mathematical meaning. Two ways we typically find bounds are as follows.

First, if f is monotone increasing then an easy upper bound for $f(x)$ is $f(u)$ for any $u \geq x$ for which we can compute $f(u)$. Similarly an easy lower bound is $f(v)$ for any $v \leq x$ for which we can compute $f(v)$. If f is monotone decreasing, you can swap the roles of u and v in finding upper and lower bounds. There are even stupider bounds that are still useful, such as $f(x) \leq C$ if f is a function that never gets above C.

Example 2.0.1. An Unexample. Suppose we wanted to estimate $\sin(1)$. The easiest upper and lower bounds are 1 and -1 respectively because $\sin$ never goes above 1 or below -1. A better lower bound is 0 because $\sin(x)$ remains positive until $x = \pi/2$ and obviously $1 < \pi/2$. You might in fact recall that one radian is just a bit under 60°, meaning that $\sin(60^\circ) = \sqrt{3}/2 \approx 0.0866\ldots$ is an upper bound for $\sin(1)$. Computing more carefully, we find that a radian is also less than 58°. Is $\sin(58^\circ)$ a better upper bound? Probably not: ***we don't know how to calculate it***, so it's not a quantity we understand better. Of course is we had an old-fashioned table of sines, and all we can remember about one radian is that it is between 57° and 58°, then $\sin(58^\circ)$ is not only an upper bound but the best one we have.



266 *Tables for Use in Trigonometry*

NATURAL TRIGONOMETRIC FUNCTIONS
SINE, COSINE, TANGENT, COTANGENT,
FOR ANGLES IN DEGREES AND DECIMALS (Continued)

Deg.	Sin	Tan	Cot	Cos	
24.0	0.4067	0.4452	2.246	0.9135	**66.0**
.1	.4083	.4473	2.236	.9128	65.9
.2	.4099	.4494	2.225	.9121	.8
.3	.4115	.4515	2.215	.9114	.7
.4	.4131	.4536	2.204	.9107	.6
.5	.4147	.4557	2.194	.9100	.5
.6	.4163	.4578	2.184	.9092	.4
.7	.4179	.4599	2.174	.9085	.3
.8	.4195	.4621	2.164	.9078	.2
.9	.4210	.4642	2.154	.9070	65.1
25.0	0.4226	0.4663	2.145	0.9063	**65.0**
.1	.4242	.4684	2.135	.9056	64.9
.2	.4258	.4706	2.125	.9048	.8
.3	.4274	.4727	2.116	.9041	.7
.4	.4289	.4748	2.106	.9033	.6
.5	.4305	.4770	2.097	.9026	.5
.6	.4321	.4791	2.087	.9018	.4
.7	.4337	.4813	2.078	.9011	.3
.8	.4352	.4834	2.069	.9003	.2
.9	.4368	.4856	2.059	.8996	64.1
26.0	0.4384	0.4877	2.050	0.8988	**64.0**
.1	.4399	.4899	2.041	.8980	63.9
.2	.4415	.4921	2.032	.8973	.8

Deg.	Sin	Tan	Cot	Cos	
30.0	0.5000	0.5774	1.7321	0.8660	**60.0**
.1	.5015	.5797	1.7251	.8652	59.9
.2	.5030	.5820	1.7182	.8643	.8
.3	.5045	.5844	1.7113	.8634	.7
.4	.5060	.5867	1.7045	.8625	.6
.5	.5075	.5890	1.6977	.8616	.5
.6	.5090	.5914	1.6909	.8607	.4
.7	.5105	.5938	1.6842	.8599	.3
.8	.5120	.5961	1.6775	.8590	.2
.9	.5135	.5985	1.6709	.8581	59.1
31.0	0.5150	0.6009	1.6643	0.8572	**59.0**
.1	[illegible]	.6032	1.6577	.8563	58.9
.2	.5180	.6056	1.6512	.8554	.8
.3	.5195	.6080	1.6447	.8545	.7
.4	.5210	.6104	1.6383	.8536	.6
.5	.5225	.6128	1.6319	.8526	.5
.6	.5240	.6152	1.6255	.8517	.4
.7	.5255	.6176	1.6191	.8508	.3
.8	.5270	.6200	1.6128	.8499	.2
.9	.5284	.6224	1.6066	.8490	58.1
32.0	0.5299	0.6249	1.6003	0.8480	**58.0**
.1	.5314	.6273	1.5941	.8471	57.9
.2	.5329	.6297	[illegible]	[illegible]	.8

Figure 2.0.2. A table of sines -- what was used before electronic calculators were widely available.

Checkpoint 4.

3 Concavity

A more subtle bound come when f is known to be concave upward or downward in some region. By definition, a concave upward function lies below its chords and a concave downward function lies above its chords.

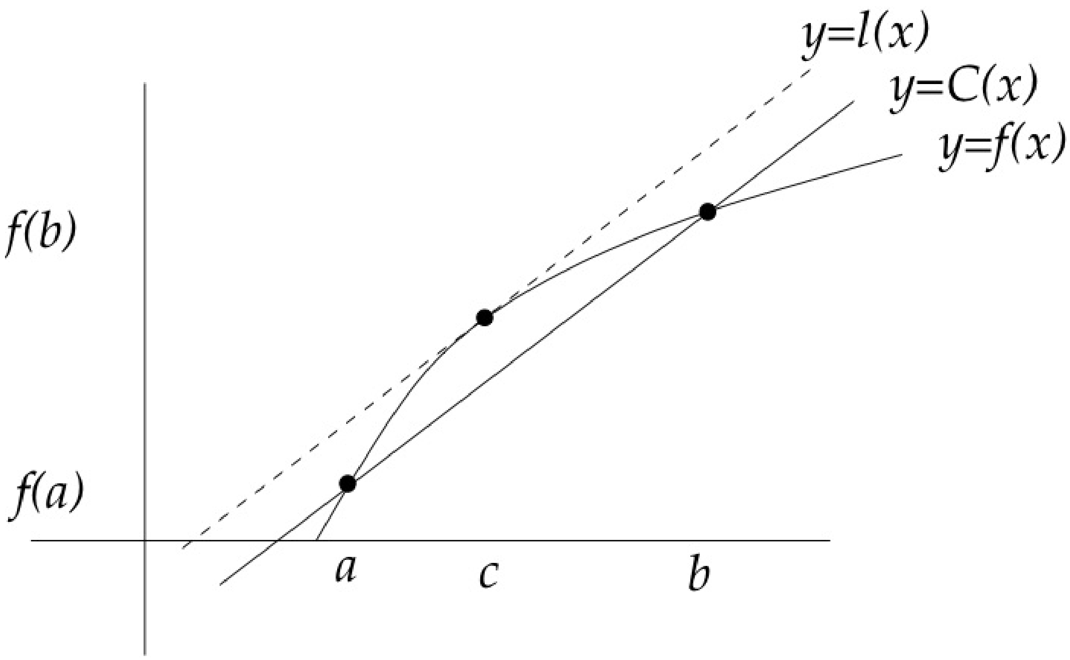


Figure 3.0.1. Chords of a graph.

Figure 3.0.1 shows a function $f(x)$ which is concave down. As long as x is in the interval $[a, b]$, we are guaranteed to have $C(x) \leq f(x)$. On the other hand, when $f'' < 0$ on an interval, the function always lies below the tangent line. Therefore $L(x)$ is an upper bound for $f(x)$ when $x \in [a, b]$ no matter which point $c \in [a, b]$ at which we choose to take the linear approximation.

In the ladder example, we were lucky that the graph was a familiar geometric shape, a quarter circle, which we know to be convex. We are able to conclude that the tangent line remains above the graph because we know geometrically that the tangent line to a circle touches the circle at one point and otherwise remains outside the circle. Calculus will give us a far more general way to determine concavity.

Checkpoint 5.

∑ () MATH 1070: Mathematics of Change, part I (article-1.html)

Calculus Group

0 Variables, functions and graphs

If we count pre-calculus/trigonometry as a pre-requisite, then functions and their graphs are a pre-pre-requisite! But that doesn't mean that you have familiarity with every aspect of these. Recognition of basic types of functions is crucial for being able to use mathematics for modeling and to handle material at the pace and level you will need. So is the ability to go back and forth between analytic expressions for functions and their graphs. So is number sense: knowing approximate values without stopping for a detailed calculation. So is knowledge of how to use physical units in a math problem. We expect most of these to be unfamiliar to many of you, and have included explanations and some homework; this may be challenging, even though it is on pre-college material. We hope it will be at least somewhat interesting!

In addition, there are some more routine things to discuss up front. In order to have a shared language, we need to agree on notation and terminology. Normally it is a good idea to read everything that is assigned; however if this notation is very familiar, you can probably just answer the self-check questions and skip the reading. We apologize for the length of this preliminary section. When the material becomes harder, the sections will be shorter.

1.1 Notation and terminology

There are several ways to conceive of a function. One is that it is a ***rule*** that takes an input and gives you an output. This is how most of us think of functions most of the time, but it is not precise (rules are sentences which may be ambiguous or underspecified). For this reason we also need a ***formal definition***. A third way to understand functions is via their ***graphs***. We now discuss all three of these ways of characterizing a function, beginning with the most formal.

> **Definition 1.1.1.**
>
> 1. A function is a set of ordered pairs with the property that no two ordered pairs have the same first element. The ***domain*** of f is defined to the set of all first elements of the ordered pairs. The ***range*** of f is defined to be the set of all second elements of the ordered pairs.
> 2. The expression $f(x)$ is defined to equal to y if the ordered pair (x, y) is in the set of ordered pairs defining f and undefined otherwise. Informally, $f(x)$ is called the ***value*** of the function f evaluated at the ***argument*** x.

Now let's say the same things verbally. The ***domain*** of a function is the set of allowed inputs; the ***range*** is the set of all outputs. We often name functions with letters; f is the typical choice, then g if another is needed, but of course we could name a function anything. While it is common to refer to the function f as $f(x)$, we will try to observe the distinction that f is the function and $f(x)$ is its ***value*** at the

argument x, meaning the output when you plug in x as an input. The condition that first coordinates are distinct corresponds to the rule producing an unambiguous answer.

Finally, to describe the function f via its graph, we interpret the ordered pairs as points in the plane, and draw this set as a curve. The condition that first coordinates are distinct corresponds to the so-called ***vertical line test***: any vertical line (vertical lines being sets with a single fixed x-coordinate but all possible y-coordinates) intersects the graph at most once.

Why be so formal? In common usage, one might encounter any of the three ways of defining or referring to a function. We don't want to drown in formality, so we usually use something only as formal as needed. Let's look at why we sometimes need formality.

Example 1.1.2. Suppose we define a function f by $f(x) := x^2 + 2$. Have we formally defined this function? It sounds as if this is the set of ordered pairs

$$\{\ldots, (-2, 6), (-1, 3), (0, 2), (1, 3), (2, 6), \ldots\}.$$

That would be if we meant the domain to be the set of all integers. Maybe instead we meant the domain to be the set of all real numbers. In that case, the "..." in the list is somewhat misleading; we should probably write the set of ordered pairs like this:

$$\{(x, x^2 + 2) : x \in \mathbb{R}\}$$

(we use the notation $\mathbb{R}$ for the real numbers and $\in$ for the "is an element of"). If this function arose in a word problem where $f(x)$ represented the value of some quantity at a time x seconds after the start, maybe it makes sense to allow only nonnegative real numbers as inputs. Formally, this would look like

$$\{(x, x^2 + 2) : x \text{ is real and nonnegative}\},$$

which could also be written

$$\{(x, x^2 + 2) : x \in [0, \infty)\}$$

or

$$\{(x, x^2 + 2) : x \geq 0\},$$

this last version assuming we understood this to mean real numbers at least zero rather than, say, integers at least zero.

A consequence of this discussion is that two functions are technically different if they have different domains, even if they have the same defining rule. We mostly don't care about this distinction unless it matters to some problem, such as trying to determine the number of solutions to an equation, as in the next exercise.

Technically, our discussion of the function $x \mapsto x^2 + 2$ referred to three different functions: one whose domain was all integers, one whose domain was all reals, and one whose domain is all nonnegative reals. You can see they are different functions: even thought the defining equation $f(x) := x^2 + 2$ is the same for all three, they are defined by different sets of ordered pairs. On the other hand, for many purposes, we don't care which of these functions was intended. We can feel free to define the function by $f(x) := x^2 + 2$ without specifying the domain unless and until we get into trouble with the ambiguity in the domain. If we try to answer a question like "How many solutions are there to $f(x) = 3$?" then we will need to be more precise about the domain.

Checkpoint 6.

In the discussion so far, we have introduced four notations you are probably familiar with, but to be completely explicit, we discuss each briefly.

Maps-to notation. Often we name a function when defining it, then refer to it by name, but we can also refer to it using the "maps-to" symbol $\mapsto$. Thus, $x \mapsto x^2 + 2$ refers to the function that we named f, above. We use this when mentioning a function but rarely when evaluating it at an argument because the notation $(x \mapsto x^2 + 2)(3)$ is an atrocity (but technically equal to 11).

Open and closed interval notation. The interval $[a, b]$ refers to all real numbers x such that $a \leq x \leq b$. When both endpoints are included, this is called a ***closed*** interval.

The interval (a, b) refers to all real numbers x such that $a \leq x \leq b$. When both endpoints are excluded, this is called an ***open*** interval.

> **Warning 1.1.3.** The notation for an open interval is exactly the same as for an ordered pair! If there is any ambiguity we will try to specify which, for example, "Let (a, b) be the open interval..."

The notations $(a, b]$ and $[a, b)$ are called *half-open* and refer to an interval with one point (the one next to the square bracket) included and one excluded.

Set-builder notation. To define a subset of some set S, we write $\{x \in S : \cdots\}$ where instead of the three dots we write a property of x that can be true or false. In some books the colon is replaced by a vertical line, the words "such that" or the abbreviation $s.t.$. If the set S is the set of all real numbers it is sometimes omitted. Thus, for example, $\{x : a \leq x < b\}$ refers to the half open interval of real numbers, $[a, b)$.

The defining colon-equal sign. We use $:=$ to mean that the quantity on the left is defined to be the quantity on the right, and a regular equal sign to mean an equation that could hold for some values of the variables and fail for others. Thus, $f(x) := x^2 + 2$ defines a function, whereas $f(x) = x^2 + 2$ is an equation which is true when a given function f, evaluated at x, has the same value as $x^2 + 2$, and false otherwise.

Checkpoint 7.

The range of a function f is defined to be the set of all possible function values. Formally we can write the range of f as the set $\{f(x) : x \in \text{domain of } f\}$.

One final remark about the basic definitions: there is an ambiguity in common usage of the word "range". Sometimes "range" is used to refer to a bigger set than in our definition, namely the set of all things of the type that the function outputs (we'll call this other set the ***target*** of the function, should we ever need to refer to it.). For example, they might say that the domain and range of a function $f(x) := x^2 + 2$ is all real numbers. It's fine to define the domain to be all real numbers, but then technically the range is the set of real numbers that are at least 2. If for some reason we want to tell the person we're talking to that we intend the outputs to always be real numbers, we can say "the target is the real numbers".

Checkpoint 8. penn.edu/~ancoop/1070/section-1.html#project-8)

Definition by cases. As we said, the most familiar way of referring to a function is as a rule for converting input to output. Usually the rule is an equation, such as $f(x) := C - x \cdot e^x$, but the rule could be verbal, for example, "Let $f(t)$ be the amount in tons of carbon dioxide emitted in t years." Sometimes we want to talk about functions that are defined by equations, but different ones in different parts of the domain. This is called ***definition by cases***. An example from a recent research paper looks like this:

$$f(x) := \begin{cases} -9x & a \leq -3 \\ 2x^2 - 3x & -3 < x < 1 \\ -a^3 & a \geq 1 \end{cases} \quad .$$

A number of useful functions can be defined in this way. For example the absolute value of x, denoted $|x|$, may also be defined in cases:

$$|x| := \begin{cases} x & x \geq 0 \\ -x & x < 0 \end{cases} \quad .$$

Remark 1.1.4. Some remarks on defining by cases.

1. Note that x and $-x$ agree at $x = 0$, so we could have grouped zero with either case. When this happens, writing

   $$|x| := \begin{cases} x & x \geq 0 \\ -x & x \leq 0 \end{cases}$$

   emphasizes this. If x and $-x$ did not agree at $x = 0$, this would be a badly formed definition.

2. There is a period following the two example definitions but not the one in the first remark. Why? Because well written math follows rules of basic grammar. These rules can be a little different on occasion, but for the most part, you should expect this text to read in complete sentences, to define variables and functions before using them, and when used within sentences, to connect and flow logically, using connecting words like "and", "because", "therefore", and punctuation such as commas and periods.

Checkpoint 9. penn.edu/~ancoop/1070/section-1.html#project-9)

Free and bound variables. In the defining statement $f(x) := x^2 + 2$, it would define the same function if instead we said $f(u) := u^2 + 2$. It is the same set of order pairs, has the same graph, etc. The variable x (or in the second case, u) is said to be a ***bound variable***. The bound variable in this case runs over all values in the domain of f. A variable that is not bound is ***free***. For example, in the definition $f(u) := u^2 + c$, the variable c is free. The definition of the function f depends on the value of c. If $c = 2$, it boils down to the previous definition. If $c = 1$ it is a different function. If c has not been assigned a value, then f is a function whose range is not the real numbers but rather algebraic expressions in the variable c.

Bound variables arise many times throughout this course, in fact throughout math and throughout life! Here is a list of some places bound variables occur in this course, the first two of which you have already seen.

- In the definition of a function
- In the definition of a subset
- In limits
- In the definition of a derivative
- In summations
- In the definition of an integral
- In notions of orders of magnitude and asymptotic equivalence
- In Taylor's theorem

Let's get some practice with identifying free and bound variables.

Definition 1.1.5. A function f is said to be ***differentiable at*** x if x is in the domain of f and $f'(x)$ exists.

Definition 1.1.6. A function f is said to be ***differentiable on the open interval*** (a, b) if (a, b) is in the domain of f and if, for all $x \in (a, b)$, the derivative $f'(x)$ exists.

Checkpoint 10. enn.edu/~ancoop/1070/section-1.html#project-10)

1.2 Some useful functions

Here are some more useful special functions.

The ***greatest integer*** function at the ***argument*** x is denoted $\lfloor x \rfloor$ defined to be the greatest integer y such that $y \leq x$. In other words, if x is an integer then $\lfloor x \rfloor = x$; if x is positive and not an integer, then $\lfloor x \rfloor$ is the "whole number you get when you write x as a decimal and ignore what comes after the decimal point"; if x is negative and not an integer, it is -1 plus what you get when you ignore the decimals. In older texts, the same function is sometimes denoted $[x]$. This square bracket notation has largely been abandoned in favor of the "floor" notation, because (especially in computer science) we also often want to use the ***ceiling*** function as well. The ceiling function at the argument x is denoted $\lceil x \rceil$ and is defined to be the least integer y such that $y \geq x$. Informally, $\lfloor x \rfloor$ rounds *down* to the nearest integer and $\lceil x \rceil$ rounds *up*.

Checkpoint 11. enn.edu/~ancoop/1070/section-1.html#project-11)

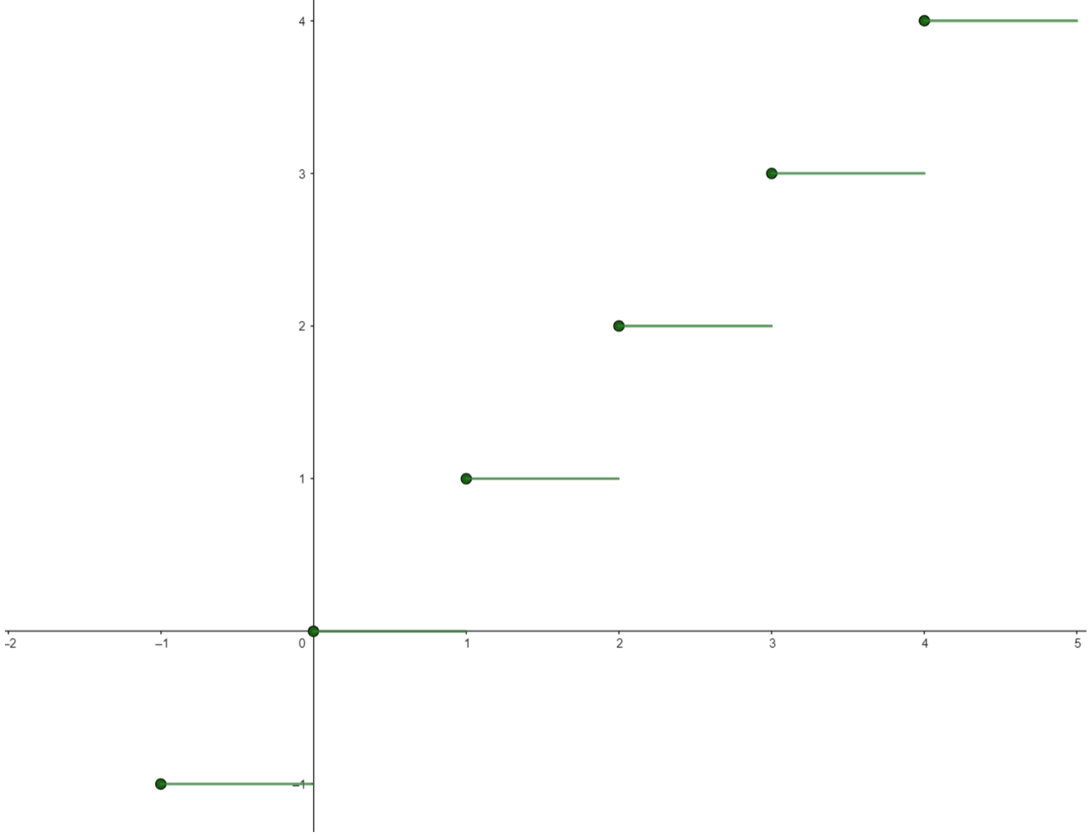


Figure 1.2.1. The floor function.

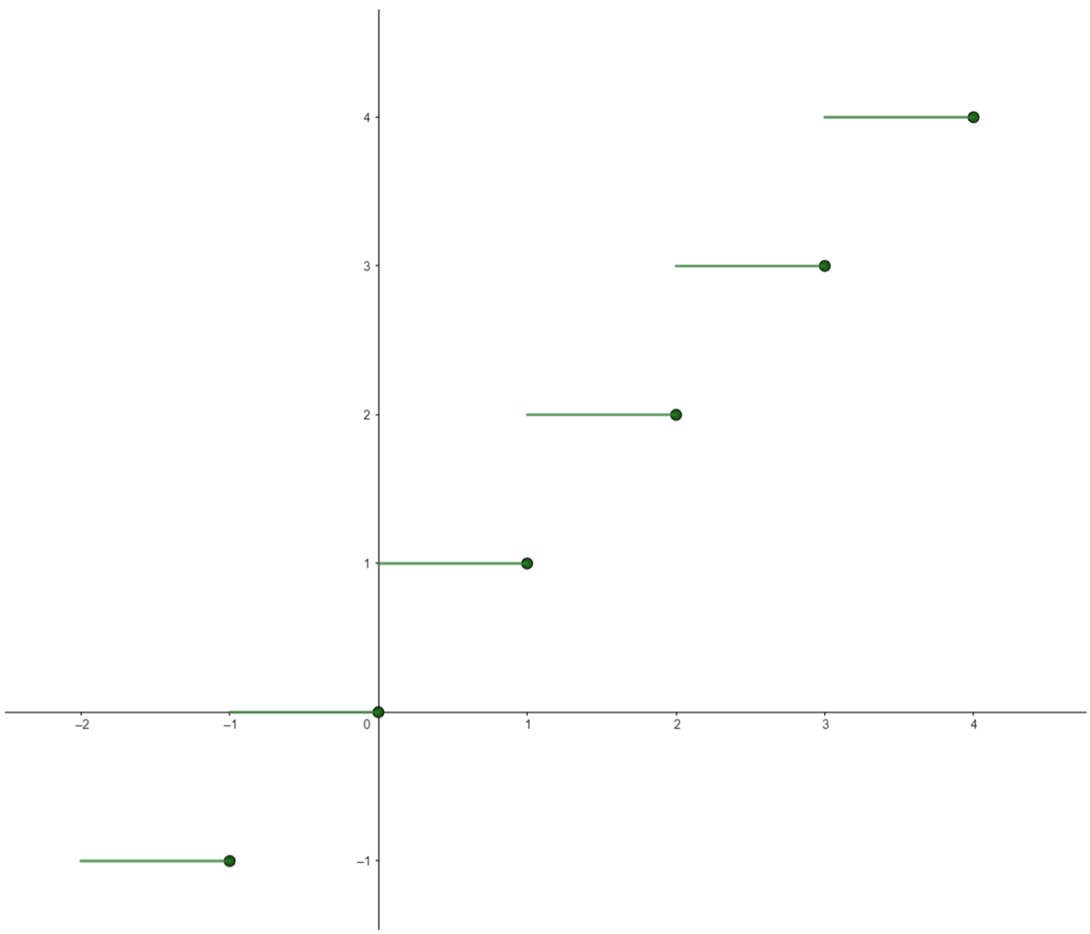


Figure 1.2.2. The ceiling function.

Another useful function is the ***sign*** function.
This is defined by

$$\operatorname{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}.$$

Another is the ***delta function***, defined by $\delta(x) = 1$ when $x = 0$ and 0 when $x \neq 0$.

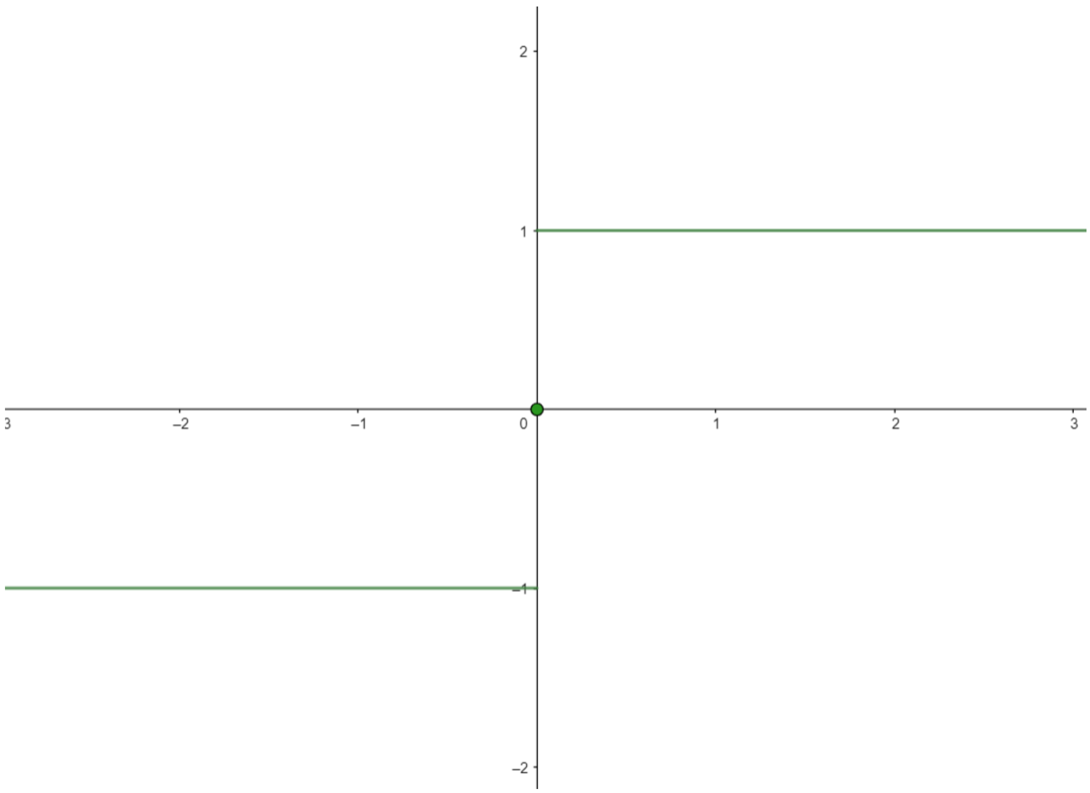


Figure 1.2.3. The sign function.

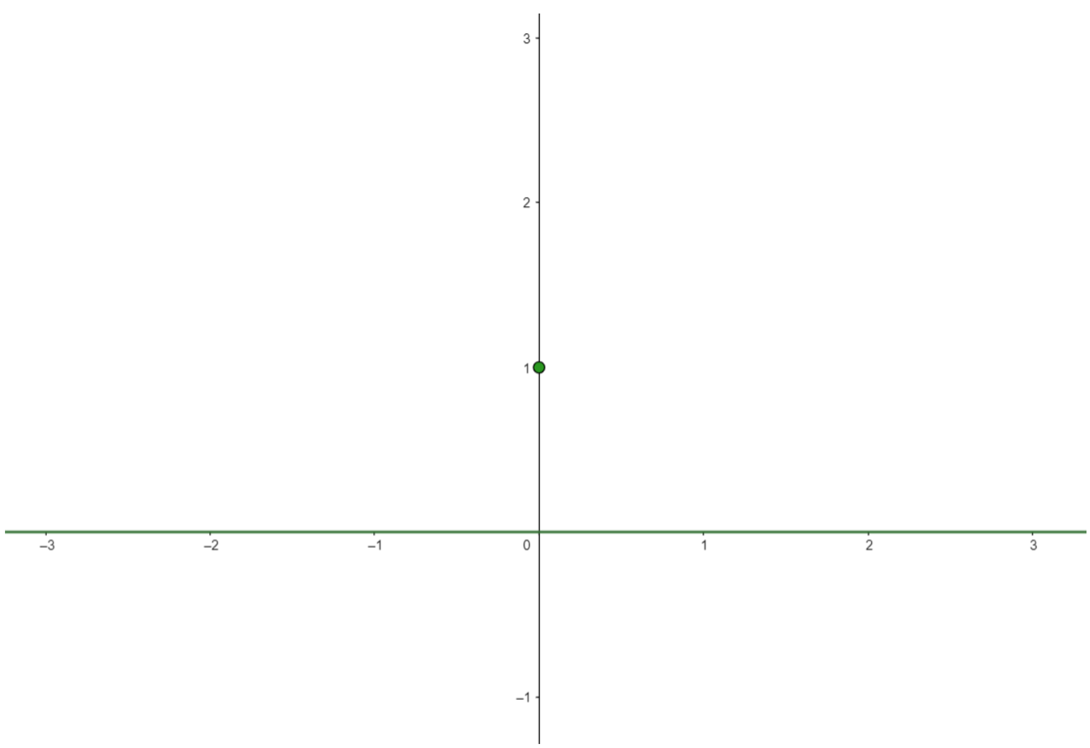


Figure 1.2.4. The delta function.

Checkpoint 12.

1.3 Properties of functions

We now list certain properties of functions to which we will often refer.

A function f is said to be ***odd*** if $f(-x) = -f(x)$ for all x in the domain of f. (It is unclear what this would mean if the domain contains x but not $-x$.) Similarly an ***even*** function f is one satisfying $f(-x) = f(x)$.

Checkpoint 13.

A function f is said to be ***increasing*** if $f(x) \leq f(y)$ for all values of x and y in the domain of f such that $x < y$. Informally, the value of an increasing function gets bigger if the argument gets bigger. If you change the requirement that $f(x) \leq f(y)$ to the strict inequality $f(x) < f(y)$, this defines the notion of ***strictly increasing***. ***Decreasing*** and ***strictly decreasing*** functions are defined analogously but with one inequality reversed: f is ***decreasing*** if $f(x) \geq f(y)$ for all x, y satisfying $x < y$.

Checkpoint 14.

We can also say when a function is increasing or decreasing on a part of the domain: f is increasing on the open interval (a, b) if the above inequality holds for all $x, y \in (a, b)$. For any point $c \in (a, b)$, we then also say that f is increasing at c. In other words, to say f is increasing at a point c means there is some $a < c < b$ such that f is increasing on the open interval (a, b).

Why must the interval be open? Here is an example showing why. Let $f(x) := x^2$ and consider the closed interval $[0, 1]$. For every $x < y$ in this closed interval $f(x) < f(y)$. If we are allowed to conclude f is strictly increasing at

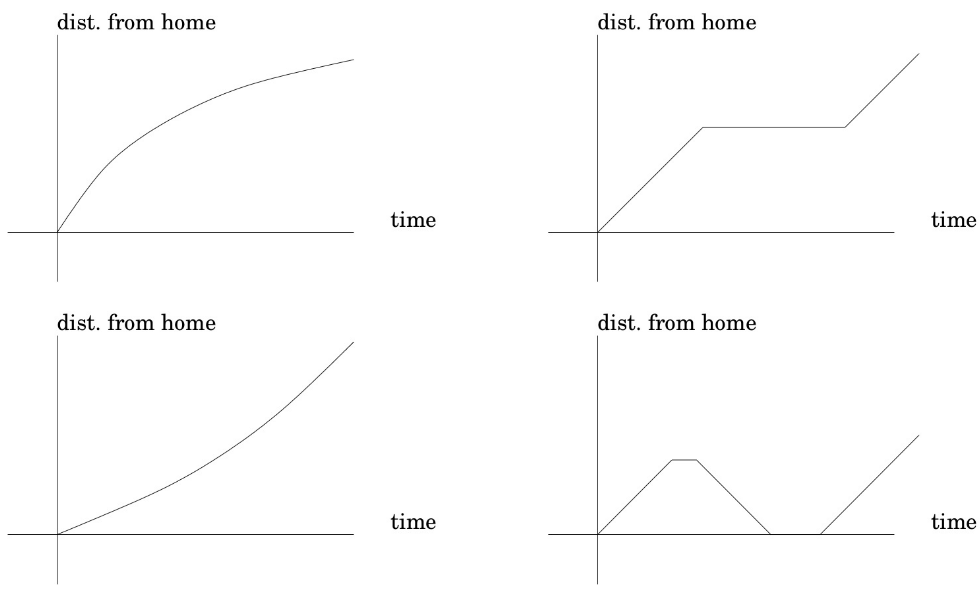
Checkpoint 15.

1.4 Graphing

As you already know, points in the plane can be labeled by ordered pairs of real numbers. As you also already know, the graph of a function f is the set points in the plane corresponding to the ordered pairs $\{(x, f(x)) : x \in \text{domain of } f\}$.

Often the graph of a function is a continuous curve, and can be quickly drawn, conveying essential information about f to the eye much more efficiently than if the reader had to wade through equations or set notation.

The following Checkpoint is about the following graphs, borrowed from Hughes-Hallet et al.:



Checkpoint 16.

Some conventions make graphs even more effective at conveying information. The axes should be labeled (more on that later) but more importantly, marked so that the scale is clear. Rather than just mark where 1 is on the horizontal and vertical axes, it is often helpful to mark any value where something interesting is going on: a discontuity, an asymptote, a maximum, or a change of cases for functions defined in cases. For example, if we graph $x \mapsto 1/(x^2 - 3x + 2)$, we should mark vertical asymptotes (a certain kind of discontinuity) on the x-axis at $x = 1$ and $x = 2$; a dashed vertical line is customary. We should mark a local maximum of -4 (marked on the y-axis) occuring at $x = 3/2$ (marked on the x-axis). Another way to do this would be to label and mark the point $(3/2, -4)$ on the graph. There is a horizontal asymptote at zero, which we would mark with a dashed horizontal line if it occurred anywhere else, but we don't because it is hidden by the x-axis. When graphing a function on the entire real line, we can't go to infinity and stay in scale, so we either go out of scale or draw a finite portion, large enough to given the idea. Choosing the latter, the resulting picture should look something like the graph in Figure 1.4.1.

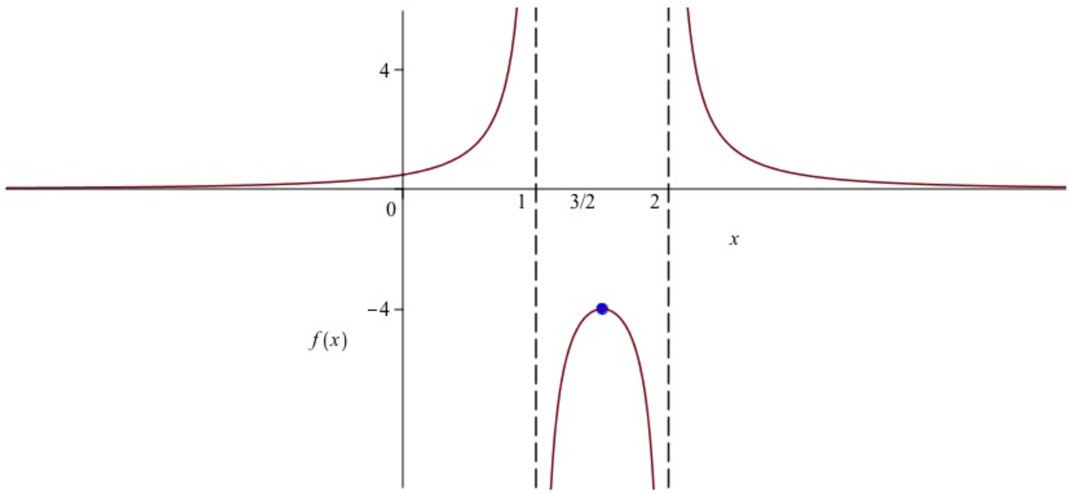


Figure 1.4.1. graph of $f(x) := 1/(x^2 - 3x + 2)$

Here follows a list of tips on graphing an unfamiliar function, call it f. The last three tips on shifting and scaling are ones we have found in the past that many students vaguely recall but get wrong, so please make sure you know them.

1. Is the domain all real numbers? If not, what is it? If the function has a piecewise definition, try drawing each piece separately.
2. Is there an obvious symmetry? If $f(-x) = f(x)$ for all x in the domain, then f is even and there is a symmetry about the y-axis. If $f(-x) = -f(x)$ then f is odd and there is 180-degree rotational symmetry about the origin.
3. Are there discontinuities, and if so, where? Are there asymptotes?
4. Try values of the function near the discontinuities to get an idea of the shape -- these are particularly important places. If the domain includes oints on both sides of a discontinuity be sure to test points on each side.
5. Try computing some easy points. Often $f(0)$ or $f(1)$ is easy to compute. Trig functions are easily evaluated at certain multiples of π.
6. Where is f positive?
7. Where is f increasing and where is it decreasing? This will be easier once you know some calculus.

8. Where is f concave upward versus concave downward? This will be a lot easier once you know some calculus.
9. Where are the maxima and minima of f and what are its values there? This will be a lot easier once you know some calculus.
10. What does f do as x approaches ∞ and $-\infty$?
11. Is there a function you understand better than f which is close enough to f that their graphs look similar?
12. Is f periodic? Most combinations of trig functions will be periodic.
13. Is the graph of f a shift of a more familiar graph? Graphing $y = f(x) + c$ shifts the graph up by c; this is pretty intuitive; if c is negative the graph shifts downward. Graphing $y = f(x + c)$ shifts the graph left or right by c. If c is positive, the graph shifts *left*.
14. Is the graph of f a rescaling of a more familiar graph? The graph $y = cf(x)$ stretches vertically by a factor of c. When $c < 1$ this is a shrink rather than a stretch.
15. The graph of $y = f(cx)$ stetches or shrinks in the horizontal direction. When $c > 1$, it is a shrink. Why? Try sketching $y = \cos x$ and on top of this sketch $y = \cos(2x)$.

Checkpoint 17. enn.edu/~ancoop/1070/section-1.html#project-17)

Checkpoint 18. enn.edu/~ancoop/1070/section-1.html#project-18)

Σ () MATH 1070: Mathematics of Change, part I (article-1.html)

Calculus Group

1 Units, proportionality and mathematical modeling

2.1 Physical units and formulas

One skill we'll need is writing formulas for functions given by verbal descriptions. Try this multiple choice question before going on.

Checkpoint 19.

Here are some more helpful facts about units.

1. You can't add or subtract quantities unless they have the same units. That would be like adding apples and oranges!
2. Multiplying (resp. dividing) quantities multiplies (resp. divides) the units.
3. Taking a power raises the units to that power. For example, if x is in units of length, say centimeters, then $3x^2$ will have units of area, in this case square centimeters. Most functions other than powers require unitless quantites for their input. For example, in a formula $y = e^{***}$ the quantity *** must be unitless. The same is true of logarithms and trig functions: their arguments are always unitless.
4. Units tell you how a quantity transforms under scale changes. For example a square inch is 2.54^2 times as big as a square centimeter.

Checkpoint 20.

Often what we can easily tell about a function is that it is proportional to some combination of other quantities, where the ***constant of proportionality*** may or may not be known, or may vary from one version of the problem to another. Constants of proportionality have units, which may be computed from the fact that both sides of an equation must have the same units.

Example 2.1.1. If the monetization of a social networking app is proportional to the square of the number of subscribers (this representing perhaps the amount of messaging going on) then one might write $M = kN^2$ where M is monetization, N is number of subscribers and k is the constant of proportionality. You should always give units for such constants. They can be deduced from the units of everything else. The units of N are people and the units of M are dollars, so k is in dollars per square person. You can write the constant as $k \frac{\$}{\text{person}^2}$.

To say A is ***inversely proportional*** to B means that A is proportional to $1/B$. If a quantity A is proportional to both quantities B and C,which can vary independently, then A must be proportional to $B \cdot C$,so $A = k\,b\,C$ for some constant of proportionality, k.

Example 2.1.2. If the expected profit on a home sale is proportional to the assessed value of the home and inversely proportional to the number of days it has been on the market, we could capture that relation as $P = kV/T$ where P is profit in dollars, V is assessed price in dollars, T is number of days on the market, and k is a constant of proportionality.

Checkpoint 21. ()enn.edu/~ancoop/1070/sec-modeling.html#project-21)

Warning 2.1.3. Sometimes in mathematical modeling, an equation represents an empirical law, which is a rough fit to some function. For example, energy loss due to the nature of alternating current is said to be inversely proportional to the 0.6 power of the frequency, but this is simply the best fit to data, not due to electromagnetic theory. In that case, units will not make sense. For example, if it is observed that the blood volume of small mammals is roughly proportional to the 2.65 power of the mammal's length, sensible units will not be assignable to the proportionality constant k in the formula $BV = kL^{2.65}$. In this case we just have to live with the fact that k has units involving fractional powers of length that won't make much sense outside of this context.

An important point when writing up your work: You don't just write $M = kN^2$ without stating the interpretations of the three variables. Also, there would not usually be a $:=$ here, because you are not defining the function $M(N) := kN^2$ as much as you are saying that two observed quantities M and N vary together in a way that satisfies the equation $M = kN^2$. There isn't a clear line here, but the style of the definition can be important in conveying to the reader what's going on.

Example 2.1.4. The present value under constant discounting is given by $V(t) = V_0 e^{-\alpha t}$ where V_0 is the initial value and α is the discount rate. What are the units of α? They have to be inverse time units because αt must be unitless. A typical discount rate is 2% per year. You could say that as "0.02 inverse years." We hope that by the end of the semester, the notion of an inverse year is somewhat intuitive.

Checkpoint 22. ()enn.edu/~ancoop/1070/sec-modeling.html#project-22)

Often quantities are measured as proportions. For example, the proportional increase in sales is the change is sales divided by sales. In an equation: the proportional increase in S is $\Delta S/S$. Here, ΔS is the difference between the new and old values of S. You can subtract because both have the same units (sales), so ΔS has units of sales as well. That makes the proportional increase unitless. In fact proportions are always unitless.

Percentage increases are always unitless. In fact they are proportional increases multiplied by 100. Thus if the proportional increase is 0.183,the percentage increase is 18.3%. In this class we aren't going to be picky about proportion versus percentage. If you say thve percentage increase is 0.183 or the proportional change is 18.3%,everyone will know exactly what you mean. But you may as well be precise.

Checkpoint 23. ()enn.edu/~ancoop/1070/sec-modeling.html#project-23)

Units behave predictably under differentiation and integration as well. We will refer back to this when we define the relevant concepts, but you may as well see a preview now. The derivative $(d/dx)f$ has units of f divided by units of x. You can see this easily on the graph in Figure 2.1.5 because the derivative is a limit of rise over run, where rise has units of f and run has units of x. The integral $\int f(x)\, dx$ has units of f times units of x. Again you can see it from a picture (Figure 2.1.6), because the integral is an area under a graph where the y-axis has units of f and the x-axis has units of, well, x.

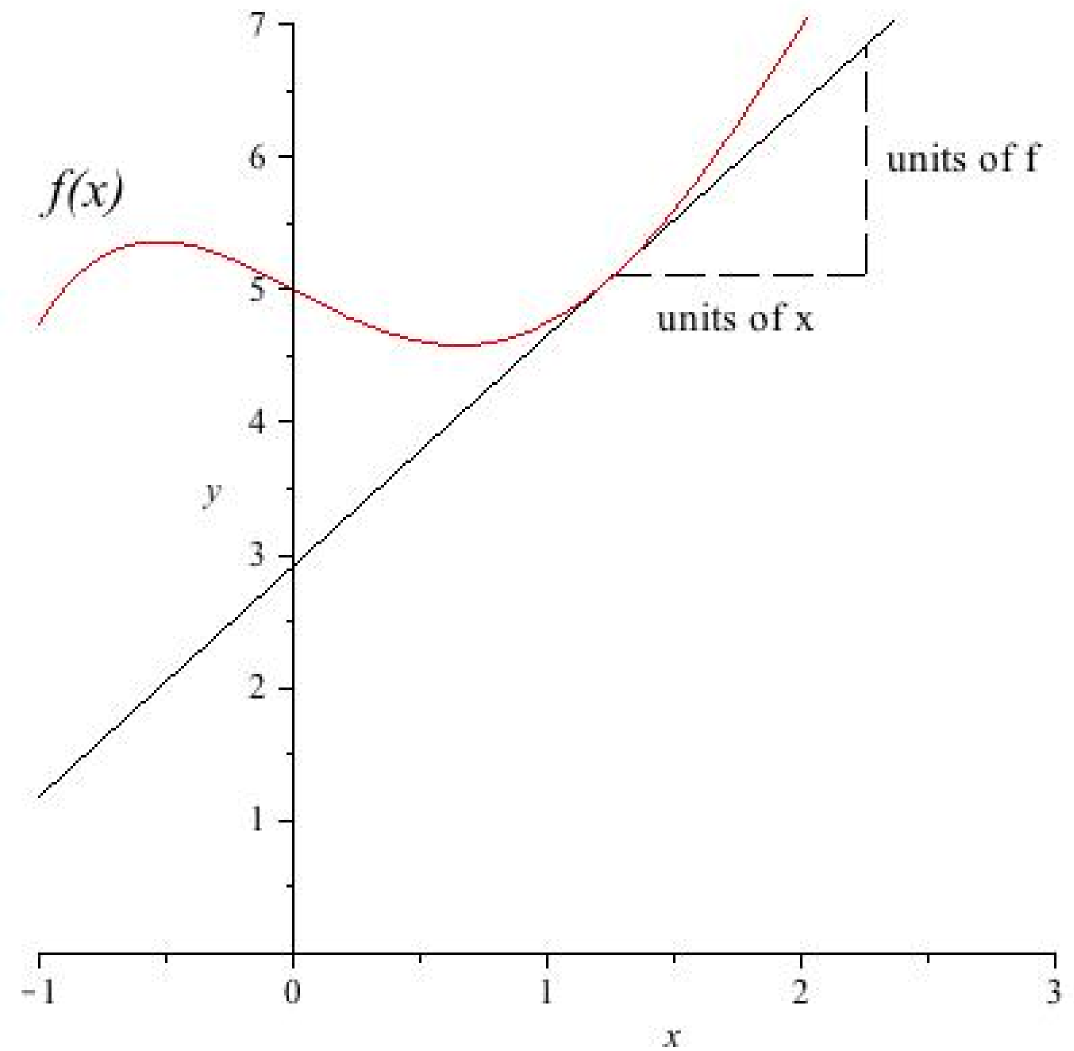


Figure 2.1.5. units of the derivative

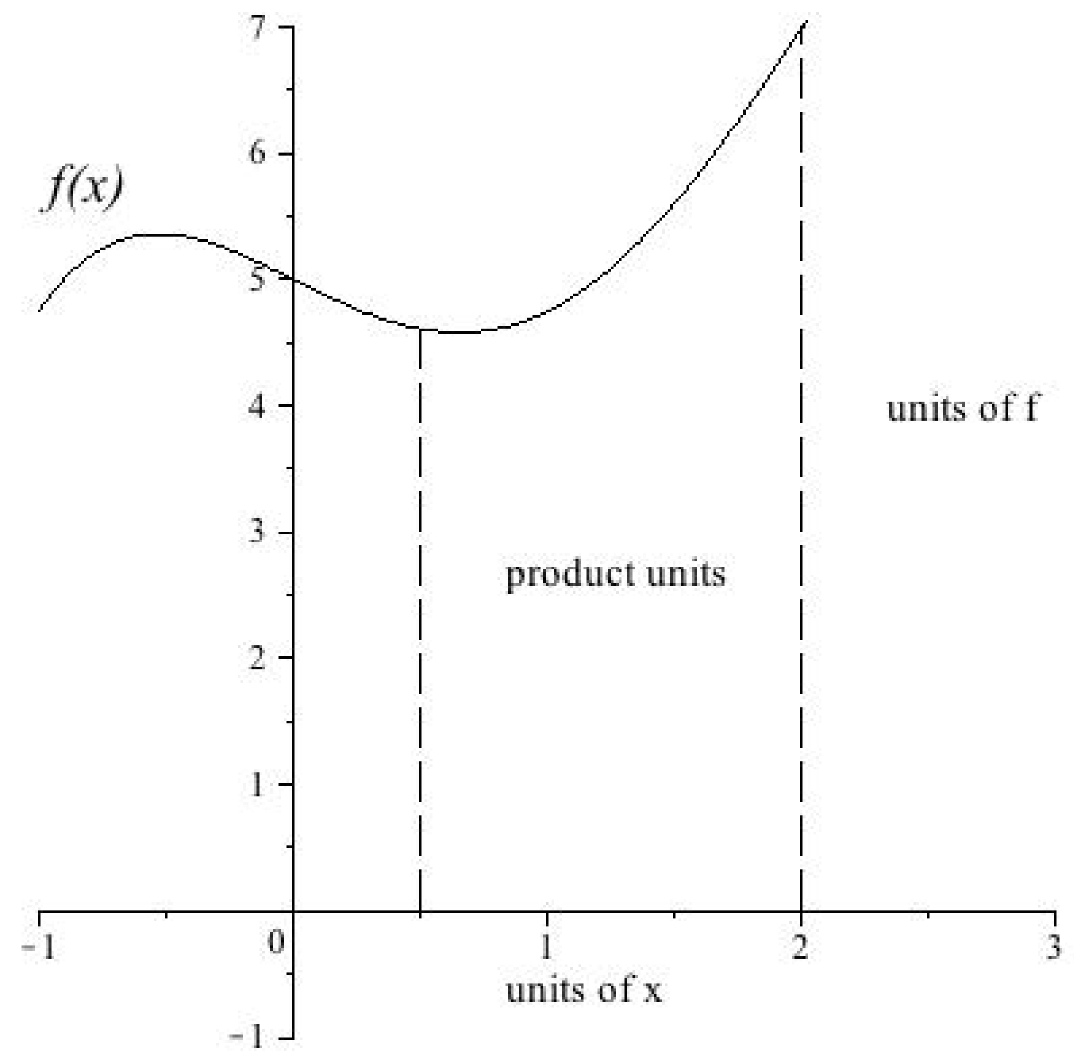


Figure 2.1.6. units of the integral

2.2 Modeling

Mathematical modeling means writing down mathematics corresponding to a given real-world (physical, chemical, biological) scenario, along with equations or other relations that could be expected to hold, at least approximately or under further assumptions.

Unpacking this, we see a number of features. First of all one must define mathematical objects in the model: variables, sets, functions, equations and so forth. Secondly, one must give ***interpretations*** of everything in the model. An interpretation tells what physical quantity is associated with each of the constants and variables and what relation is meant by each function. Real-world quantities include units, so this part always involves stating units.

Thirdly, often one needs to add ***assumptions*** about the scenario. These say the circumstances under which would you expect the mathematics to be correct for the model. These assumptions are determined by the real world; they are not mathematical. Lastly, if there are questions given in the scenario, it is necessary to say what part of the mathematics answers the question(s). After this, what is left is a math problem: solve for the quantities that answer the questions.

Modeling assumptions are the sorts of things you'd ask a biologist about (if you're modeling a biological phenomenon), an economist about (if you're modeling an economic phenomenon), etc.

In the following example, we have underlined parts of the modeling exercise that reflect the outline we have given, such as naming of variables, interpretation, units and hypotheses.

Example 2.2.1.

Scenario

Galileo observes that objects falling a short time seem to fall a distance that was proportional to the square of the time and independent of the object: 4 feet for an object falling half a second, 9 feet after three quarters of a second, 16 feet after one second, and so forth. Galileo decides to measure the Tower of Pisa by dropping a stone from the top of the tower and measuring the time it takes for him to hear it hit the ground. Make a model for this and use it to estimate the elapsed time Galileo measure between dropping the stone and hearing the sound.

Model

Let $f(t)$ be the distance *in feet* that an object falls in t *seconds*, starting from rest. The wording of the scenario tells us that $f(t) = c\,t^2$ where c has units of *feet per seconds squared*. This assumes we *set* $t = 0$ *at the time of release* and measure distance *from the point of release*. We are asked to determine t such that $f(t) = h$, where h is the height of the Tower of Pisa. Equivalently, we need to find $f^{-1}(h)$. We assume that the model is accurate. What that means in this case is that we can ignore things such as ***air resistance*** and the ***time lag*** for the sound of impact to get back to Galileo's ear.

Solution:

We look up the height of the Tower of Pisa to find that $h = 186$ feet. We solve for c given Galileo's data for small distances and find that $c = 16$ (for example: $f(1/2) = c(1/2)^2 = 4$ implies $c = 16$). We can solve directly for $16t^2 = 186$ or we can compute the inverse function to f yielding $f^{-1}(x) = \sqrt{x/16}$ and substitute 186 for x. Either way we get $t = \sqrt{186/16} \approx 3.40$. It may sound pedantic, but probably we should justify our choice of the positive square root by saying the whole experiment only covers time after the release, that is, $t \geq 0$.

Were the physical assumptions warranted? Many objects would be slowed by air resistance over such a distance. Probably Galileo would have had to drop something like a rock in order for the fall not to have been significantly slowed. Looking up the speed of sound, it would take an extra $1/4$ second to register the sound. Probably Galileo could measure time to within greater accuracy that $1/4$ second, so this assumption is definitely shaky.

Checkpoint 24.

Using units to interpret a model. Sometimes we're given (or derive) a formula whose meaning isn't immediately clear; often the units can be our guide here. As an example, consider the *body mass index* (BMI), which medical researchers and doctors frequently use as a measure of obesity. BMI was invented by Adolphe Quetelet, a Belgian statistician, in the mid-1800s; it's still used today. (Think of how few medical tools from that period are still in use. . .)

A patient's BMI is computed by the following formula:

$$\text{BMI} = \frac{\text{mass}}{(\text{height})^2}$$

If we measure mass in kilograms and height in meters, the units of BMI are kg/m^2. That doesn't seem very revelatory.

But now consider: for objects of uniform density (which humans aren't, but close enough), we have

$$\text{mass} = \text{density} \times \text{volume}$$

Substituting, we get

$$\text{BMI} = \text{density} \times \frac{\text{volume}}{(\text{height})^2}$$

Now, broadly speaking, most people have approximately the same density. So when we compare two patients' BMIs, we're really comparing the quantity

$$\frac{\text{volume}}{(\text{height})^2}$$

which has units $m^3/m^2 = m$. That is, roughly speaking BMI measures a *length*! But the length of what?

Here's were we need another rough biological fact: most peoples' width is proportional to their height; that is,

$$\text{width} = k \times \text{height}$$

for some constant of proportionality k. If we pretend that a person's shape is a rectangular prism, then the volume of this prism would be

$$\text{volume} = \text{width} \times \text{height} \times \text{depth}$$

Because width is proportional to height, that works out to approximately

$$\text{volume} = k(\text{height})^2 \times \text{depth}$$

so that the quantity $\frac{\text{volume}}{(\text{height})^2}$ is really

$$\frac{\text{volume}}{(\text{height})^2} = \frac{k \times (\text{height})^2 \times \text{depth}}{(\text{height})^2} = k \times \text{depth}$$

That is to say, what BMI really measures is how deep -- perhaps better to say, how thick -- a person is.

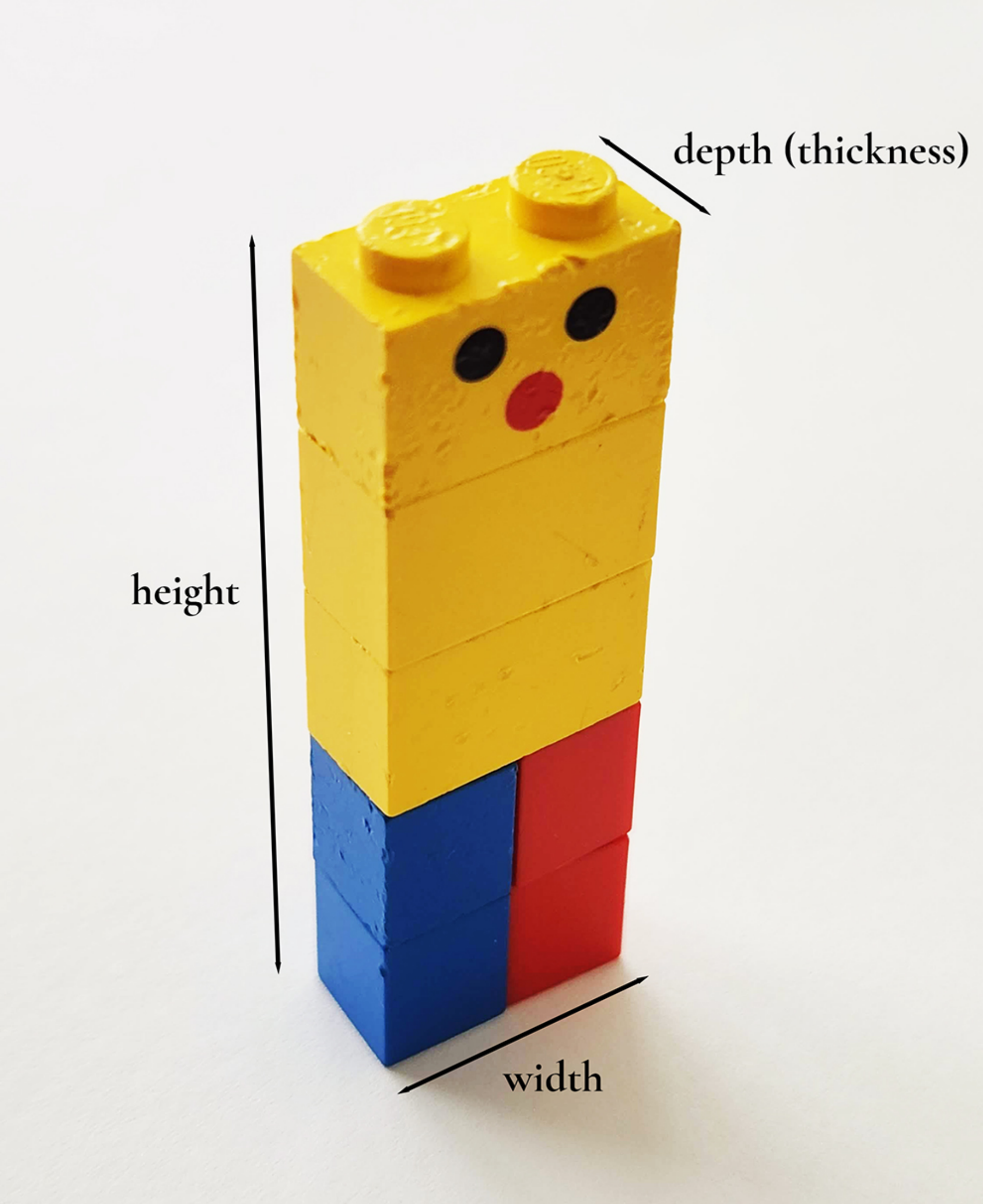


Figure 2.2.2. BMI assumes the body is basically a rectangular prism.

Checkpoint 25.

2.3 Inverse functions

One method of solving the problem of Galileo's experiment involved an inverse function. Let's be explicit about the definitions involved. The inverse function of a function f is the function that answers the question,

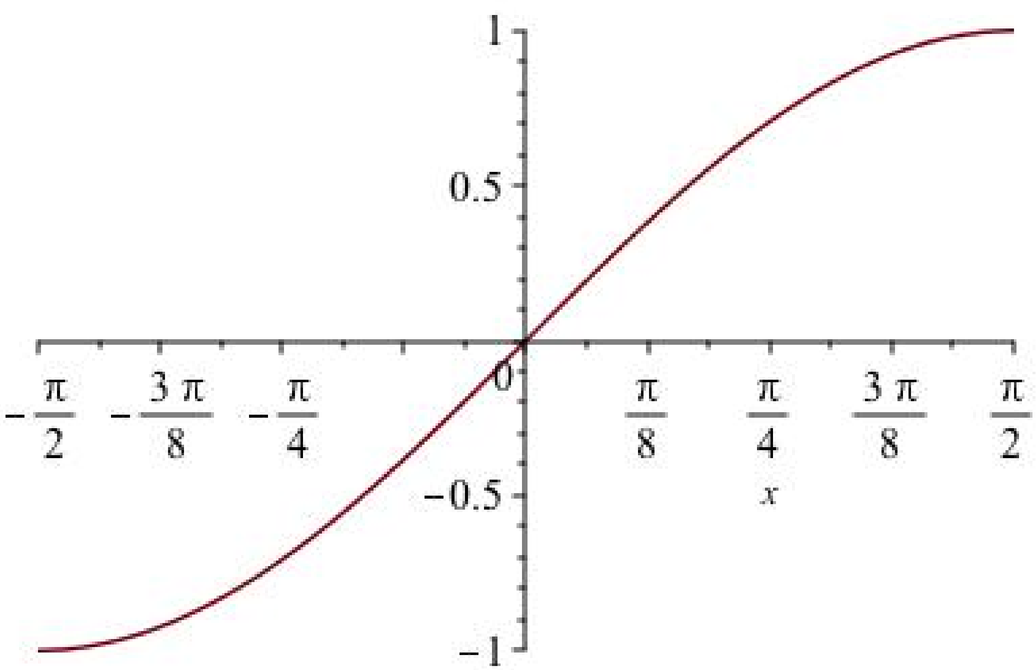
> *What input do I need to get the given output?*

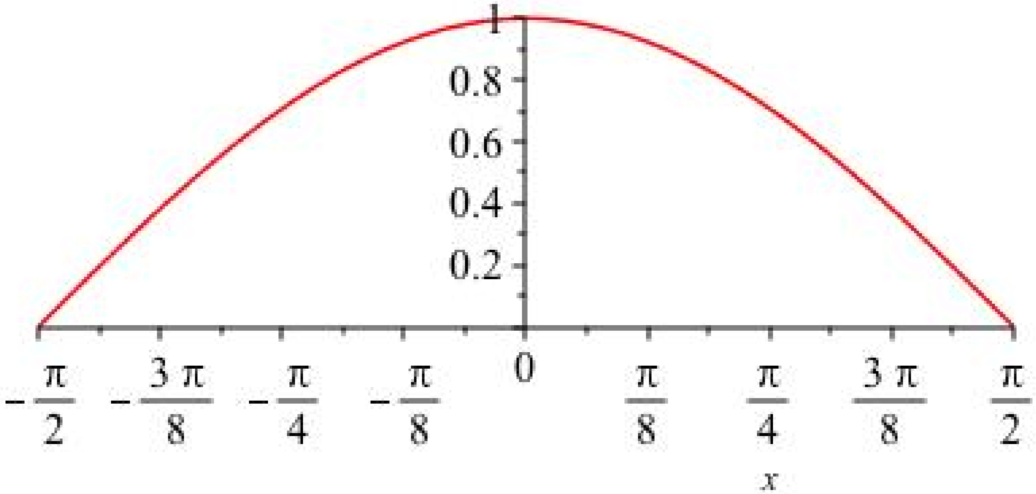
In other words, if g is the inverse function of f then $g(y)$ is whatever value x satisfies $f(x) = y$. If there is more than one answer to this, then f has no inverse function; however, you can usually restrict the domain so there is only one answer. If there is no answer, that's not a problem, it just means that y is not in the domain of g. This happens when y is not in the range of f. Thus, the domain of g is the range of f. Likewise, the range of g is any possible answer to the question above, therefore any x in the domain of f.

Checkpoint 26.

The usual notation for the inverse function to f is f^{-1}. This is terrible notation because it is the same as the notation for the -1 power of f,also known as $1/f$. We tried changing the inverse function notation to f^{inv} for the purposes of this class, but then students were confused when they saw f^{-1}. We will stick with the terrible notation, and mention it when confusion might arise.

There is a standard way that the domain is restricted on trig functions so that the inverse function can be defined. For $\sin$ and $\tan$ it is $[-\pi/2, \pi/2]$. The function $\cos$ when restricted to $[-\pi/2, \pi/2]$ is not one-to-one; the standard choice for $\cos$ is $[0, \pi]$. These are arbitrary conventions, but are probably built in to your calculator, so we had better adopt them. Also, along with $\sin^{-1}$,$\cos^{-1}$ and $\tan^{-1}$,the conventional names $\arcsin$, $\arccos$ and $\arctan$ are also used.





1
0.8
0.6
0.4
0.2
0
-π/2
-3π/8
-π/4
-π/8
π/8
π/4
3π/8
π/2
x

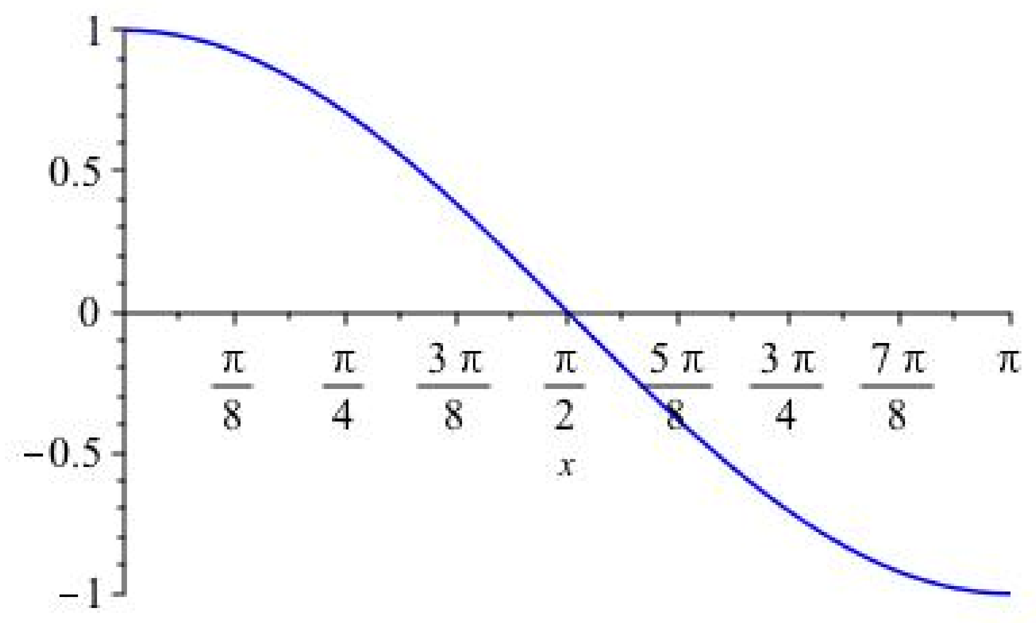


Figure 2.3.1. $\sin$ is one-to-one on $[-\pi/2, \pi/2]$ (top) but $\cos$ is not (left) so we move the window to $[0, \pi]$ (right)

Checkpoint 27.

Inverse functions occur naturally in mathematical modeling. For example, if $f(t)$ represents how many miles you can walk in t hours, then $f^{-1}(x)$ represents how many hours it takes you to walk x miles. Note that in this explanation, x is a bound variable; we could have used any other name, such as t again, only it helps readability if we use names such as t for time and x for distance.

Checkpoint 28.

Remark 2.3.2.

1. The concept of an inverse function appears to be harder than most people realize. For example, a number of years ago, calculus students were given the problem to compute a formula for the inverse function to $\sinh(x)$,the hyperbolic sine function. This was already not easy: (only half the students got it) but then we asked them to find a number u such that $\sinh(u) = 1$. Almost no one got it, despite that fact that this was supposed to be the easy part -- they just needed to plug in 1 to their inverse sinh function! By definition, $\sinh^{-1}(1)$ is a number u such that $\sinh(u) = 1$. The moral of the story is, don't lose sight of the meaning of an inverse function when doing computations with them.

2. How does the graph of an inverse function relate to the graph of a function?
   The roles of x and y have switched. When the first and second coordinate of an ordered pair are switched, the point reflects across the diagonal line $y=x$. Thus, the graph of the inverse function is the original graph (on the appropriate domain) reflected across the diagonal.

Logarithm Cheat Sheet. The inverse function to $f(x):=e^x$ is called the ***nautral logarithm*** $\ln$. All that means is: the equations

$$y=e^x$$

and

$$\ln y = x$$

mean exactly the same thing. Logarithms look scary, but all you have to remember is: when you see a logarithm on one side, just convert the whole equation to exponentials.

It may be useful to have some approximate values of $\ln$ in mind so that the function seems less mysterious. The following values are accurate to about 1%:

$$\begin{aligned} e &\approx 2.7 \\ \ln(2) &\approx 0.7 \\ \ln(10) &\approx 2.3 \\ \log_{10}(2) &\approx 0.3 \\ \log_{10}(3) &\approx 0.48 \\ e^3 &\approx 20 \end{aligned}$$

Checkpoint 29.

In case you need them, here are some other useful quantities to within 1%:

$$\begin{aligned} \pi &\approx \frac{22}{7} \\ \sqrt{10} &\approx \pi \\ \sqrt{2} &\approx 1.4 \\ \sqrt{1/2} &\approx 0.7 \\ e^8 &\approx 3000 \end{aligned}$$

Also useful sometimes: $\sqrt{3}=1.732\ldots$ and $\sqrt{5}=2.236\ldots$ both to within about 0.003%.

Checkpoint 30.

Squares and Powers of 2 Cheat Sheet. If you know the powers of 2 you can do the same thing with $\log_2$ that you can do with $\log_{10}$. It will be helpful for you be at least somewhat familiar with them -- for example, to recognize them on sight. You should also recognize the first twenty squares:

$$1,4,9,16,25,36,49,64,81,100,121,144,169,196,225,256,289,324,361,400\,.$$

No kidding, when you come across one of these numbers under a radical, you know immediately it can be factored out.

Here are the powers of 2.

$$
\begin{aligned}
2^0 &= 1\\
2^1 &= 2\\
2^2 &= 4\\
2^3 &= 8\\
2^4 &= 16\\
2^5 &= 32\\
2^6 &= 64\\
2^7 &= 128\\
2^8 &= 256\\
2^9 &= 512\\
2^{10} &= 1,024\\
2^{11} &= 2,048\\
2^{12} &= 4,096\\
2^{13} &= 8,192\\
2^{14} &= 16,384\\
2^{15} &= 32,768\\
2^{16} &= 65,536\\
2^{20} &\approx 1,000,000\\
2^{30} &\approx 1,000,000,000\\
2^{100} &\approx 10^{30}
\end{aligned}
$$

2.4 Exponential and logarithmic relationships

Taking a break from heady stuff, let's do some quick computations with logs.

The log cheatsheet is there to encourage you to use logs for quick computations. The squares and powers of two are just for fun (OK it was written by geeks). We're going to take a quick break from concepts to get the hang of computing with logs.

Example 2.4.1. What is the probability of getting all sixes when rolling 10 six-sided dice? It's 1 in 6^{10} but how big is that? If we use base-10 logs, we see that

$$\log_{10}(6^{10}) = 10\log_{10}6 = 10(\log_{10}(2) + \log_{10}(3)) \approx 10(.78) = 7.8.$$

So the number we're looking for is approximately $10^{7.8}$ which is $10^7 \times 10^{0.8}$ or $10,000,000$ times a shade over $10^{.78}$,this latter quantity being very close to 6 according to the one-digit logs you computed. So we're looking at a little over sixty million to one odds against.

Checkpoint 31.

These are not just random examples, it is always the best way to get a quick idea of the size of a large power. When the base is 10 we already know how many digits is has, but when the base is something else, we quickly compute $\log_{10}(b^a) = a \cdot \log_{10}(b)$.

Example 2.4.2. Why is the value $\ln(10) \approx 2.3$ on your log cheatsheet so important? It converts back and forth between natural and base-10 logs. Remember, $\log_{10} x = \ln x / \ln 10$. Thus the constant $\ln 10$ is an important conversion constant that just happens to be closer than it looks (the actual value is $2.302\ldots$). So for example,

$$e^8 \approx 10^{8/2.3} \approx 10^{3.5} = 1000 \times 10^{0.5} \approx 3,000\,.$$

Checkpoint 32.

Recall in the definition of e,the slope of the graph e^x at $(0,1)$ is 1, therefore the tangent line approximation is

$$e^x \approx 1 + x\,.$$

This approximation is very good when $x < 0.1$. Converting to logs gives the version

$$x \approx \ln(1+x)\,,$$

which the following example illustrates the utility of.

Example 2.4.3. Suppose, your company grows in value by 6% each year for 20 years. By what factor C does the value increase over this time? The answer is 1.06^{20},but about how big is that? For a quick answer, take logs. Using the fact that $\ln 1.06 \approx 0.06$ (that's the logarithmic version of our estimate, with $x = .06$), we see that $\ln C = \ln(1.06^{20}) \approx 20 \times 0.06 = 1.2$. We'd rather have this in base ten, so we compute

$$\log_{10} C = \ln C / \ln 10 \approx \ln C/2.3 \approx 1.2/2.3 \approx 0.5\,,$$

maybe a little bigger like 0.52 or so. Looking at the log cheatsheet shows this means C should be between 3 and 4, somewhat closer to 3. In fact to two significant figures, the growth factor is 3.2.

Checkpoint 33.

The multiplicative frame. If you ask someone to state a relationship between the numbers 20 and 30, the most common answer is that 30 is ten more than 20. A more fundamental answer is that 30 is 50% more, or equivalently that 30 is three halves of 20. The section on proportionality is designed to emphasize multipicative thinking over additive thinking. ***Additive thinking is more common only because we find it computationally easier to add than to multiply.*** Saying that multiplicative thinking is more fundamental is not a precise mathematical statement, so there's no way to prove it. One reason to believe it is that mulitplicative statements remain the same no matter what units you use (as long as the 30 and the 20 are in the same units).

Checkpoint 34.

Exponentials and logarithms are built to express multiplicative facts. In fact the additive laws of exponentiation and logarithms basically convert multiplicative facts to additive facts, thereby converting the more fundamental fact to the type you can compute more easily.

Much of what you learn on topic of exponential and logarithmic relationships involves questions such as:

If you observe that $\ln x$ has increased by about 0.7, what does this mean about the increase that has occurred in x?

A tip about setting up equations representing functional relationships: when a quantity has different values at different times such as "before" and "after", using one variable to represent both quantities can lead to mess and confusion. Better to use different names such as x_1 and x_2, or x_{init} and x_{final}, or possibly x and x', etc. Using this idea on the question above sets up an equation like this: $\ln x_2 \approx \ln x_1 + 0.7$. From here, exponentiating leads to

$$x_2 \approx e^{\ln x_1 + 0.7} = x_1 \cdot e^{0.7} \approx 2x_1 .$$

So, if you observe $\ln x$ increasing by about 0.7, you will know that x had approximately doubled. This is what it means that logarithms transfer multiplicative scales to additive ones. A multiplicative relation such as doubling transfers to an additive relation, namely addition of about 0.7.

Checkpoint 35.

One more thing to keep in mind about logarithms and exponentials is that they do not scale with units. If we change the units of x from inches to centimeters, and if $y = e^x$, then in the new units $y' = e^{2.54x'} = y^{2.54}$. The new exponential appears to be the old one to the 2.54 power. What does that even mean? It is a tipoff that x should not be exponentiated: anything other than a unitless constant is likely to be meaningless when exponentiated. The same is true for logarithms and trig functions.

If $\log x$ increases at a constant additive rate, then x increases at a constant multiplicative rate. What does this mean?

If a quantity Q increases at a constant additive rate, it means that if you wait one unit of time, Q always increases by the same additive amount. In fact, between any two times s and t the increase will be $c(t-s)$.

Checkpoint 36.

If a quantity Q increases at a constant *multiplicative* rate, it means waiting one unit of time always multiples Q by the same amount, and in general, between times s and t, the factor by which Q increases will be c^{t-s} where c is the factor by which Q increases in one unit of time.

Checkpoint 37.

To get back to the question of what it means about logs relating additive to multiplicative growth, if $\log x = a + bt$ (constant additive growth over time) then $x = e^{a+bt} = e^a e^{bt} = AB^t$ where $A = e^a$ and $B = e^b$. This is constant multiplicative growth.

Constant multiplicative growth rates occur in a lot of applications. This is also called *exponential growth* because the formula for a quantity growing multiplicatively is Ae^{bt} (also e^{a+bt} or AB^t). When $b < 0$, it is called exponential ***decay*** or ***decrease***.

Here are a few examples.

Equilibrating temperatures If an item is hotter or colder than its environment then the temperature difference between the object and its environment, as a function of time, decreases exponentially (here, in Ae^{bt},the coefficient b is negative).

Accumulating money At a fixed interest rate, money grows exponentially. So, unfortunately does debt (just put a minus sign on the money).

Population Populations tend to grow this way (again unfortunately, in most cases).

Radioactive substances decay exponentially.

DNA The portion of DNA remaining unmutated decays exponentially.

Present value analysis If there is a fixed discount rate, there is exponential decrease of the present value for revenue at future times.

Time series data Not always, but often, the correlation decays exponentially.

If we get to assume a nice clean exponential model, and can observe at more than one time point, then exponential growth/decay models are nearly as easy to solve as linear growth models (a highlight of eighth grade math). You should learn this both conceptually and as a mindless skill. In Math 104, this is taught as an elementary example of differential equations. We'll get to those next semester, but really this kind of growth/decay model is far more fundamental and should be discussed now. Here's an example of how the computation goes.

Example 2.4.4. A viral infection is spreading exponentially through the community. On the first day that the outbreak had a name, there were 25 infections. A week later there were 40 infections. How many infections will there be in another two weeks? When will the number of infections reach 200,000, which is the size of the entire local population?

Solution #1 (plug in logs) Let $N(t)$ denote the number of infections after t weeks. Our model is $N(t) = Ae^{bt}$. The given information is that plugging in $t = 0$ and $t = 1$ give $N = 25$ and $N = 40$ respectively. Because $e^0 = 1$,we have $25 = A$,while $40 = Ae^b$. This gives $e^b = 40/25 = 8/5$,hence $b = \ln(8/5)$. In another two weeks we will have $t = 3$,so

$$N(3) = 25e^{3\ln(8/5)}.$$

When $N = 200,000$ we have $25e^{t\log(8/5)} = 200,000$ hence

$$e^{t\ln(8/5)} = \frac{200,000}{25} = 8,000 \text{ hence } t = \frac{\ln 8000}{\ln(8/5)} \approx 19.12.$$

Solution #2 (growth factor) If we use the growth factor B in the equation AB^t instead of the exponential constant b in Ae^{bt} we may get away without logs. In a week the increase was from 25 to 40, a factor of $8/5$ so clearly $B = 8/5$. Thus $N = 25(8/5)^t$. In three weeks we have $N(3) = 25(8/5)^3 = 512/5 = 102.5$.

Evidently the expression $25e^{3\ln(8/5)}$ can be simplified! The time needed to get to 200,000, a growth factor of 8000, is t such that $(8/5)^t = 8000$. This is, by definition $\log_{8/5} 8000$, which is equal to $\log_b 8000/\log_b(8/5)$ for any base b. If we pick $b = 10$,we get

$$t = \frac{\log 8000}{\log(8/5)} \approx 19.12\,.$$

Notice that that the answer given by plugging in logs is a special case of the growth factor approach, using base e. But the ratio of two logs is the same in any base! In Solution 2, we can use what we know about $\log_{10}$ to compute

$$\frac{\log 8000}{\log(8/5)} = \frac{\log 8 + \log 1000}{\log 8 - \log 5} = \frac{\log 8 + 3}{\log 8 - \log 5}$$

Having converted everything to logs of single digits, we can approximate by

$$3.9/(0.9 - 0.7) = 19.5.$$

So it should take between nineteen and twenty weeks to saturate the city. That's close enough to the more precise answer that we obtained from a calculator.

Checkpoint 38. (https://www2.math.upenn.edu/~ancoop/1070/sec-modeling.html#project-38)

The following exercise is a little more involved than the usual self-check exercise. It will be more readily solved if you realize that computations in the exponential model for equilibrating temperatures work better if you keep track of the temperature *difference* between substance and the environment rather than the temperature of the substance on its own.

Checkpoint 39. (https://www2.math.upenn.edu/~ancoop/1070/sec-modeling.html#project-39)

2 Limits

You might not think limits would show up in a calculus course oriented toward application. Wrong! There are a lot of reasons why you need to understand the basic of limits. You should know these reasons, so here they are.

1. You have already seen they show up in the definition of powers and logarithms when the exponent is not rational.
2. The definition of derivative (instantaneous rate of change) is a limit.
3. The number e is defined by a limit.
4. Continuous compounding is a limit.
5. Limits are needed to understand improper integrals, such as the integrals of probability densities.
6. Infinite series, which we will discuss briefly, require limits.
7. Discussing relative sizes of functions is really about limits.

3.1 Definitions of limit

You should learn to understand limits in four ways:

Intuitive The limit as $x \to a$ of $f(x)$ is the numerical value (if any) that $f(x)$ gets close to when x gets close to (but does not equal) a. This is denoted $\lim_{x\to a} f(x)$. If we only let x approach a from one side, say from the right, we get the one-sided limit $\lim_{x\to a^+} f(x)$.

Please observe the syntax: If I tell you a function f and a value a then the expression $\lim_{x\to a} f(x)$ takes on a numerical value or "undefined". The variable x is a bound variable; it does not have a value in the expression and does not appear in the answer; it stands for a continuum of possible values approaching a. The variable a is free and does show up in the answer; for example $\lim_{x\to a} x^2$ is equal to a^2.

Pictorial If the graph of f appears to zero in on a point (a, b) as the x-coordinate gets closer to a, then b is the limit, even if the actual point (a, b) is not on the graph. For example, suppose $f(x) = \frac{x^2-4}{x-2}$. Canceling the factor of $x-2$ from top and bottom, you can see this is equal to $x+2$, except when $x = 2$ because then you get zero divided by zero. Functions like this are not just made up for this problem. They occur naturally when solving simple differential equations, where indeed something different might happen if $x = 2$.

The graph of f has a hole in it, which we usually depict as an open circle, as in the left side of Figure 3.1.1. The value of $\lim_{x \to 2} f(x)$ is 2, even though f is undefined precisely at 2.

In this example the function f behaved very nicely everywhere except~2, growing steadily at a linear rate. The center figure shows the somewhat less well behaved function $g(x) := x \sin(1/x)$. This function is undefined at zero. As x approaches zero, the function wiggles back and forth an infinite number of times, but the wiggles are smaller and smaller. Intuitively, the value of the function g seems to approach zero as x approaches zero. Pictorially we see this too: zooming in on $x = 0$ in the right-hand figure, corroborates that $g(x)$ approaches zero.

We can take limits at infinity as well as at a finite number. The limit as $x \to \infty$ is particularly easy visually: if $f(x)$ gets close to a number C as $x \to \infty$ then f will have a horizontal asymptote at height C. Thus $3 + \frac{1}{x}$, $3 + e^{-x}$ and $3 + \frac{\sin x}{x}$ all have limit 3 as $x \to \infty$, as shown in Figure 3.1.3.

Formal

The precise definition of a limit is a little unexpected if you've never seen it before. We don't define the value of $\lim_{x \to a} f(x)$. Instead, we define when the statement $\lim_{x \to a} f(x) = L$ is true. It can be true for at most one value L. If there is such a L, we call this the limit. If there is no L, we say the limit does not exist. When asked for the value of $\lim_{x \to a} f(x)$, you should answer with either a real number, or "DNE", for "does not exist". We won't have to spend a lot of time on the formal definition. You should see and grasp it at least once. Use of the Greek letters ε and δ for the bound variables is a strong tradition.

Computational

We'll focus a lot on how to *compute* limits, given a formula for a function or some other information about it. This involves rules which allow us to express a complicated limit in terms of several more straightforward limits.

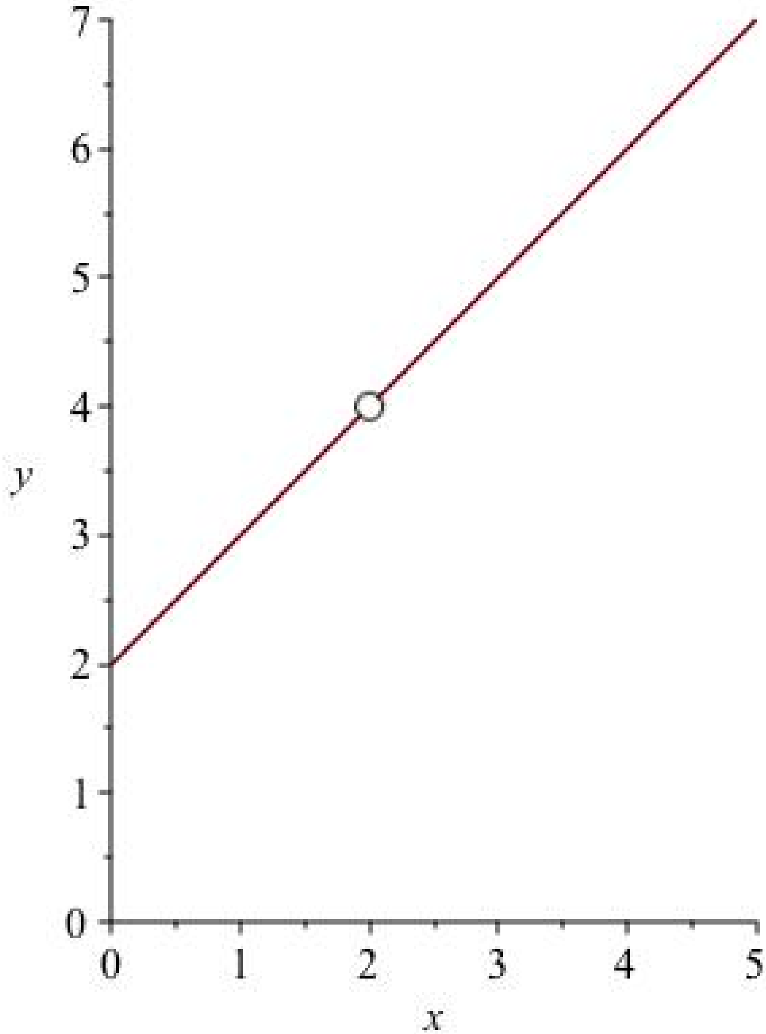


Figure 3.1.1. Left: $f(x) = x + 2$ except that f is undefined at $x = 2$;

Figure 3.1.2. A "wiggly" function which has a limit at $x = 0$. Zoom in and out to explore it.

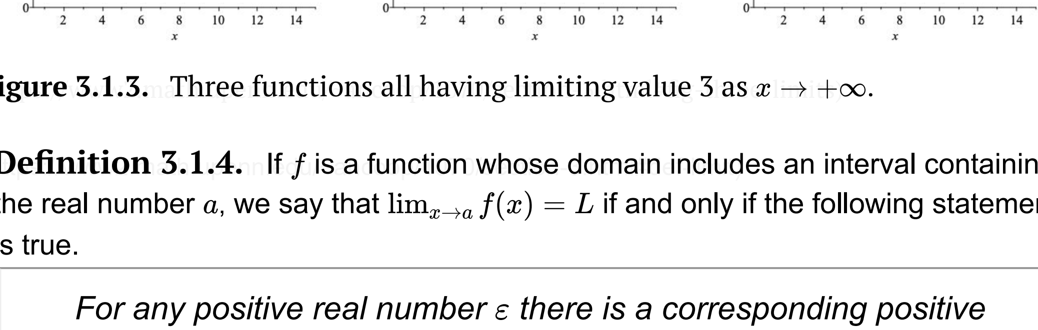


Figure 3.1.3. Three functions all having limiting value 3 as $x \to +\infty$.

Definition 3.1.4. If f is a function whose domain includes an interval containing the real number a, we say that $\lim_{x \to a} f(x) = L$ if and only if the following statement is true.

> *For any positive real number ε there is a corresponding positive real δ such that for any x other than a in the interval $[a - \delta, a + \delta]$, $f(x)$ is guaranteed to be in the interval $[L - \varepsilon, L + \varepsilon]$.*

In symbols, the logical implication that must hold is:

$$0 < |x-a| < \delta \Longrightarrow |f(x) - L| < \varepsilon\,.$$

Remark 3.1.5. Loosely speaking, you can think of ε as an acceptable error tolerance in the y value and δ as the precision with which you control the input (i.e., the x value). The limit statement says, you can meet even the pickiest error tolerance provided you can tune the input precisely enough.

Why is this a difficult definition? Chiefly because of the quantifiers. The logical form of the condition that must hold is: *For all* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that for all* $x \in [a-\delta, a+\delta]$, $\cdots$. This has three alternating quantifiers (for all... there exists... such that for all...) as well as an if-then statement after all this. Experience shows that most people can easily grasp one quantifier "for all" or "there exists", but that two is tricky: "for all ε there exists a $\delta\ldots$". A three quantifier statement usually takes mathematical training to unravel.

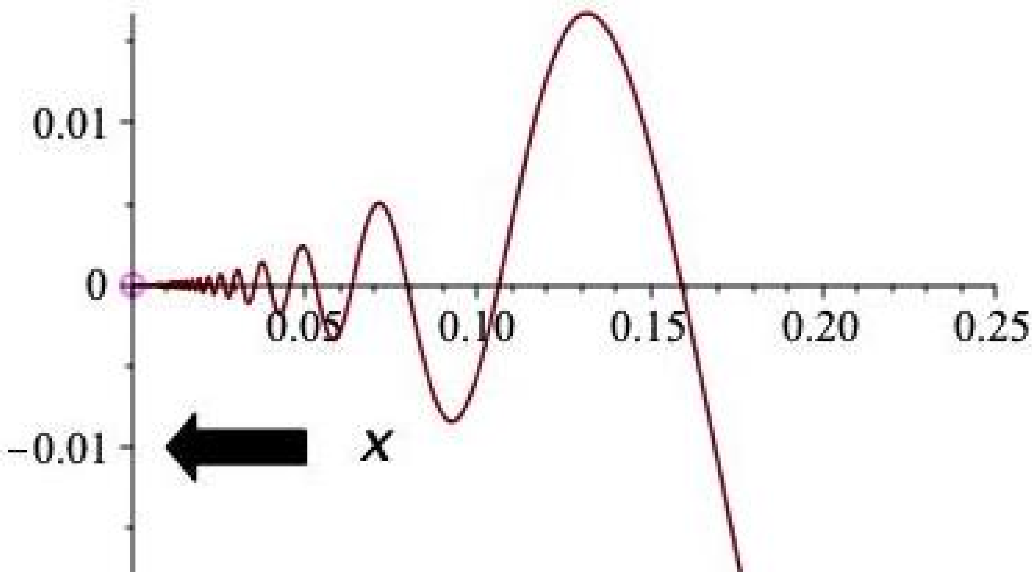
Some people find it easier to conceive of the formal definition as a game. Alice is trying to show it's true. Bob is trying to show it's false. Alice says to Bob, no matter what ε you give me, I can find a δ to make the implication true. (The implication is that all x-values fitting into Alice's δ-interval will give values of $f(x)$ inside Bob's interval.) Now they play the game: Bob tries to come up with a value of ε so small as to thwart Alice. Then Alice has to say her δ. If she can always do so (assuming Bob has not made a blunder in overlooking the right choice of ε) she wins and the limit is L. If not (unless Alice has overlooked a δ that would have worked), Bob has won and the limit is not L.

3.2 Variations

Before introducing computational apparatus for limits, we need to finish the definitions by defining some variations: one-sided limits, limits at infinity and "limits of infinity" (which are in quotes because technically they are not limits at all).

One-sided limits. Change the definition so that $f(x)$ is only required to approach L when $x \to a$ if x is greater than a. We say x "approaches a from the right," thinking of a number line. If the value of $f(x)$ approaches L when x approaches a from the right, we say that the limit from the right of $f(x)$ at $x = a$ is L, and denote this $\lim_{x\to a^+} f(x) = L$. If we require $f(x)$ to approach L when x approaches a but only for those x that are less than a, this is called having a limit from the left and is denoted $\lim_{x\to a^-} f(x) = L$.

Remark 3.2.1. Just like wind directions (North wind, South wind, etc.), one-sided limits are named for the direction they come from, not the direction x is moving. Thus, $\lim_{x\to 0^+}$ is evaluated by letting x approch zero from the positive direction, as shown to the right.



Checkpoint 40.

Checkpoint 41.

Both kinds of one-sided limits require something less stringent, so the statement $\lim_{x\to a} f(x) = L$ automatically implies both $\lim_{x\to a^+} f(x) = L$ and $\lim_{x\to a^-} f(x) = L$. Likewise, if $f(x)$ is forced to approach L when x approaches a from the right, but also when x approaches a from the left, then this covers all x, and the (unrestricted) limit will be L. If you want, you can summarize this as a theorem -- wait, no it's too puny, let's make it a proposition. We won't be referring to this too often, but here it is.

Proposition 3.2.2. For every function f and real numbers a and L,

$$\lim_{x\to a} f(x) = L \text{ if and only if } \lim_{x\to a^+} f(x) = L \text{ and } \lim_{x\to a^-} f(x) = L\,.$$

In words, a limiting value for a function exists at a point if and only if the two one-sided limits exist are are equal.

Checkpoint 42.

Example 3.2.3. Let $f(x) = \lfloor x \rfloor$, the greatest integer function. Let's evaluate the one-sided limits and two-sided limit at a couple of values. First, take $a = \pi$, you know, the irrational number beginning $3.14\ldots$. If we just look near this value, say between 3.1 and 3.2, it is completely flat: a constant function, taking the value~3 everywhere. So of course the limit at $x = \pi$ will also be~3. This is the same by words or pictures; see Figure 3.2.4.

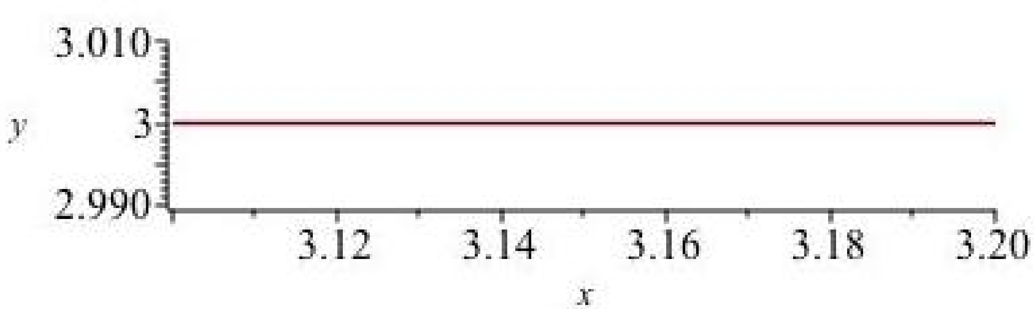


Figure 3.2.4. An interval where the greatest integer function is constant.

By the formal definition, no matter what ε is chosen, you can take $\delta = 0.1$, say, and $f(x)$ will be within ε of~3 because it will be exactly~3. So the limit is~3, hence so are both one-sided limits as in the picture just above.

Now take x to be an integer, say $a = 5$. The limit from the right looks like it did before, with $f(x)$ taking the value~5 for every sufficiently close x (here sufficient means within~1) greater than~5. On the other hand, when x is close to~5 but less than~5, we will have $f(x) = 4$, as in the picture below. Thus,

$$\lim_{x\to 5^+} f(x) = 5$$
$$\lim_{x\to 5^-} f(x) = 4$$
$$\lim_{x\to 5} f(x) \text{DNE}$$

The two-sided limit does not exist because the two one-sided limits are not equal; see Figure 3.2.5 .

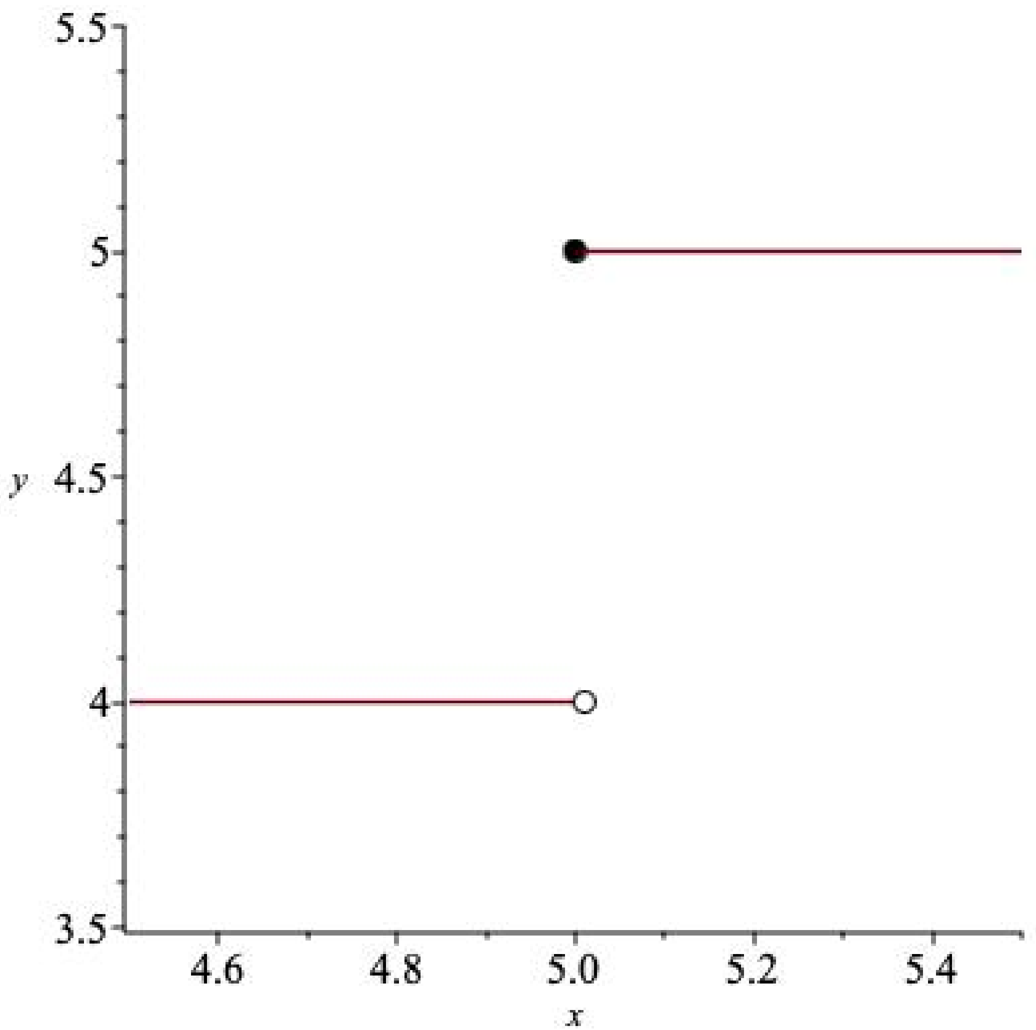


Figure 3.2.5. An interval where the greatest integer function is discontinuous.

Checkpoint 43. (https://www.math.upenn.edu/~ancoop/1070/section-3.html#project-43)

Limits at infinity. You have already seen the pictorial and verbal version of a limit at infinity. Here is the formal definition. It repeats a lot of the definition of a limit at $x = a$. The only difference is that instead of having to come up with an interval

$[a-\delta, a+\delta]$ guaranteeing $f(x)$ is within ε of the limit, you have to come up with an "interval near infinity". This turns out to mean an interval $[M, \infty)$. In other words, there must be a real number M guaranteeing $f(x)$ is with ε of L when $x > M$.

Remark 3.2.6. Informally, "close to infinity" turns into "sufficiently large". In the tolerance/accuracy analogy, getting $f(x)$ to be close to L to within the acceptable tolerance will result from guaranteed largeness of the input rather than guaranteed closeness to a.

Definition 3.2.7. We say that $\lim_{x\to\infty} f(x) = L$ if and only if L is a real number and:

> *For any positive real number ε (think of this as acceptable tolerance in the y value) there is a corresponding real M (think of this as guaranteed minimum value for x) such that for any x greater than M, $f(x)$ is guaranteed to be in the interval $[L-\varepsilon, L+\varepsilon]$.*

In symbols, the logical implication that must hold is:

$$x > M \Longrightarrow |f(x) - L| < \varepsilon\,.$$

If a real number L exists satisfying this, we write $\lim_{x\to\infty} f(x) = L$. Sometimes to be completely unambiguous, we put in a plus sign: $\lim_{x\to+\infty} f(x) = L$.

Checkpoint 44.

Limits at $-\infty$ are defined exactly the same except for a single inequality that is reversed. Now the implication that must hold is that for some (possibly very negative) M,

$$x < M \Longrightarrow |f(x) - L| < \varepsilon\,.$$

When this holds, we write $\lim_{x\to-\infty} f(x) = L$. When no such L exists, we write $\lim_{x\to-\infty} f(x)$ DNE.

Example 3.2.8. Let $f(x) := \frac{x}{\sqrt{1+x^2}}$. Because $\sqrt{1+x^2}$ is a little bigger than $|x|$ but almost the same when x or $-x$ is large, this function satisfies

$$\lim_{x\to\infty} f(x) = 1$$
$$\lim_{x\to-\infty} f(x) = -1$$

As with limits at finite x values, one shouldn't write $\lim_{x\to\infty} f(x) =$ DNE for grammar.

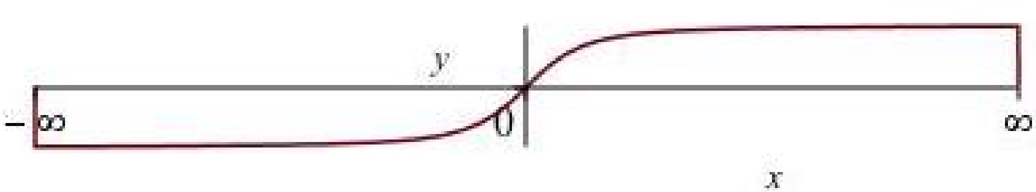


Figure 3.2.9. graph of $x/\sqrt{1+x^2}$

The graph of this function is shown in Figure 3.2.9. It has horizontal asymptotes at 1 and -1. This suggests how to define a horizontal asymptote.

Definition 3.2.10. A function f or its graph is said to have a ***horizontal asymptote*** at height b if $\lim_{x\to\infty} f(x) = b$ or $\lim_{x\to-\infty} f(x) = b$.

Checkpoint 45.

"Limits" of infinity. Consider the function $f(x) = 1/x^2$, defined for all real numbers except zero. What happens to $f(x)$ as $x \to 0$? By our definitions, $\lim_{x\to 0} 1/x^2$ ***does not exist***. But we can see that $f(x)$ "goes to infinity". Because infinity is not a number, the limit technically does not exist. However, it is useful to classify DNE limits as ones where the function approaches ∞ (or $-\infty$) versus ones where there is no consistent behavior.

Remark 3.2.11. This time, instead of staying within a tolerance of ε in the output, we make the output sufficiently large (greater than any given N) or small. We do this by guaranteeing δ accuracy in the input (for limits as $x \to a$) or by making the input sufficiently large or small (limits as $x \to \pm\infty$).

Formally, this turns into the following defintion.

Definition 3.2.12. If f is a function and a is a real number, we say that $\lim_{x\to a} f(x) = +\infty$ if for every real N there is a $\delta > 0$ such that $0 < |x - a| < \delta$ implies $f(x) > N$.

Again, if we reverse the last inquality to require that $f(x) < N$ (and N can be a very negative number) we get the definition for a limit of negative infinity. Please remember these are all subcases of limits that don't exist! If you show that a limit is infinity, you have shown that the limit does not exist (and you have specified a particular reason it doesn't exist).

Example 3.2.13. Let's check that $\lim_{x\to 0} 1/x^2 = +\infty$. Given a positive real number N, how can we ensure $f(x) > N$? Answer: for positive numbers, f is decreasing and $f(x) = N$ precisely when $x = 1/\sqrt{N}$. Therefore, if we keep x positive but less than $1/\sqrt{N}$ then $f(x)$ will be greater than N. We have just shown that $\lim_{x\to 0^+} 1/x^2 = +\infty$. Similarly, when x is negative, if we keep x in the interval $(-1/\sqrt{N}, 0)$ we ensure $1/x^2 > N$. So $\lim_{x\to 0^-}$ is also $+\infty$. Both one-sided limits are $+\infty$, therefore

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

Don't forget, it follows from the limit being $+\infty$ that

$$\lim_{x\to 0} \frac{1}{x^2} \text{ does not exist.}$$

For one-sided limits and limits at infinity, the DNE case also includes a case where the limit would be said to be infinity. Stating all these would be repetitive. Try one, to make sure you agree it's straightforward.

Checkpoint 46.

Checkpoint 47.

Limit of a sequence. A special case of limits at infinity is when the domain of f is the natural numbers. When f is only defined at the arguments $1, 2, 3, \ldots$, it is more usual to think of it as a sequence $b_1, b_2, b_3, \ldots$, where $b_k := f(k)$. The definition of a limit at infinity can be applied directly, resulting in the definition of the limit of a sequence.

Definition 3.2.14. [limit of a sequence] Given a sequence $\{b_n\}$ and a real number L we say $\lim_{n\to\infty} b_n = L$ if and only if for all $\varepsilon > 0$ there is an M such that $|b_n - L| < \varepsilon$ for every $n > M$.

Remark 3.2.15. Often we use letters such as n or k to denote integers and x or t to denote real numbers. Therefore, by context, $\lim_{n\to\infty} 1/n$ denotes the limit of a sequence while $\lim_{t\to\infty} 1/t$ denotes the limit at infinity of a function. Formally we should clarify and not count on the name of a variable to signify anything! But because the two definitions agree, often we don't bother.

Checkpoint 48.

Pictorially, if a sequence has a limit L, then for every pair of parallel horizontal lines, however narrow, enclosing the height L, the sequence must eventually stay between them. This is shown in Figure 3.2.16.

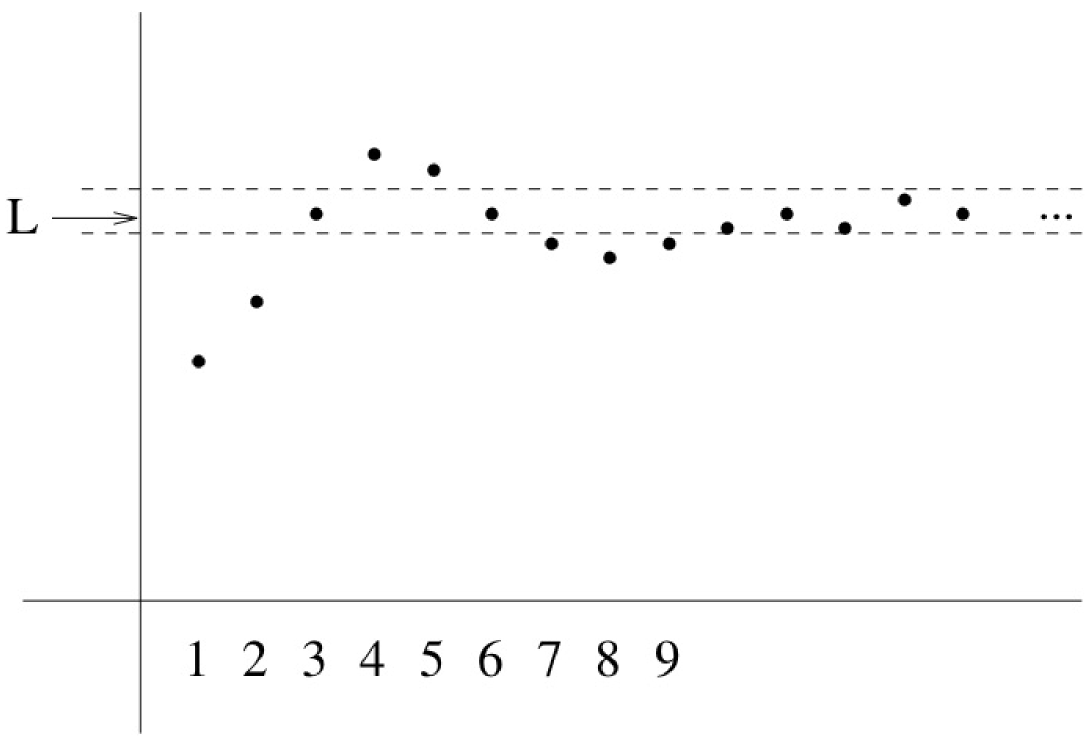


Figure 3.2.16. For these two parallel lines, once $k > 9$, the height b_k is between the lines.

As you will see, Proposition 3.4.1 and Proposition 3.4.2 give ways to determine limits of more complicated functions once you understand limits of some basic functions. Here is another piece of logic that can help do the same thing. You will prove it in your homework.

Theorem 3.2.17. Let a be a real number or $\pm\infty$ and let f, g and h be functions satisfying $f(x) \le g(x) \le h(x)$ for every x. If $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} h(x) = L$ then also $\lim_{x\to a} g(x) = L$. If we know only that $\lim_{x\to a^+} h(x) = \lim_{x\to a_-} h(x) = L$ then we can conclude $\lim_{x\to a^+} g(X) = L$, and same for limits from the left.

The same fact is true of sequences: if $a_n \le b_n \le c_n$ for these three sequences and the first and last sequence converge to the same limit L, then so does the middle one. We will not do anything with this now, but will get back to this fact in a week or two. The next exercise brushes up on the logical syntax of limits.

Checkpoint 49. (https://www2.math.upenn.edu/~ancoop/1070/section-3.html#project-49)

3.3 Continuity

Definition 3.3.1. A function f is said to be ***continuous at the value*** a if the limit exists and is equal to the function value, in other words, if $\lim_{x\to a} f(x) = f(a)$.

Intuitively, this means the limit at a exists and there is no hole: the function is actually defined at a and wasn't given some weird other value. To illustrate what we mean, consider this picture of a function that is discontinuous at $x = 2$ even though $\lim_{x\to 2} f(x)$ exists and so does $f(2)$, because the values don't agree.

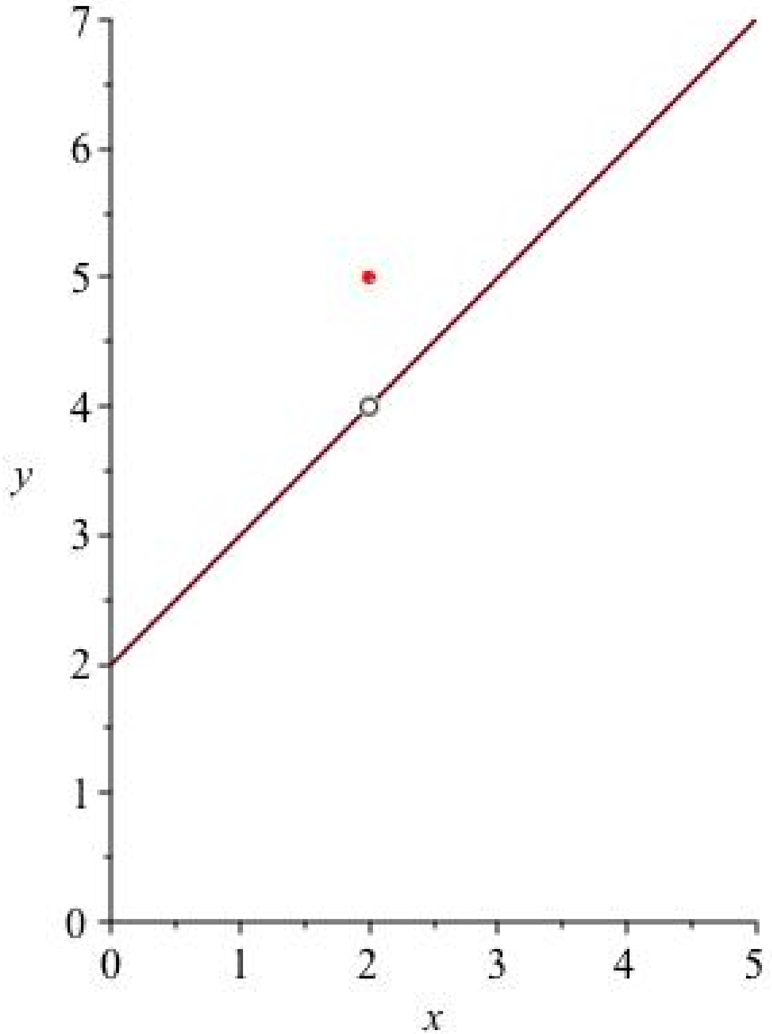


Figure 3.3.2.

Checkpoint 50.

Continuity on regions.

Definition 3.3.3. A function is said to be ***continuous on an open interval*** (a, b) if it is defined and continuous at every point of (a, b).

A function is said to be ***continuous on an closed interval*** (a, b) if it is defined and continuous at every point of (a, b), with only one-sided contintuity required at a^+ and b_-.

A function f is said to be just plain ***continuous*** if it is continuous on the whole real line.

Checkpoint 51.

Before going on to use the notion of continuity to help us compute limits, we will state one famous result which will seem either stupid and obvious or deep and tricky.

Theorem 3.3.4. Intermediate Value Theorem. Let f be a continuous function defined on the closed interval $[a, b]$ and suppose that y is any value between the values $f(a)$ and $f(b)$. Then there is some number c in the interval $[a, b]$ satisfying $f(c) = y$.

This says, basically, a continuous function can't get from one value to another without hitting everything in between. The theorem is most often used when there is a number we can only define this way. For example, let $f(x) := e^x/x$, which is an increasing function on the half-line $[1, \infty)$. We want to say "let c be the value for which $f(c) = 3$." How do we know there is one? Well, $f(1) = e$, which is less than 3, and $f(3) \approx 6.695$ which is greater than 3. So there must be an argument between 1 and 3 where f takes value 3. There can be only one because f is strictly increasing (you can prove after another two sections).

3.4 Computing limits

Computing a limit by verifying the formal definition is a real pain. There is computational apparatus that allows us to compute limits of many functions once we know limits of a few simple ones. One approach we have seen in textbooks is to give a list of rules that work. It looks something like this.

Proposition 3.4.1. If $\lim_{x\to a} f(x) = L$ and c is a real number then

$$\lim_{x\to a} cf(x) = cL$$
$$\lim_{x\to a} f(x)^c = L^c \text{ provided } L > 0$$

Proposition 3.4.2. If f and g are functions and a, K, L are real numbers with $\lim_{x\to a} f(x) = K$ and $\lim_{x\to a} g(x) = L$, then

$$\lim_{x\to a} f(x) + g(x) = K + L$$
$$\lim_{x\to a} f(x) - g(x) = K - L$$
$$\lim_{x\to a} f(x) \cdot g(x) = K \cdot L$$
$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{K}{L} \text{ provided } L \neq 0$$

Example 3.4.3. Suppose f is a polynomial: $f(x) = b_n x^n + \cdots + b_1 x + b_0$. What is $\lim_{x\to a} f(x)$? We hope you think this is a really boring example. Of course, the polynomial is continuous (picture in your mind the graph of a polynomial) and so we expect $\lim_{x\to a} f(x) = f(a)$. But why?

First note that we can evaluate the limit at a of the monomial x^k as a^k using the second conclusion of Proposition 3.4.1. We can evaluate the limit at a of each monomial $b_k x^k$ as $b_k a^k$ by applying the first conclusion of Proposition 3.4.1 with $c = b_k$ and $f(x) = x^k$.

So each term in the polynomial has the right behavior. We still need to combine them all by addition. This is allowed to us by Proposition 3.4.2.

The fact that polynomial limits can be computed just by "plugging in" the value $x = a$ comes up a lot, so we record it as a propostion.

Proposition 3.4.4. Polynomials are continuous. The limit of a polynomial f at a is always given by $f(a)$.

Checkpoint 52. enn.edu/~ancoop/1070/section-3.html#project-52)

Checkpoint 53. enn.edu/~ancoop/1070/section-3.html#project-53)

So that Proposition 3.4.1 Proposition 3.4.2 don't look like arbitrary rules from out of nowhere, you should realize they can be proved, and in fact follow from one basic theorem.

Theorem 3.4.5. If the function f has a limit L at $x = a$ and the function H is continuous at L then $H \circ f$ will have the limit $H(L)$ at $x = a$. Formally,

$$\lim_{x\to a} f(x) = L \text{ implies } \lim_{x\to a} H(f(x)) = H(L) \text{ provided is continuous at } L\,.$$

Why do the Proposition 3.4.1 and Proposition 3.4.2 follow from this principle? Let $H(x)$ be the continuous function cx. Then $H \circ f$ is $cf(x)$ and we recover the first conclusion of Proposition 3.4.1. Setting $H(x) := x^c$ recovers the second conclusion.

Checkpoint 54. enn.edu/~ancoop/1070/section-3.html#ex-changevar)

Checkpoint 55. enn.edu/~ancoop/1070/section-3.html#project-55)

Some more techniques and tricks. This course is more about using limits than it is about computational technique, but you should at least see some of the standard techniques for cases that go beyond what's in Proposition 3.4.1 and Proposition 3.4.2.

Suppose you need to evaluate $\lim_{x\to a} f(x)/g(x)$. If both f and g have nonzero limits at a, say L and M, then Proposition 3.4.2 tells you

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

In fact if $L = 0$ but $M \neq 0$, this still works. If $M = 0$ but $L \neq 0$, then the question of evaluating $\lim_{x\to a} f(x)/g(x)$ also has an easy answer.

Checkpoint 56. enn.edu/~ancoop/1070/section-3.html#project-56)

The remaining case, when $L = M = 0$, can be enigmatic. Calculus provides one solution you will see in a few weeks (L'Hôpital's Rule), but you can often solve this with algebra. If you can factor out $(x - a)$ from both f and g, you may get a simpler expression for which at least one of the functions has a nonzero limit.

Example 3.4.6. enn.edu/~ancoop/1070/section-3.html#eg-limit-factor)

What is

$$\lim_{x\to 5} \frac{x^2 - 25}{x^2 - 5x}\ ?$$

Both numerator and denominator are continuous functions with values of zero (hence limits of zero) at~5. That suggests dividing top and bottom by $x - 5$, resulting in $\lim_{x\to 5} \frac{x+5}{x}$. Both numerator and denominator are continuous functions so we can just evaluate and get $10/5$ so the answer is 2.

You don't have to worry about having divided top and bottom by zero in Example 3.4.6 because the limit depends on values of $f(x)$ for x ***near 5*** but ***not at*** $x = 5$ ***itself***. The value of $f(5)$, whether or not exists, has no bearing on the

Sometimes you have to do a little algebra to simplify. Here's an example of one of the most common simplification tricks.

Example 3.4.7. What is $\lim_{x\to 0} \frac{\sqrt{x+1}-1}{x}$? Multiplying and dividing by the so-called conjugate expression, where a sum is turned into a difference or vice versa, gives

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

The numerator and denominator are continuous at $x = 0$ with nonzero limits of 1 and 2 respectively, so the limit is equal to $1/2$.

This algebra trick occurs so commonly throughout mathematics that you should always think about conjugate radicals every time you see an expression with a square root added to or subtracted from something!

Further tricks can wait until you've learned some more background. Although limits are needed to define derivatives, you can then use derivatives to evaluate more limits (L'Hôpital's Rule). Similarly, limits are used to define orders of growth, which can then be used to evaluate more limits.

3 Derivatives

In the Preface (sec-zero.html), we caught a glimpse of the usefulness of linear approximations. If all relationships were linear, life would be easy. But most of the models we have developed so far are nonlinear! In this chapter and the ones that follow, we'll dig deeper into how to compute and use linear approximations to nonlinear relationships.

4.1 Concept of the derivative

It's easy to define your average speed for a trip: take the number of miles, divide by the number of hours, and there's your average speed in miles per hour. If you journey at constant speed, then that's also your speed at every moment of the trip. Most of us do not travel at constant speed. What is your speed then? How do you define it? How do you measure it? How do you compute it if you know some equation for your position at time t?

The concept of instantaneous speed is subtle. It is what spurred the invention of calculus over a few decades near the year 1700. It is a very general notion. Average speed is distance traveled per total time. Instantaneous speed is some instantaneous version.

Checkpoint 57.

If you replace "distance traveled" by "production price" and "time elapsed" by "units produced" you get the notions of average production cost per unit; marginal cost per unit is the instantaneous version. The list of applications is endless. Mathematically, they are all the same: if $f$is a function and $x_0$and $x_1$are starting and ending arguments for f, then the average change in $f$over the interval is $(f(x_1) - f(x_0))/(x_1 - x_0)$; the instantaneous rate of change of $f$with respect to $x$is called the ***derivative*** of f with respect to x and denoted $f'(x)$.

Checkpoint 58.

In this section we will see how to understand f' both physically and mathematically. We will continue to use instantaneous speed as a running example of the physical concept, and instantaneous rate of change of $f(x)$ with respect to x as the corresponding mathematical concept.

Remark 4.1.1. We can take the slope of the function f at any point. Taking it at x gives a value we call $f'(x)$.That means that f' is a function: give it an argument x and it will produce the slope of f at that point. It will be helpful to keep in mind that the derivative operator takes as input functions f and produces as output their derivatives f'.Operator is a fancy word for a function whose input and output are functions rather than numbers. Taking derivatives is a ***linear operator***. This is captured in Proposition 5.1.1 through Proposition 5.1.3 below.

Checkpoint 59.

4.2 Definitions

Most functions we use in mathematical modeling have unique tangent lines at most points. The slope of the tangent line to the graph of $f$at the point $(x, f(x))$ seems like one reasonable definition of $f'(x)$.In rare cases, such as you have already seen, we can use geometry to prove there is exactly one line tangent to the graph of $f$at a point and compute the slope.

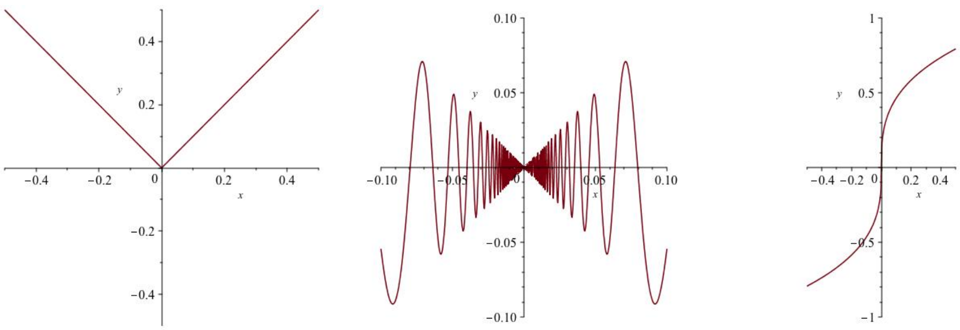


Figure 4.2.1. graphs of $|x|$, $x\sin(1/x)$ and $\sqrt[3]{x}$

Unfortunately, there are not many functions for which the graph is a well known geometric object. In most cases we can't use geometry to conclude that there is a tangent line, that there is only one tangent line, or what the slope of this line is, if indeed there is exactly one. Keeping this in mind, we will use limits to come up with a definition that works for most functions and, when it does not work, as in the examples in Figure 4.2.1, gives an indication of why. In cases when it does not work, in fact we would probably agree that there is no good way to make sense of the instantaneous slope.

Checkpoint 60.

We can take average slopes over any interval we want. The slope over the interval $[a, b]$ is the slope of the ***secant line*** passing through $(a, f(a))$ and $(b, f(b))$.This is also called the ***difference quotient*** of f at the arguments a and b.What happens when one endpoint of the interval is x and the other is very close to x? Pictorially, it looks the slope get very close to the slope of the tangent line at $(x, f(x))$. Figure 4.2.2 gives a tool for exploring this secant-line-to-tangent-line approximation.

Figure 4.2.2. Secant lines approaching a tangent line. You can input your favorite function and slide c, which is where the tangent line is computed. Slide a and b to see the secant line. What happens as we move a and b closer to c?

Definition using limits. The derivative is a mathematical definition meant to compute the slope of the tangent line at a. Definition 4.2.3, however, only talks about limits of slopes of secants, not of tangents. Do you think these two notions will always coincide? There isn't a right answer to this.

Checkpoint 61.

Definition 4.2.3. Let f be a function whose domain contains an interval around the point a. Define

$$f'(a) := \lim_{b\to a} \frac{f(b) - f(a)}{b - a} \tag{4.2.1}$$

if the limit exists, and say that $f'(a)$ is undefined if the limit does not exist. Because we want to emphasize that $b - a$ is going to zero, we often define $h := b - a$ and rewrite the definition as

$$f'(a) := \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}\ . \tag{4.2.2}$$

The two equations (4.2.1) and (4.2.2) are algebraically equivalent.

Checkpoint 62.

Example 4.2.4. Let $f(x) = x^2$. Let's see the definition to try to compute $f'(1)$. By definition, this is

$$\lim_{b \to 1} \frac{f(b) - f(1)}{b - 1}$$

Evaluating the numerator gives

$$\lim_{b \to 1} \frac{b^2 - 1}{b - 1} = \lim_{b \to 1} b + 1 = 2$$

The first equality is true because we can cancel the factors of $b - 1$ (remember, the limit looks at values of b near 1 but not equal to 1). The second equality is true because we can evaluate the limit of the polynomial $b + 1$ at $a = 1$ by plugging in 1 for b (Proposition 3.4.4).

Checkpoint 63.

Notation. We already agreed to use a prime after the function name as one way to denote a derivative. Thus the derivative of f is f', the derivative of g is g', the derivative of Γ is Γ', etc. We may need to refer to the derivative of a function when it has not been given a name. One could imagine something like the notation $(cx)'$ for the derivative of the function "multiply by c", or perhaps the more precise $(x \mapsto cx)'$.

To avoid ambiguity, we use the notation $\frac{df}{dx}$ for the derivative of f with respect to x. This is better than f' when there is more that one variable that could be differentiated. that one variable that could be differentiated. You can also write this as $\frac{d}{dx} f$ when f is a big long cumbersome expression, for example,

$$\frac{d\left(\frac{e^{x^2-1}\sin x}{1+x}\right)}{dx} \quad \text{is the same as} \quad \frac{d}{dx}\left(\frac{e^{x^2-1}\sin x}{1+x}\right)$$

Then there is the question of how to write $f'(a)$, the value of the function f' at argument a, in this notation. Should we write $\frac{d\,f(a)}{dx}$ or $\frac{df}{dx}(a)$? The second is better, for example, $\frac{d(x^3-3x+1)}{dx}(a)$, because the first looks like you are differentiating a constant. Another common way of writing this is $\left.\frac{d\,(x^3-3x+1)}{dx}\right|_{x=a}$.

Checkpoint 64.

The notation $\frac{df}{dx}$, also written df/dx, arose because Leibniz had an interpretation where df was an infinitesimally small change in the quantity f and dx was an infinitesimally small change in the quantity x. A lot of computations in

Further interpretations: error propagation and marginal effect. You have seen examples in which derivatives represent speed. More generally, the derivative of a function of time represents the rate of change of the quantity per time. Here are some other things derivatives commonly represent.

Suppose you have a formula $f(x)$ involving a quantity x that is measured, but with measurement error. Then $f'(x)$ tells you how much error you get in f per amount of error in measuring x.

Example 4.2.5. A 4×8 foot board is cut parallel to the long side to obtain a 3×8 board. The accuracy of the cut is $1/4$ inch. What is the accuracy of the area, in square feet? Writing $A = \ell \times w$ and differentiating gives $dA/dw = \ell$ =8 feet in our case. Therefore, the error in area (in square feet) is 8 feet times the measurement error in the width (in linear feet). Plugging in a measurement error of $1/4$inch, which equals $1/48$ feet, we see the area is accurate to within $8\text{ ft} \times \frac{1}{48}\text{ ft} = \frac{1}{6}\text{ ft}^2$.

The symbol Δ is the upper case Greek letter Delta an often used to denote change in a quantity or error in a measurement.

Checkpoint 65.

Another interpretation is the marginal effect of the variable on the function. For example, if $f(x)$ represents the cost of producing x barrels of refined oil, then $f'(x)$ is the marginal cost of production of more oil. Unless f is linear, this will depend on x. The marginal cost of further production usually depends on the present level of production.

4.3 First and second derivatives, and sketching

Knowledge of the deriviatve can help you sketch a function more accurately. The very first practice problem asked you to incorporate slope information into a sketch. Sketching is as much an art as a science, but there are methodical ways to use information about the function and its derivatives.

To begin with, knowing where the derivative is positive and negative determines whether it is sloping up or down as you move right. In other words, the sign of the derivative indicates whether the function is increasing or decreasing. Where the sign of the derivative changes from positive to negative as you move right, the function changes from increasing to decreasing. That means someone hiking on the graph of the function from left to right has been walking upwards and now begins to walk downward; see Figure 4.3.1.

Figure 4.3.1. Slide the point c along the graph to watch the hiker get to the hilltop. What happens to f'?

Checkpoint 66.

Because transitions in the sign of f'correspond to hilltops and valley floors, finding values of $x$that are maxima and minima for $f(x)$ involves finding values of $x$for which $f'(x) = 0$.We discuss this at greater length in Unit 7 (sec-max-min.html). For the purposes of sketching, the moral of the story is: know where f' is positive and where it is negative, and use this to depict a function that is increasing and decreasing in the right places.

Checkpoint 67.

The second derivative. A Japanese proverb says, "The other side also has another side." The function f has a derivative. f' is also a function. Therefore, ***The derivative also has a derivative.*** Not quite so poetic, but very useful for sketching functions. The derivative of the derivative is called the second derivative, denoted f'', or $\frac{d^2 f}{dx^2}$.The sign of the derivative says where a function is increasing or decreasing, therefore the *sign* of f'' indicates where the *slope* f' is *increasing* or *decreasing*. We use italics here as a visual reminder that there are a number of levels (original function, first derivative, second derivative) and attributes (positive/negative, increasing/decreasing) and it's easy to get mixed up what corresponds to what. We hope Table 4.3.2 helps clarify *whose* positivity says what about *whom*.

Table 4.3.2. Interpreting what the signs of the second, first, and zeroth derivatives of f mean for f.

f'' is	f' is	f is	graph of f
positive	increasing	of increasing slope	concave upward
negative	decreasing	of decreasing slope	concave downward
	positive	increasing	upward
	negative	decreasing	downward
		positive	above the horizontal axis
		negative	below the horizontal axis

Remark 4.3.3. The placement of the 2 in the numerator of $\frac{d^2 f}{dx^2}$ may seem strange, but it reflects something important: (d/dx) is a differential operator, and (d^2/dx^2) is the result of applying this operator twice. This becomes important in later courses such as Math 114.

Checkpoint 68. (https://www2.math.upenn.edu/~ancoop/1070/section-4.html#project-68)

Of course, not every function is differentiable, and not every derivative is itself differentiable, so f'' may not exist even if f' exists.

We have talked informally about functions that are concave up or down. It is time to give a definition. In fact we give two definitions, one algebraic and one pictorial. The pictorial one is in fact more general because it works when f' does not exist. When f' exists on (a, b), then the two definitions agree.

Definition 4.3.4. concavity. When the function f' exists and is increasing, we say that f is concave upward.

When the function f' exists and is decreasing, we say that f is concave downward.

Definition 4.3.5. concavity (pictorial definition). If $(a, b)$is an open interval in the domain of $f$and if for every pair of numbers $x, y \in (a, b)$the graph of f on (a, b) lies below the line segment connecting $(x, f(x))$to $(y, f(y))$, we say that f is concave upward on (a, b).

Checkpoint 69. (https://www2.math.upenn.edu/~ancoop/1070/section-4.html#project-69)

To summarize, if f'' exists on (a, b) then the sign of f'' determines the concavity of f. If f'' doesn't exist or you can't compute it, use Definition 4.3.4.

Points of inflection. We never formally defined a tangent line. One definition would be "A line that touches a graph of a function at precisely one point and stays on one side of the graph other than this."

Checkpoint 70. (https://www2.math.upenn.edu/~ancoop/1070/section-4.html#ex-tangent-line)

As you can see, the intuitive definition of tangent line is subject to unanticipated judgment calls. This motivates a more formal definition.

Definition 4.3.6. tangent line. If f is differentiable at a, the tangent line to f at a is defined to be the line $(y - f(a)) = m(x - a)$ where $m = f'(a)$.

Checkpoint 71. (https://www2.math.upenn.edu/~ancoop/1070/section-4.html#project-71)

One confusing case is when the second derivative is zero. What happens to the concavity at such a point? Often it switches from up to down or *vice versa*. Wherever concavity switches is called a ***point of inflection***. The geometric concept of an inflection point does not require calculus, though the notion seems not to have been discussed much before the advent of calculus.

Checkpoint 72. (https://www2.math.upenn.edu/~ancoop/1070/section-4.html#project-72)

Checkpoint 73. (https://www2.math.upenn.edu/~ancoop/1070/section-4.html#project-73)

4 Computing derivatives

There are a lot of tools for computing derivatives that are relatively easy to remember and use. These tools are theorems -- they can all be derived from the definition via limits and some computation. You will get familiar enough with these that you will happily use them without thinking. The structure of this chapter is philosophically backwards: we give you nearly all the tools right away, then give arguments and explanations for some of them, postponing some of the arguments and explanations until we have developed a few more tools. We do this because calculus is so much more fun when you know enough do a few computations!

5.1 Tools for computing derivatives

There are two kinds of tools for computing derivatives. We'll use the following terminology to keep them straight.

formulas *Formulas* are statements about the derivative of a particular function. For example, there is a formula for the derivative of the sine function: $\frac{d}{dx}\sin(x) = \cos(x)$.

rules *Rules* are ways to combine information about derivatives of some functions in order to compute the derivatves of other functions.

The idea will be to leverage the *formulas* to help us compute the derivative of any function which can be built from our stable of basic functions.

The rules. If we build a function out of others by adding, subtracting, multiplying, or composing, and we happen to know the derivatives of all the parts, we can express the derivative of the function we've built in terms of the derivatives of the parts.

Proposition 5.1.1. sum rule. Let f and g be differentiable functions. Then $(f+g)' = f' + g'$.

Proposition 5.1.2. difference rule. Let f and g be differentiable functions. Then $(f-g)' = f' - g'$.

Proposition 5.1.3. multiplication by a constant. Let f be a differentiable function and c be a constant. Then $(cf)' = cf'$.

Checkpoint 74.

Proposition 5.1.4. product rule. Let f and g be differentiable functions. Then $(fg)' = f'g + g'f$.

Proposition 5.1.5. quotient rule. Let f and g be differentiable functions. Then for any x such that $g(x) \neq 0$,

$$\frac{d}{dx}\frac{f(x)}{g(x)} = \frac{gf' - fg'}{g^2}$$

all functions on the right-hand side evaluated at x.

Proposition 5.1.6. chain rule. Let f and g be differentiable functions. Let a be a real number inside an open interval in the domain of g such that $g(a)$ is inside an open interval in the domain of f. Then

$$\frac{d}{dx} f(g(x))\Big|_{x=a} = \left(\frac{df}{dx}\Big|_{x=g(a)}\right)\left(\frac{dg}{dx}\Big|_{x=a}\right)$$

We can write this more compactly as

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

the longer version can help unravel any confusion.

The formulas. We list a few formulas that are either obvious from the definition or are ones you've worked out already.

Proposition 5.1.7. some easy derivatives. Let c be any real constant. Then,

$$\frac{d}{dx}\, c = 0$$

$$\frac{d}{dx}\, cx = c$$

$$\frac{d}{dx}\, x^2 = 2x$$

$$\frac{d}{dx}\, \sqrt{x} = \frac{1}{2\sqrt{x}} \quad \text{for } x > 0\,.$$

Checkpoint 75.

Proposition 5.1.8. powers and transcendental functions. In the following list, if no restrictions are given on x, then the statement holds for all real x.

1. $$\frac{d}{dx}\, x^n = nx^{n-1} \quad \text{when } n \text{ is a positive integer}$$
2. $$\frac{d}{dx}\, x^r = rx^{r-1} \quad \text{when } x \neq 0 \text{ and } r \text{ is any nonzero real number}$$
3. $$\frac{d}{dx}\, e^x = e^x$$
4. $$\frac{d}{dx}\, a^x = a^x \cdot \ln a \text{ for } a > 0 \text{ and all real } x$$
5. $$\frac{d}{dx}\, \ln x = \frac{1}{x} \text{ for } x > 0$$

6. $$\frac{d}{dx} \sin x = \cos x$$

7. $$\frac{d}{dx} \cos x = -\sin x$$

8. $$\frac{d}{dx} \tan x = \sec^2 x \text{ when this is finite}$$

9. $$\frac{d}{dx} \arcsin x = \frac{1}{\sqrt{1-x^2}}$$

10. $$\frac{d}{dx} \arccos x = \frac{-1}{\sqrt{1-x^2}}$$

11. $$\frac{d}{dx} \arctan x = \frac{1}{1+x^2}$$

Checkpoint 76. enn.edu/~ancoop/1070/sec-comp-der.html#project-76)

Checkpoint 77. enn.edu/~ancoop/1070/sec-comp-der.html#project-77)

You are probably pretty experienced at taking apart algebraic expressions into sums and differences of products and quotients of simpler expressions. Here are some more exercises to check that you can do this and then apply the differentiation rules above.

Checkpoint 78. enn.edu/~ancoop/1070/sec-comp-der.html#project-78)

Taking apart algebraic expressions into compositions of functions, as is needed for the chain rule, can be a little trickier.

Example 5.1.9. In order to differentiate $(1+x^2)^{1/3}$ you need to recognize this as a composition $f(g(x))$ with $f(x) = x^{1/3}$ and $g(x) = 1+x^2$. The chain rule tells us that the derivative of $(1+x^2)^{1/3}$ at $x = a$ will be given by

$$\left(\frac{d}{dx} x^{1/3} \bigg|_{x=1+a^2} \right) \cdot \left(\frac{d}{dx} (1+x^2) \bigg|_{x=a} \right). \tag{5.1.1}$$

The derivative of $x^{1/3}$ is $(1/3)x^{-2/3}$ be the power rule (the second identity in Proposition 5.1.8); the derivative of $1+x^2$ is $0+2x = 2x$ by the sum rule and the power rule. This shows (5.1.1) to equal

$$\left(\frac{1}{3} x^{-2/3} \bigg|_{x=1+a^2} \right) \cdot \left(2x|_{x=a} \right) = \frac{1}{3}(1+a^2)^{-2/3}(2a)$$

The next few exercises check on your understanding of the chain rule. The first two tell you how to choose f and g. The last two do not.

Checkpoint 79. enn.edu/~ancoop/1070/sec-comp-der.html#project-79)

Checkpoint 80. enn.edu/~ancoop/1070/sec-comp-der.html#project-80)

Checkpoint 81. enn.edu/~ancoop/1070/sec-comp-der.html#project-81)

Checkpoint 82.

5.2 Arguments and proofs

Proofs are for convincing others, as well as for deciding whether you know something for sure, in all cases. The next two exercises ask for opinions on whether or not a proof is needed. There's no right answer, but we expect you to give a good sense of why or why not.

Checkpoint 83. the sum rule: obvious or not?

Checkpoint 84. the chain rule: obvious or not?

In case some of you answered that it was not obvious, here is a mathematical proof. In most of the upcoming proofs, we need to use the definition of the derivative as a limit of difference quotients We don't need to use the ε-δ definition of limit, just known facts about limits.

Let $h = f + g$. By definition

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

The difference quotient on the right-hand side simplifies to $\frac{f(x)-f(a)}{x-a} + \frac{g(x)-g(a)}{x-a}$. This is a sum of two things. The limit of the sum is the sum of limits, therefore

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

As you can see, the logic broke this down into small steps, justified by facts we have accumulated. The proof didn't add a whole lot to our understanding, although it does help to nail down the fact that this holds whenever $f'(a)$ and $g'(a)$ exist, without exceptions for when one of them is zero, or undefined for values other than a, or anything like that.

We'll ask you to do one of these on your own, then not bother you with proofs of things that are borderline obvious.

Checkpoint 85.

A close up look at the product rule. We mentioned earlier what units a derivative has, but never discussed why. Now is a good time. Taking the limit of an expression gives something with the same units. The derivative is the limit of a difference quotient $(f(x+h) - f(x))/h$. The numerator is the difference between two things with the same units, namely the units of the value of f. The denominator has units of the argument of f. So the difference quotient has units of the value of f divided by the argument of f. For example, if $f(t)$ is distance traveled in the time t, then f' has units of distance per time (such as MPH).

Why is $(fg)'$ not equal to $f'g'$?There are many reasons, one of which is the units. In an application, the values of f and g might have different units, but if both are being differentiated with respect to x then they must have the same input units. The units of $(fg)'$ are, as we have just seen, units of f times units of g divided by units of x, the argument. Unfortunately $f'g'$ has the units of f/x times the units of g/x, so one too many units of x in the denominator.

We now present three arguments for the product rule. When we're done, we'll take a poll of which is most convincing.

If f is a constant, so all the change in the product fg comes from changes in g, then we have seen $(fg)' = f \cdot g'$. If g is a constant, then similarly, $(fg)' = gf'$. In reality, both are changing, so the rate of change of the area is the sum of these two individual rates.

Suppose $f(t)$ is the length in meters of a growing rectangular blob at time t seconds, and $g(t)$ is its width. How fast is the area growing at time t?

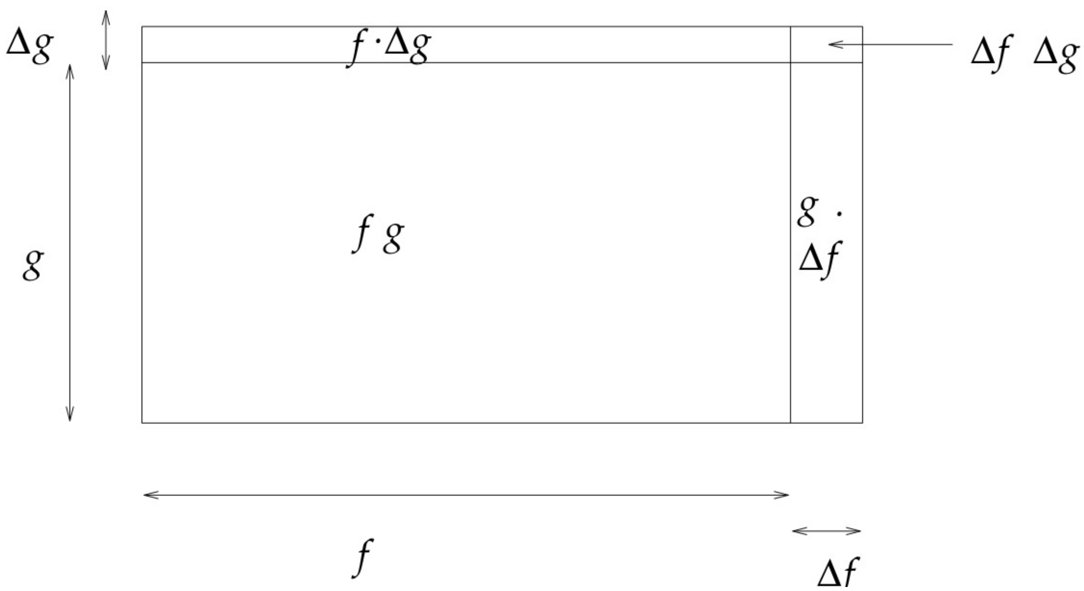


Figure 5.2.1. Pictorial proof of the product rule

Figure 5.2.1 shows the classical pictorial argument. When time increases by a small quantity Δt, both f and g increase by small quantities, which we respectively call Δf and Δg, and the area increases by $f\Delta g$ plus $g\Delta f$ plus $(\Delta f)(\Delta g)$. We know that Δf is approximately $f'(t)\Delta t$, because in the limit as $\Delta t \to 0$, the ratio $\Delta f/\Delta t$ converges to $f'(t)$. Similarly, $\Delta g \approx g'(t)\Delta t$. From the picture, you can see that $\Delta(fg) = f\Delta g + g\Delta f + (\Delta g)(\Delta f)$. So

$$\frac{\Delta fg}{\Delta t} = f\frac{\Delta g}{\Delta t} + g\frac{\Delta f}{\Delta t} + \frac{(\Delta f)(\Delta g)}{\Delta t}$$

Taking limits on the right hand side as $\Delta t \to 0$ gives $f'g + g'f + \lim_{\Delta t \to 0}(\Delta f)(\Delta g)/\Delta t$. This last limit should be zero. Why? Say $f'(t) = a$ and $g'(t) = b$. Then $\Delta f \approx a\Delta t$ and $\Delta g \approx b\Delta t$, so

$$\lim_{\Delta t \to 0} \frac{(\Delta f)(\Delta g)}{\Delta t} \approx \lim_{\Delta t \to 0} \frac{f'(t)(\Delta t)g'(t)(\Delta t)}{\Delta t} = \lim_{\Delta t \to 0} f'(t)g'(t)(\Delta t)$$

which is zero.

The simplest algebraic proof of the product rule is a bit more out of the blue because it relies on this trick:

$$f(x+h)g(x+h) - f(x)g(x) = f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)$$

and hence

$$\frac{f(x+h)g(x+h)-f(x)g(x)}{h} = f(x+h)\frac{g(x+h)-g(x)}{h} + g(x)\frac{f(x+h)-f(x)}{h}$$

The trick was, we added and subtracted $f(x+h)g(x)$ in order to be able to separate the original difference quotient into two pieces, both of which look a function times a simpler difference quotient. Taking limits and using the fact that limits of sums are sums of limits, and the same for products, gives

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

Checkpoint 86.

A physics proof of the derivative of the sine function. Suppose a toy car is moving around a circular track of radius one meter, so that its speed is constant 1 meter per second; the coordinates of the point are $x = \cos t, y = \sin t$. By definition of radian, its angle with respect to the horizontal increases at a rate of one radian per second. The northward (y-direction) speed is the derivative of $\sin t$. Suppose at time x a gate opens up and the car stops turning to stay on the track and coasts straight onward at its present speed of 1. Its northward speed during the time $[x, x+1]$ is the derivative of the sine function at time x. To evaluate this, we just have to check how far northward the car went from time x to $x+1$. This is just analytic geometry. The car goes one unit tangent to the circle during this time interval from the point $(\cos x, \sin x)$ (B in Figure 5.2.2) to the point $(\cos t - \sin t, \sin t + \cos t)$ (A in the figure). Therefore the derivative of $\sin$ is $\cos$. For free, we also get (by looking at the x coordinate) that the derivative of $\cos$ is $-\sin$.

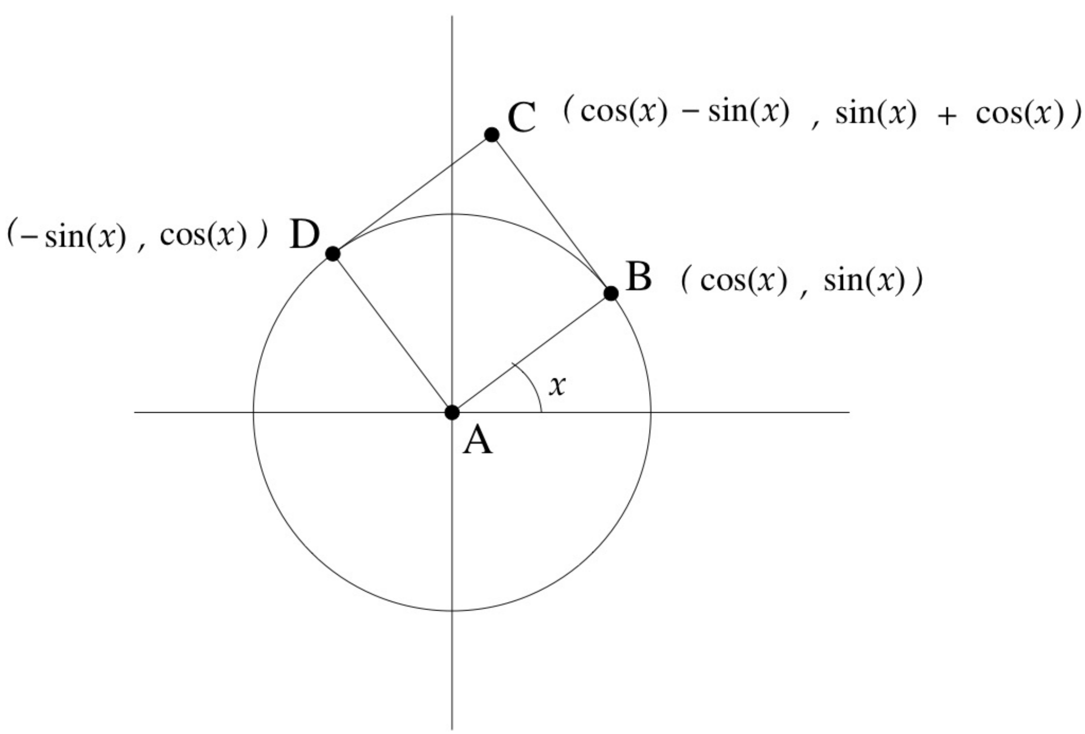


Figure 5.2.2. ABCD is a square of side 1 tangent to the unit circle as shown. At time x the car is at point B, making angle x with the x-axis. From time x to time $x+1$ the

car travels in a straight line to C.

The chain rule. The easiest way to make sense of the chain rule is in terms of related rates. Think of x, u and y as physical quantities related by rules. If you change x, it changes u. The specific rule is $u = g(x)$. If you change u it changes y. The specific rule is $y = f(u)$.

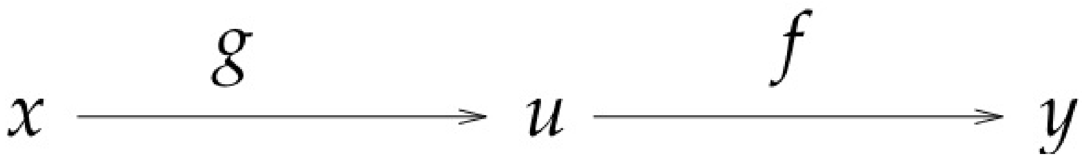


Figure 5.2.3.

What does this mean quantitatively? The rate of change of u with respect to x is $g'(x)$. This is illustrated on the left side of Figure 5.2.4, where the infinitesimal changes dx and du are depicted. The slope of the hypotenuse of the small triangle is $g'(x)$, where in the diagram, the value of x is roughly $1/2$. On the right side of the figure, we see that this small change in u leads to a proportionate small change in y. The ratio, dy/du is equal to $f'(u)$. One question remains: at what value of u is this ratio evaluated? In the figure, it appears $u \approx 1/8$. More precisely, if we originally took x to be $1/2$, the u value will be $f(1/2)$. In other words, the value from the u-axis (vertical in the first graph) is copied to the second graph (where the u-axis is now the horizontal axis). In other words, f' is evaluated at u, which is $g(x)$. Thus $dy/dx = du/dx \cdot dy/du|_{u=g(x)}$.

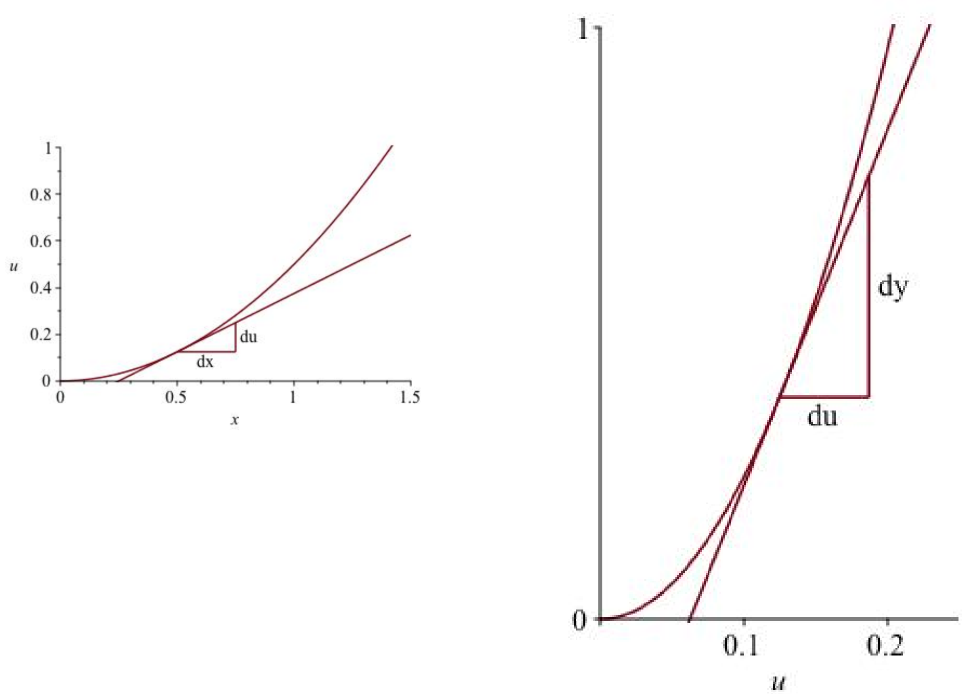


Figure 5.2.4. x affects u, which in turn affects y

If we want to make this into a formal proof, we might start by writing

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

If $g(a+h)$ could be replaced by the tangent line approximation $g(a) + hg'(a)$ then the proof would finish easily: letting $\varepsilon := hg'(a)$,

$$\lim_{h \to 0} \frac{f(g(a) + hg'(a)) - f(g(a))}{h} = \lim_{\varepsilon \to 0} \frac{f(g(a) + \varepsilon) - f(g(a))}{(\varepsilon / g'(a))} = g'(a) f'(g(a))$$

It is indeed true that the tangent line approximation is close enough to g itself to make this work, but proving that takes a trickier argument than we want to go into here.

∑ () **MATH 1070: Mathematics of Change, part I (article-1.html)**

Calculus Group

5 Asymptotic Analysis and L'Hopital's Rule

When you learn about complex numbers, they seem in one sense like make-believe but in another sense like ordinary math because they obey clear rules. Learning about infinity is different. The word is in the vocabulary of most children, but no one knows the rules. Is infinity part of math? Part of philsophy? Science fiction? It turns out infinity does obey some very clear rules, as long as you decide to define it as a limit. (Trust mathematicians to take the fun out of it!)

Suppose, in addition to the real numbers, we include the numbers $+\infty$, $-\infty$ and UND (for undefined). These are the possible limits a function can have. The goal is to create combining rules for limits under the basic operations: addition, subtraction, multiplication, division and taking powers. One rule is that once something is undefined, it stays that way. Limits that DNE could turn out to be $\pm\infty$ rather than UND, but once a limit gets classified as UND, nothing can be inferred about what you get when you add it to something, multiply it, etc. Thus, UND + 3, UND $-\infty$, $-\infty\cdot$ UND and UND / UND are all undefined.
For the remaining situations, the definition is as follows.

De nition 6.0.1. If a, b and L are real numbers or $\pm\infty$, and $\odot$ is an operation, we say $a \odot b = L$ if for every f and g such that $\lim_{x\to 0} f(x) = a$ and $\lim_{x\to 0} g(x) = b$, it is always true that $\lim_{x\to 0} f(x) \odot g(x) = L$.

Example 6.0.2. $\infty + 3$. The definition allows us to check if any guess is right. Let's guess $\infty + 3 = \infty$. To check this, we check whether $\lim_{x\to a} f(x) + g(x)$ is always equal to ∞ when f and g are functions for which $\lim_{x\to a} f(x) = \infty$ and $\lim_{x\to a} g(x) = 3$.

Indeed, if $f(x)$ gets larger than any specified number M when x gets close to a and $g(x)$ gets close to~3, then $f(x) + g(x)$ will get larger than $M + 3$, in other words, will get larger than any given number, which is the definition of the limit being $+\infty$.

We conclude that

$$``\infty + 3 = \infty''.$$

\infty + 3 = \infty". \end{equation*}

Checkpoint 87.

6.1 L'Hôpital's Rule

L'Hôpital's Rule allows us resolve UND limits in some cases. In other words, it can determine a limit of an expression such as $f+g$ or f/g or f^g, etc., when this limit is not determined just by knowing the limit of f and the limit of g. These cases of UND limits are often called ***indeterminate forms***. For example, $0/0$ is an indeterminate form because when $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$, the limit $\lim_{x\to a} f(x)/g(x)$ can turn out to be any real number, or $+\infty$, or $-\infty$, or undefined. The basic version of L'Hôpital's Rule involves just the one indeterminate form $0/0$.

L'Hôpital's Rule is named for Guillaume François Antoine, Marquis de l'Hôpital, a French noble and mathematician who worked in the late 1600s. L'Hôpital was the author of the world's very first calculus textbook.

Theorem 6.1.1. L'Hôpital's Rule, first version. Let f and g be functions differentiable on an interval containing the point a, except possibly at the point a, where f and g are not required to be defined. Suppose f and g both have limit zero at a and suppose g' is nonzero on the interval. If $\lim_{x\to a} f'(x)/g'(x) = L$ for some finite L, then the limit $\lim_{x\to a} f(x)/g(x)$ exists and is equal to L.

Example 6.1.2. L'Hôpital's Rule computes $\lim_{x\to 0} \sin(x)/x$ much more easily than you did in your homework. Let $f(x) = \sin x$, $g(x) = x$ and $a = 0$ and observe that the continuous functions f and g both vanish at zero, hence $\lim_{x\to 0} f(x) = \lim_{x\to 0} g(x) = 0$. Therefore,

$$\lim_{x\to 0} \frac{\sin x}{x} = \lim_{x\to 0} \frac{\cos x}{1} = \frac{\cos(0)}{1} = 1\,.$$

You might wonder, when we first evaluated this limit, why did we do it the hard way? Remember, we did not and will not prove L'Hôpital's Rule. For this reason it's good to see some things that can be done without it.

Checkpoint 88.

There are two common mistakes in applying L'Hôpital's Rule. One is trying to use it the other way around. If f/g has a limit at a, that doesn't mean f'/g' does, or that these even exist. The other is to try to use it when f or g has a nonzero limit at a. For example, if $\lim_{x\to a} f(x) = 5$ and $\lim_{x\to a} g(x) = 3$ then $\lim_{x\to a} f(x)/g(x) = 5/3$ (the nonzero quotient rule) and is probably not equal to $\lim_{x\to a} f'(x)/g'(x)$.

Checkpoint 89.

More general versions. If the hypotheses hold only from one side, for example $\lim_{x\to a^+} f(x) = \lim_{x\to a^+} g(x) = 0$, then the conclusion still holds on that side: if $\lim_{x\to a^+} f'(x)/g'(x) = L$ then $\lim_{x\to a^+} f(x)/g(x) = L$. Also, the limit can be taken at $\pm\infty$ and nothing changes.

Proposition 6.1.3. Improved L'Hôpital's Rule.

1. *Suppose f and g are differentiable on an open interval (a, b), with f and g both having limit zero at a. Suppose that $g' \neq 0$ on (a, b) and $\lim_{x\to a^+} f'(x)/g'(x) = L$. Then $\lim_{x\to a^+} f(x)/g(x) = L$.*

2. *Suppose f and g are differentiable on an open interval (b, a) with f and g both having limit zero at a. Suppose that $g' \neq 0$ on (b, a) and $\lim_{x \to a^-} f'(x)/g'(x) = L$. Then $\lim_{x \to a^-} f(x)/g(x) = L$.*
3. *Suppose f and g are differentiable on an open interval (b, ∞) with f and g both having limit zero at infinity. Suppose that $g' \neq 0$ on (b, ∞) and and $\lim_{x \to \infty} f'(x)/g'(x) = L$. Then $\lim_{x \to \infty} f(x)/g(x) = L$. The same holds for limits at $-\infty$, replacing the interval with $(-\infty, b)$.*

Checkpoint 90. (https://www.math.upenn.edu/~ancoop/1070/section-6.html#project-90)

Turning other indeterminate forms into $0/0$. For each of the other indeterminate forms, there's a trick to turn it into the basic form 0/0. (https://www.math.upenn.edu/~ancoop/1070/section-6.html#p-909)

The case $0 \cdot \infty$

Suppose $\lim_{x \to a} f(x) = 0$ and $\lim_{x \to a} g(x) = \infty$. How can we compute $\lim_{x \to a} f(x) \cdot g(x)$? We know that $\lim_{x \to a} 1/g(x) = 1/\infty = 0$. Therefore, an easy trick is to replace multiplication by g with division by $1/g$. Letting h denote $1/g$, we have

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

which is the correct form for L'Hôpital's Rule.

Example 6.1.4. What is $\lim_{x \to 0^+} x \cot x$? Letting $f(x) = x$ and $g(x) = \cot x$ we see this has the form $0 \cdot \infty$. Letting $h(x) = 1/g(x) = \tan x$ we see that

$$\lim_{x \to 0^+} x \cot x = \lim_{x \to 0^+} \frac{x}{\tan x} = \lim_{x \to 0^+} \frac{x}{\sin x} \cdot \cos x \, .$$

The limit at 0 of $x/\sin x$ is~1 and the limit of the continuous function $\cos x$ is $\cos(0) = 1$, therefore the answer is $1 \cdot 1 = 1$.

The case ∞/∞

You could invert both f and g, writing $\frac{f(x)}{g(x)}$ as $\frac{1/g(x)}{1/f(x)}$. There is a reasonable chance that L'Hôpital's Rule can be applied to this. There is also another version of L'Hôpital's Rule specifically for this case.

Theorem 6.1.5. L'Hôpital's Rule for ∞/∞. If both f and g tend to ∞ or $-\infty$ as $x \to a$, then

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

whenever the right-hand side is a real number or $\pm\infty$.

Checkpoint 91. (https://www.math.upenn.edu/~ancoop/1070/section-6.html#project-91)

The cases 1^∞, 0^0 and ∞^0

The idea with indeterminate powers is to take the log, compute the limit, then exponentiate. The reason this works is that e^x is a continous function. Theorem 3.4.5 says that if $\lim_{x \to a} h(x) = L$ then $\lim_{x \to a} e^{h(x)} = e^L$.

The way we will use this when evaluating something of the form $\lim_{x\to a} f(x)^{g(x)}$ is to take logarithms. Algebra tells us $\ln f(x)^{g(x)} = g(x)\ln f(x)$. If we can evaluate $\lim_{x\to a} g(x)\ln f(x) = L$ then we can exponentiate to get $\lim_{x\to a} f(x)^{g(x)} = e^L$.

Checkpoint 92.

Example 6.1.6. continuous compounding.
Suppose you have a million dollars earning a 12\% annual interest rate for one year. You might thing after a year you will have 1.12 million dollars. But no, things are better than that. The bank compounds your interest for you. They realize you could have cashed out after half a year with 1.06 million and reinvested for another half year, giving you 1.1236 million, which doesn't seem so different but is actually 3600 dollars more. You could play this game more frequently, dividing the year into n periods and earning $12\%/n$ interest n times, so your one million becomes $(1+0.12/n)^n$ million.

With computerized trading, you could make the period of time a second, or even a microsecond. Does this enable you to claim an unboundd amount of money after one year? To answer that, let's compute the amount you would get if you compounded {\mathbb{E}m continuously}, namely $\lim_{n\to\infty}(1+0.12/n)^n$. Taking logs gives $\ln(1+0.12/n)^n = n\ln(1+0.12/n)$. Changing to the variable $x := 1/n$,

$$\begin{aligned}\lim_{n\to\infty} n\ln(1+0.12/n) &= \lim_{x\to 0}\frac{\ln(1+0.12x)}{x} \\ &= \lim_{x\to\infty}\frac{(d/dx)\ln(1+0.12x)}{(d/dx)x} \quad \text{(L'Hôpita} \\ &= \lim_{x\to\infty}\frac{0.12/(1+0.12x)}{1} \quad \text{(use the cha} \\ &= 0.12\end{aligned}$$

Therefore, $\lim_{n\to\infty}(1+0.12/n)^n = e^{0.12} \approx 1.12749685$ million dollars. That's better than the \$120,000 you earn without compounding, or the \$3,600 more than that you earn compounding once, but it's not infinite, it's just another \$3896.85 better.

Checkpoint 93.

Repeated use of L'Hôpital's Rule. Sometimes when trying to evaluate $\lim_{x\to a} f(x)/g(x)$ you find that $\lim_{x\to a} f'(x)/g'(x)$ appears a bit simpler, but you still can't tell what it is. You might try L'Hôpital's Rule twice. If $f'(x)$ and $g'(x)$ tend to zero as $x\to a$ (if they don't, you can probably tell what the limit is), then you can use f' in place of f and g' in place of g in L'Hôpital's Rule. If you can evaluate the limit of $f''(x)/g''(x)$ then this must be the limit of $f(x)/g(x)$. You can often do a little better if you simplify $f'(x)/g'(x)$ to get a new numerator and denominator whose derivatives will be less messy.

Example 6.1.7. Repeated L'Hôpital's Rule makes another limit that was formerly painful into a piece of cake: $\lim_{x\to 0}(1-\cos x)/x^2$. Both numerator and denominator are zero at zero, so we apply L'Hôpital's Rule to see that the limit is equal to $\lim_{x\to 0}\sin x/(2x)$. You can probably remember what this is, but in case not, one more application of L'Hôpital's Rule shows it to be equal to $\lim_{x\to 0}\cos x/2 = \cos(0)/2 = 1/2$.

Checkpoint 94.

6.2 Orders of growth at infinity

Often in mathematical modeling, one hears statements such as "This model produces a much smaller growth rate than the other model, as time gets large." This statement sounds vague: how much is "much smaller" and what are "large times"? In this section we will give a precise meaning to statements such as this one.

Why are we spending our time making a science out of vague statements? Answer:

1. people really think this way, and it clarifies your thinking to make these thoughts precise;
2. a lot of theorems can be stated with these as hypotheses;
3. knowing the science of orders of growth helps to fulfill the Number Sense mandate because you can easily fit an unfamiliar function into the right place in the hierarchy of more familiar functions.

We focus on two notions in particular: when one function is ***much*** bigger/smaller/closer than another, and when two functions are ***asymptotically equal***.

Comparisons at infinity. Mostly we will be comparing functions of x as $x \to \infty$. Let f and g be positive functions.

Definition 6.2.1.

1. We say the function f is ***asymptotic to*** the function g, short for "asymptotically equal to", if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1\,.$$

   This is denoted $f \sim g$.

2. The function f is said to be ***much*** smaller than g, or to ***grow much more slowly*** if

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

   This is denoted $f \ll g$. Typically this notation is used only when g is positive.

Example 6.2.2. Is it true that $x^2 + 3x$ is asymptotically equal to x^2? Intuitively it should be true because $3x$ is a lot smaller than x^2 when x is large (in fact, it is *much smaller*) so adding it to x^2 should be negligible. We check that

$$\lim_{x\to\infty} \frac{x^2 + 3x}{x^2} = \lim_{x\to\infty} 1 + \frac{3}{x} = 1\,,$$

therefore indeed $x^2 + 3x \sim x^2$.

Checkpoint 95.

Example 6.2.3. Let's compare two powers, say x^3 and $x^{3.1}$. Are they asymptotically equivalent or does one grow much faster? Taking the limit at infinity we see that $\lim_{x\to\infty} x^3/x^{3.1} = \lim_{x\to\infty} x^{-0.1} = 0$. Therefore, $x^3 \ll x^{3.1}$. This is shown on the left side of Figure 6.2.4.

Figure 6.2.4. Comparing growth rates. The green graph is $y = ax^r$; the blue is $y = bx^s$. Be sure to zoom in and out.

Try using the tool in Figure 6.2.4 to compare $10x^{1.3}$ with $0.1x^{1.5}$? It looks at first like $10x^{1.3}$ remains much greater than $0.1x^{1.5}$. Doing the math gives

$$\lim_{x\to\infty} \frac{10x^{1.3}}{0.1x^{1.5}} = \lim_{x\to\infty} 100x^{-0.2} = 0\,.$$

Therefore, again, $10x^{1.3} \ll 0.1x^{1.5}$. Whether or not you care what happens beyond 10^{41} depends on the application, but the math is pretty clear: if $a < b$, then $Kx^a \ll Lx^b$ for any positive constants K and L.

Discussion. This is a general rule: the function $g(x) + h(x)$ will be asymptotic to $g(x)$ exactly when $h(x) \ll g(x)$. Why? Because $(g(x) + h(x))/g(x)$ and $h(x)/g(x)$ differ by precisely 1. It follows that if $g(x) + h(x) \sim g(x)$ then

$$1 = \lim_{x\to\infty} \frac{g(x)+h(x)}{g(x)} = \lim_{x\to\infty} 1 + \frac{h(x)}{g(x)} = 1 + \lim_{x\to\infty} \frac{h(x)}{g(x)} \qquad \text{hence} \qquad \lim_{x\to\infty} \frac{h(x)}{g(x)} = 0\,,$$

or in other words, $h \ll g$. The chain of identies runs backward as well: $g + h \sim g$ if and only if $h \ll g$.

Another principle is that if $f \sim g$ and $h \sim \ell$ then $f \cdot h \sim g \cdot \ell$. This is literally just the product rule for limits. The same is true for nonzero quotients, for the same reason.

Example 6.2.5. We know $x + 1/x \sim x$ and $2 - e^{-x} \sim 2$, therefore

$$\frac{x+1/x}{2-e^{-x}} \sim \frac{x}{2}\,.$$

These two facts give important techniques for estimating. They allow you to clear away irrelevant terms: in any sum, every term that is much less than one of the others can be eliminated and the result will be asymptotic to what it was before. You can keep going with products and quotients.

Example 6.2.6. Find a nice function asymptotically equal to $\sqrt{x^2+1}$. The notion of "nice" is subjective; here it means a function you're comfortable with, can easily estimate, and so forth.

> *Because $1 \ll x^2$ we can ignore the~1 and get $\sqrt{x^2}$ which is equal to x for all positive x. Therefore, $\sqrt{1+x^2} \sim x$.*

Beware though, if $f \sim g$ and $h \sim \ell$, it does not follow that $f + h \sim g + \ell$ or $f - h \sim g - \ell$. Why? See the following self-check exercise.

Checkpoint 96. enn.edu/~ancoop/1070/section-6.html#project-96)

It should be obvious that the relation $\sim$ is symmetric: $f \sim g$ if and only if $g \sim f$. Formally,

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1 \iff \lim_{x\to\infty} \frac{g(x)}{f(x)} = 1$$

because one is the reciprocal of the other. On the other hand, the relation $f \ll g$ is anti-symmetric: it is not possible that both $f \ll g$ and $g \ll f$.

It is good to have an understanding of the relative sizes of common functions. Here is a summary of some basic facts from this lesson, practice problems and homework problems.

Proposition 6.2.7. edu/~ancoop/1070/section-6.html#proposition-14)

1. *Positive powers all go to infinity but at different rates, with the higher power growing faster.*
2. *Exponentials grow at different rates and every exponential grows faster than every power.*
3. *Logarithms grow so slowly that any power of* $\ln x$ *is less than any positive power of* x.

6.3 Comparisons elsewhere and orders of closeness

Everything we have discussed in this section has referred to limits at infinity. Also, all our examples have been of functions getting large, not small, at infinity. But we could equally have talked about functions such as $1/x$ and $1/x^2$, both of which go to zero at infinity. It probably won't surprise you to learn that $1/x^2$ is much smaller than $1/x$ at infinity.

Checkpoint 97.

These same notions may be applied elsewhere simply by taking a limit as $x \to a$ instead of as $x \to \infty$. The question then becomes: is one function much smaller than the other as the argument approaches a? In this case it is more common that both functions are going to zero than that both functions are going to infinity, though both cases do arise. Remember: at a itself, the ratio of f to g might be $0/0$ or ∞/∞, which of course is meaningless, and can be made precise only by taking a limit as x approaches a.

The notation, unfortunately, is not built to reflect whether $a = \infty$ or some other number. So we will have to spell out or understand by context whether the limits in the definitions of $\ll$ and $\sim$ are intended to occur at infinity or some other specificed location, a.

Example 6.3.1. Let's compare x and x^2 at $x = 0$. At infinity, we know $x \ll x^2$. At zero, both go to zero but at possibly different rates. Have a look at Figure 6.3.2. You can see that x has a postive slope whereas x^2 has a horizontal tangent at zero. Therefore, $x^2 \ll x$ as $x \to 0^+$. You can see it from Figure 6.3.2 or from L'H{\^o}pital:

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

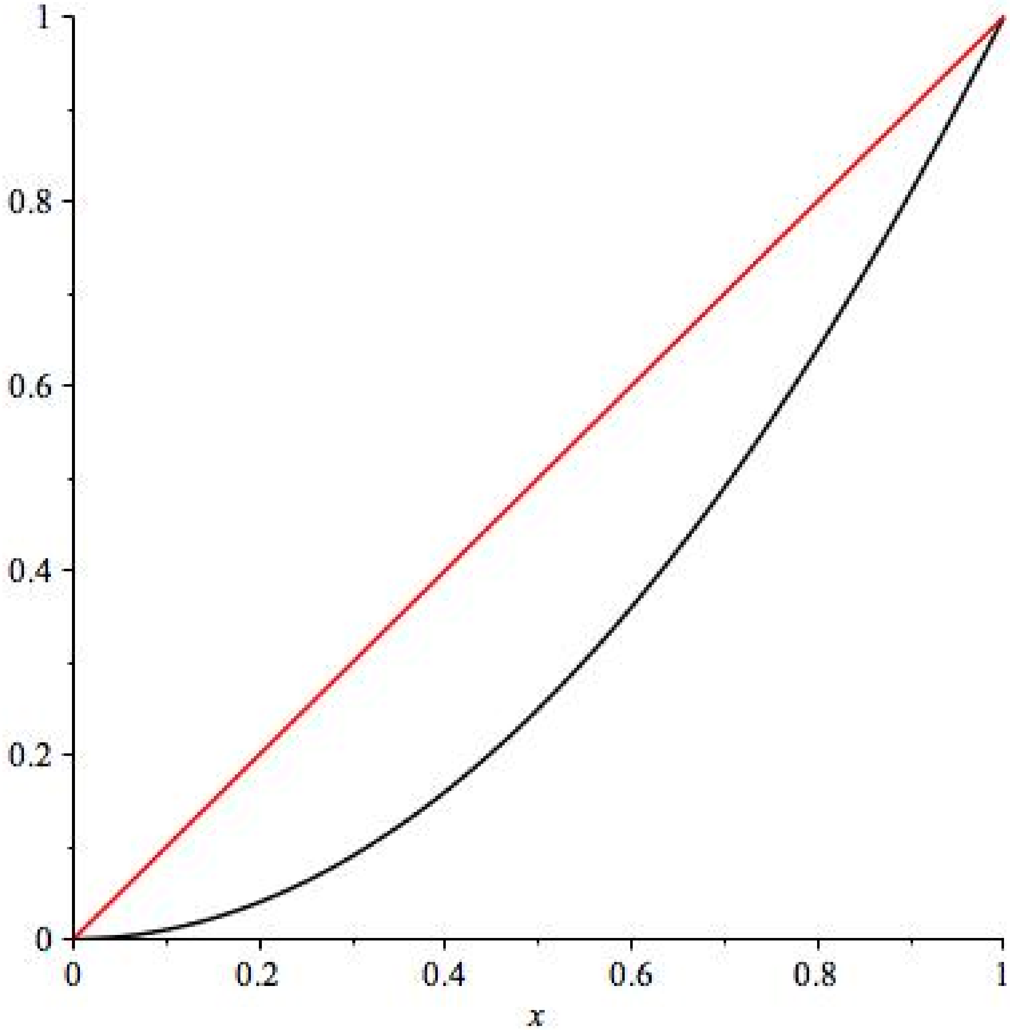


Figure 6.3.2. Comparing x (red) and x^2 (black) at $x = 0$

Example 6.3.3. What about x^2 and x^4 near zero? Both have slope zero. By eye, x^4 is a lot flatter. Maybe $x^4 \ll x^2$ near zero. It is not clearly settled by the picture (do you agree? see Figure 6.3.4), but the limit is easy to compute.

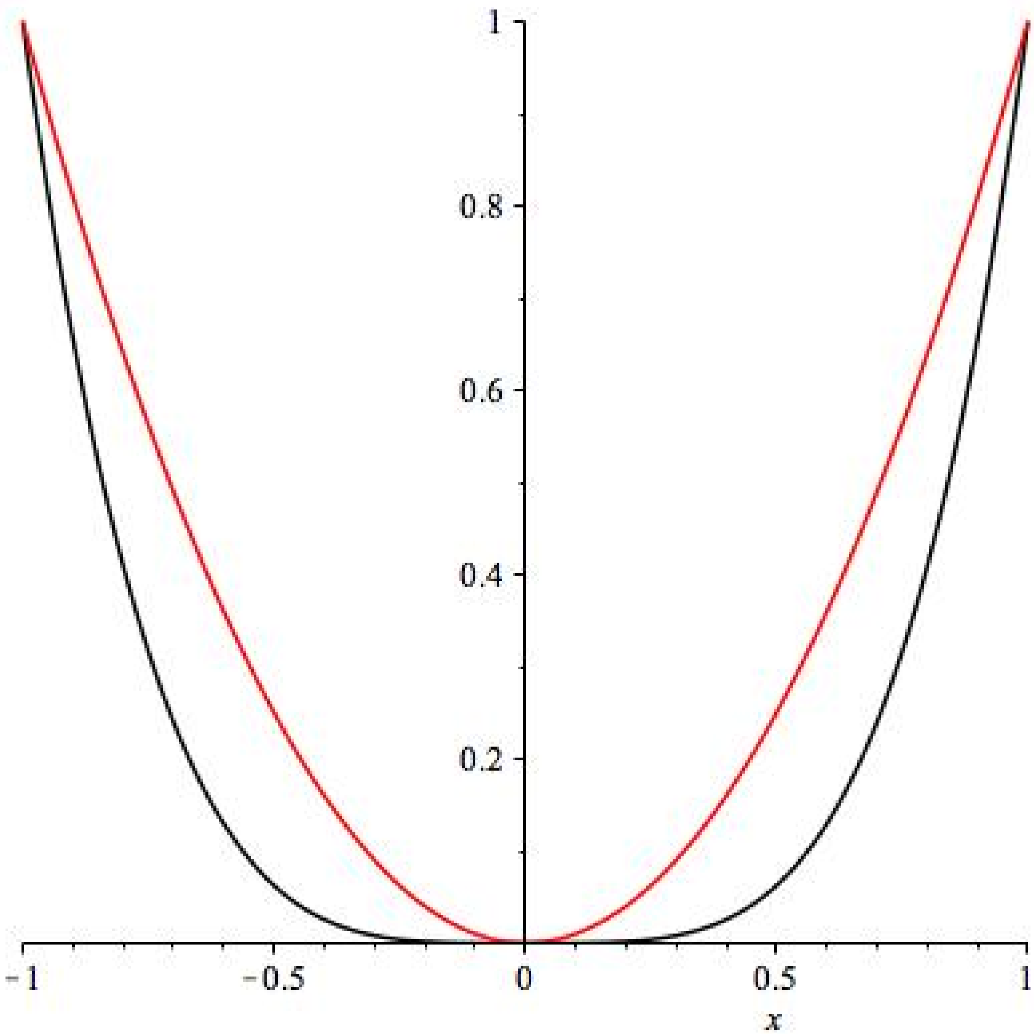


Figure 6.3.4. Comparing x^2 (red) and x^4 (black) at $x = 0$

Checkpoint 98.

Here is a less obvious example, still with powers.

Example 6.3.5. Let's compare $\sqrt{x}$ and $\sqrt[3]{x}$ near zero. See Figure 6.3.6. Is one of these functions much smaller than the other as $x \to 0^+$? Here, the picture is pretty far from giving a definitive answer!

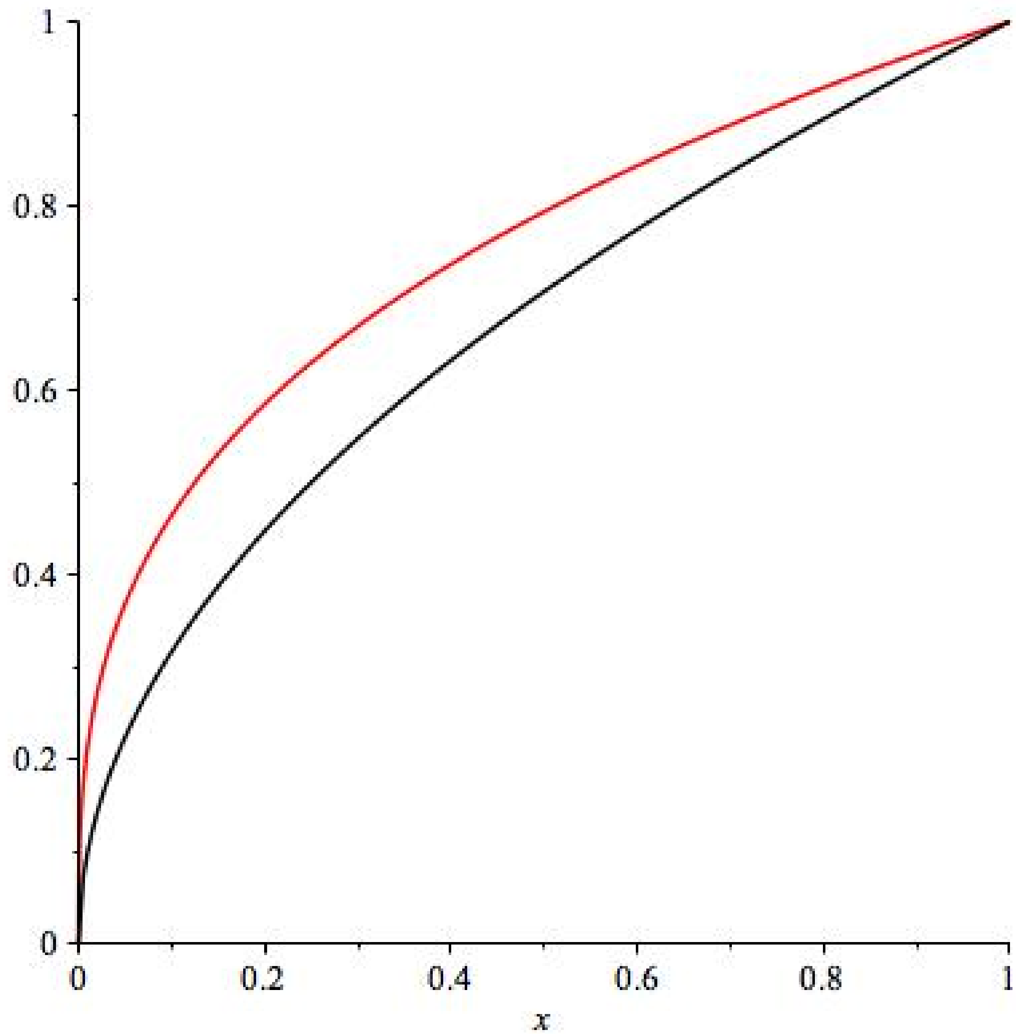


Figure 6.3.6. Comparing $\sqrt{x}$ (black) and $\sqrt[3]{x}$ (red) at $x = 0$

We try evaluating the ratio: $f(x)/g(x) = x^{1/2}/x^{1/3} = x^{1/2-1/3} = x^{1/6}$. Therefore,

$$\lim_{x\to 0^+} \frac{f(x)}{g(x)} = \lim_{x\to 0^+} x^{1/6} = 0$$

and indeed $x^{1/2} \ll x^{1/3}$. Intuitively, the square root of x and the cube root of x both go to zero as x goes to zero, but the cube root goes to zero a lot slower (that is, it remains bigger for longer).

Checkpoint 99.

Suppose f and g are two nice functions, both of which are supposed to be approximations to some more complicated function H near the argument a. The question of whether $f - H \ll g - H$, or $g - H \ll f - H$, or neither as $x \to a$ is particularly important because it tells us whether one of the two functions f and g is a much better approximation to H than is the other. We will be visiting this question shortly in the context of the tangent line approximation, and again later in the context of Taylor polynomial approximations.

"For sufficiently large x". Often when discussing comparisons at infinity we use the term "for sufficiently large x". That means that something is true for every value of x greater than some number M (you don't necessarily know what M is). For example, is it true that $f \ll g$ implies $f < g$? No, but it implies $f(x) < g(x)$ for sufficiently large x. Any limit at infinity depends only on what happens for sufficiently large x.

Example 6.3.7. We have seen that $\ln x \ll \sqrt{x-5}$. It is not true that $\ln 6 < \sqrt{6-5}$ (the corresponding values are about 1.8 and 1) and it is certainly not true that $\ln 1 < \sqrt{1-5}$ because the latter is not even defined. But we can be certain that $\ln x < \sqrt{x-5}$ for sufficiently large x. The crossover point is between 10 and 11.

6 Optimization

In this section, we'll apply the tool of differentiation to a broad sort of problem called ***optimization***. The idea is, given a mathematical model, to come up with the largest, smallest, cheapest, most profitable, lowest-risk, fastest, safest, or whatever -est word applies to the situation.

Many of you have seen max-min problems before. If not, pay attention! Finding the maximum or minimum of a function is one of the crowning achievements of calculus, with applications in literally any field where mathematical modeling is done.

7.1 Definitions of Minima and Maxima, and their existence

The following definitions give precise meaning to notions you have probably already seen. Some vocabulary may be new but none of it is rocket science.

Definition 7.1.1.

minimum A point $x \in [a,b]$ such that $f(x) \leq f(y)$ for all $y \in [a,b]$ is where f achieves its minimum (plural: minima). The value $f(x)$ is the minimum. This is also called a ***global*** or ***absolute minimum on*** $[a,b]$.

maximum A point $x \in [a,b]$ such that $f(x) \geq f(y)$ for all $y \in [a,b]$ is where f achieves its maximum (plural: maxima). This is also called a ***global*** or ***absolute maximum on*** $[a,b]$. The value $f(x)$ is the maximum.

extremum The word for a something that is a minimum or maximum is extremum (plural: extrema) or extreme value.

local minimum A value x such that $f(x) \leq f(y)$ for all y in some open interval I containing x, which could be a lot smaller than the whole interval (a,b), is where f achieves its local minimum. The terms ***local maximum*** and ***local extremum*** are defined analogously.

critical point a point where f' is zero or undefined.

Note 7.1.2. A somewhat subtle but important note about the language here: the ***input value*** x is *where the extremum is achieved*; the extremum itself is always an ***output value***.

In applications, we sometimes want to know just *what the maximum is* (an output); sometimes just *where the maximum occurs* (an input); and sometimes both. By way of example:

- If I want to build a building to house my flying squirrels, I need to know *what* the maximum height they're capable of flying is, but I don't really care when they get to that height.
- If I need to build a window which admits the most possible light, what I care about is how to set the dimensions (an input), but the amount of light actually let in (in lumens, say) isn't really needed.
- If I'm running a widget factory and I want to know what production level will maximize my profit, the input *where* the maximum occurs (a number of widgets per hour) is important, but for fiscal planning I also need to know *what* that maximum (a number of dollars) actually is.

As you may have noticed, we'll reserve the word *what* to refer to the extremum itself (the output value), and talk about *when*, *where*, or *how* that extremum occurs.

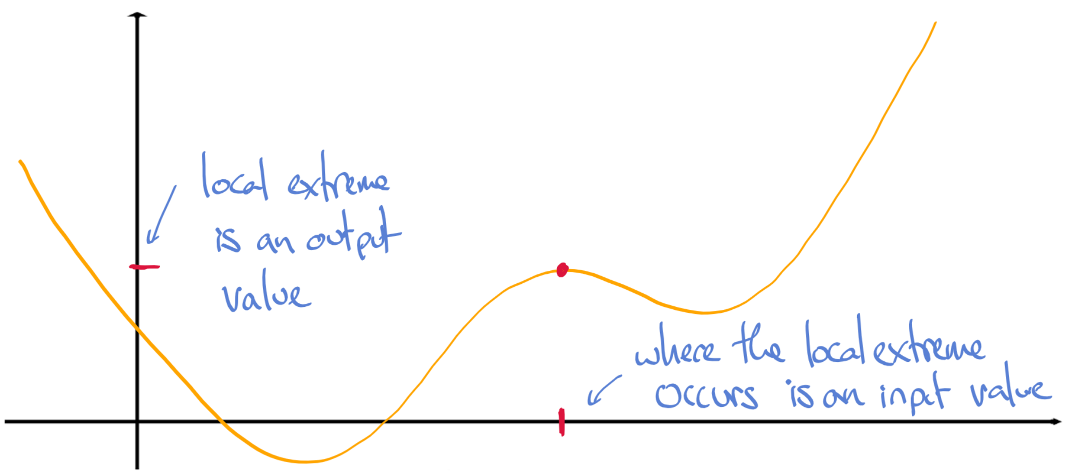


Figure 7.1.3. Extrema *are* output values; they *occur at* input values.

Before we start looking for extrema, it might occur to you to question whether they exist. Better not to go on a wild goose chase.

Checkpoint 100.

Now that you have seen some scenarios where functions have no absolute extrema on an interval, here is a theorem guaranteeing the opposite.

Theorem 7.1.4. *Let f be a continuous function on the closed interval $[a, b]$. Then f has at least one absolute minimum on $[a, b]$ and at least one absolute maximum on $[a, b]$.*

Like the Intermediate Value Theorem, this theorem requires mathematical analysis to prove; we would say it was obvious were it not for the counterexamples we paraded by you in the self-check exercises!

Checkpoint 101.

7.2 The first derivative and extrema

Theorem 7.2.1. Fermat's Theorem. *Suppose a function f has a minimum at a point c in some open interval I. If f is differentiable at c then $f'(c) = 0$.*

The proof of this theorem is accessible and conceptually relevant. This result should seem very credible an intuitive level. If $f'(c) > 0$ then moving to the left from c to $c - \varepsilon$ should produce a greater value of f. Likewise, if $f'(c) < 0$ then moving to the right should produce a greater value. This is the most intuitive justification we could write down, though not exactly airtight.

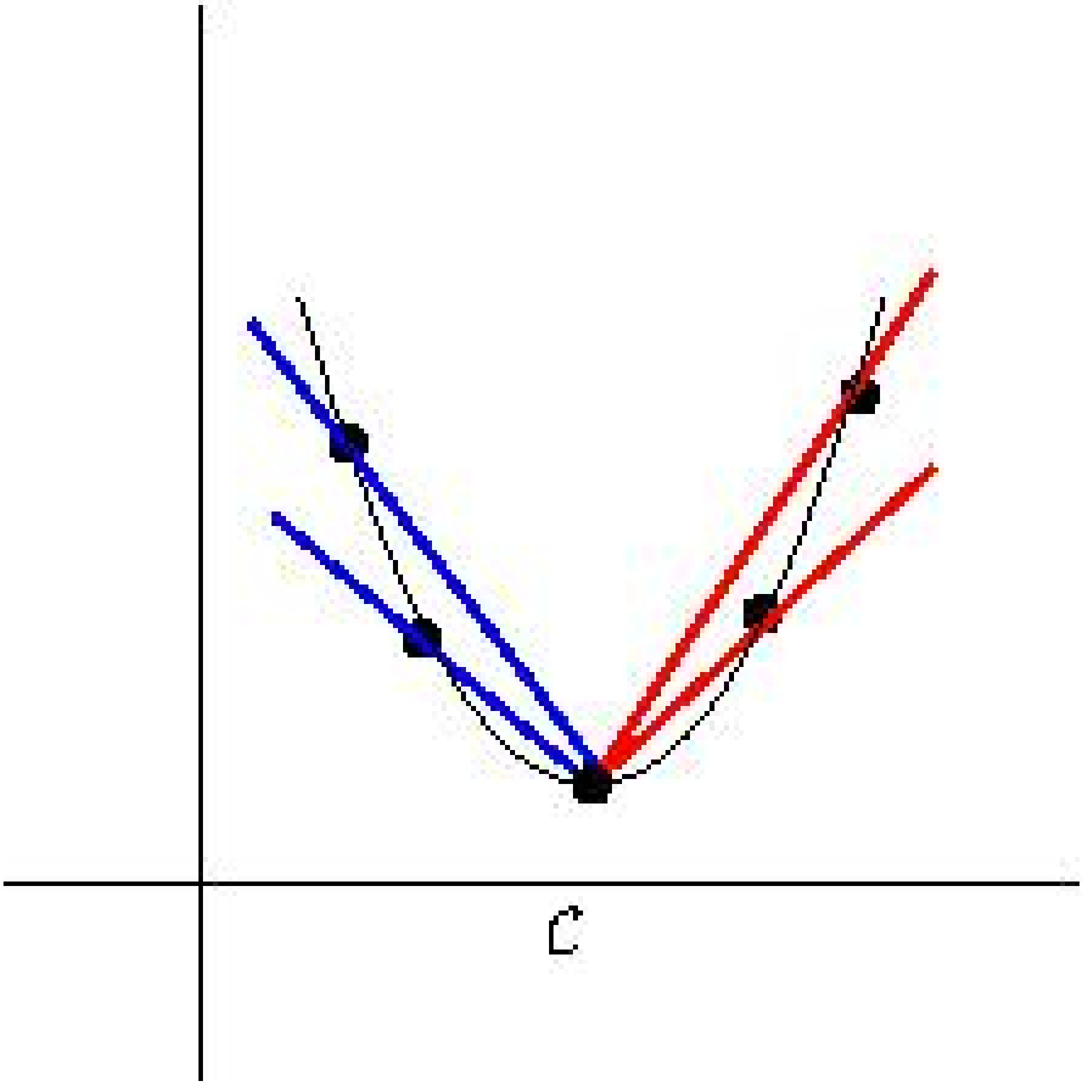


Figure 7.2.2. difference quotients between c and points to the right (red) are positive; those to the left (blue) are negative

Here is a more airtight argument. Because f is differentiable at c, the one-sided derivatives exist and are equal. The derivative from the right is $\lim_{x \to c^+} \frac{f(x)-f(c)}{x-c}$; because c is a minimum, both top and bottom of this fraction are positive (the numerator could be zero). The limit of nonnegative numbers is nonnegative, hence $f'(c_+) \geq 0$; see Figure 7.2.2. Similarly, $f'(c_-)$ is a limit in which each term is nonpositive, thus $f'(c_-) \leq 0$. For these to be equal, both must equal zero. This finishes the proof.

Checkpoint 102.

Let's get the logic straight. It is of the form {\tt minimum $\Rightarrow f' = 0$}. The converse is not necessarily true: {\tt $f' = 0 \Rightarrow$ minimum}. Nevertheless, everyone's favorite procedure for finding minima is to set f' equal to zero. Why does this work, or rather, when does this work?

Let $a < b$ be real numbers. First of all, does f even have a minimum on $[a, b]$? In fact there are counterexamples in Figure 7.2.3.

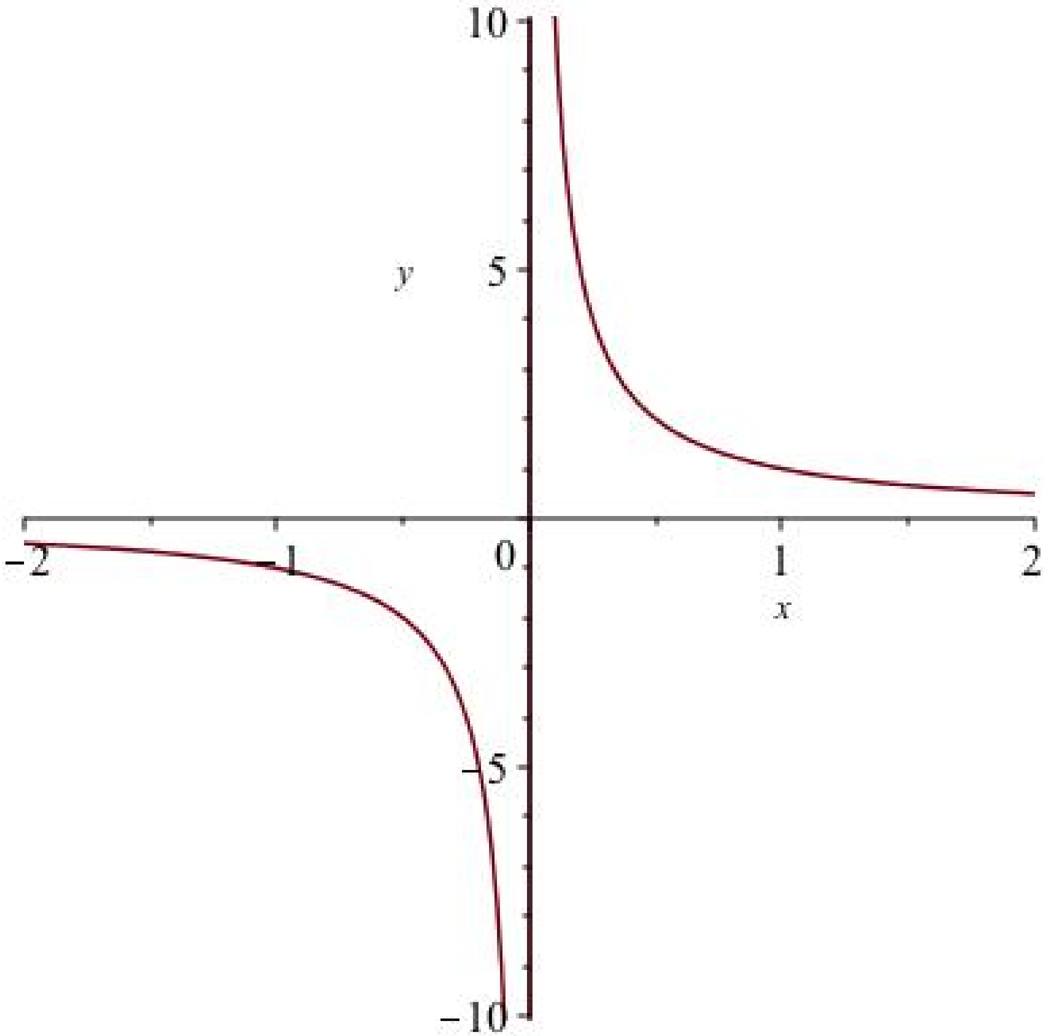


Figure 7.2.3. The function $1/x$ has no minimum on the interval $[-2, 2]$

From Theorem 7.1.4, if f is defined and continuous on a closed interval $[a, b]$, then indeed f has to have a minimum somewhere on $[a, b]$. We can find it by using Theorem 7.2.1 to rule out where it's not: if $a < c < b$ and $f'(c) \neq 0$, then definitely the minimum does not occur at c. Where can it be then? What's left is the point a, the point b, every point where f' is zero, and every point where f' does not exist. An identical argument shows the same is true for the maximum. Summing up:

> **Theorem 7.2.4.** *Suppose f is continuous on $[a, b]$ and differentiable everywhere on (a, b) except for a finite number of points $c_1, \ldots, c_k$. Then the minimum value of f on $[a, b]$ occurs at one or more of the points $\{a, b, c_1, \ldots, c_k,$ anywhere $f' = 0\}$, and nowhere else. The maximum also occurs at one or more of these points and nowhere else.*

Checkpoint 103.

Remark 7.2.5. Being differentiable except for a number of points (call them $c_0, \ldots, c_k$) is sometimes called being ***piecewise differentiable***, because the function is differentiable in pieces, the pieces being the intervals $(c_0, c_1), (c_2, c_2), \ldots, (c_{k-1}, c_k)$.

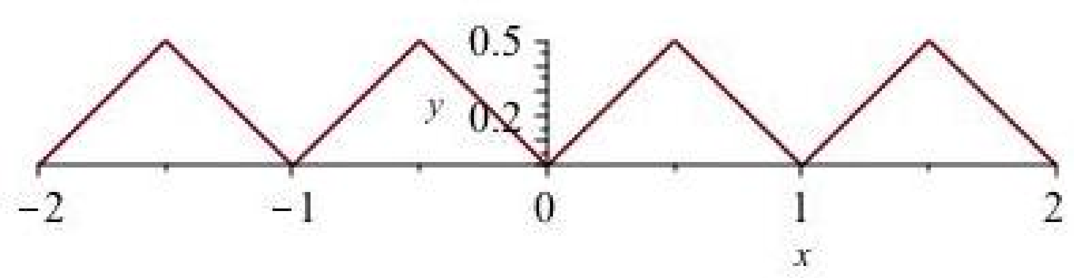


Figure 7.2.6.

Checkpoint 104.

You can write Theorem 7.2.4 as a procedure if you want. Even if you're looking only for the minimum or only for the maximum, the procedure is the same so it will find both.

1. Make sure f is continuous on $[a, b]$; if not, abort procedure.
2. Write down all $x \in (a, b)$ where $f'(x) = 0$.
3. Add to the list all $x \in [a, b]$ where $f'(x)$ DNE.
4. Add to the list the points a and b.
5. For every point x on the list, compute $f(x)$; the greatest value on this second list (the output list) will be the maximum; the least will be the minimum.
6. Using the two lists side by side, we can determine *where* the extrema occurred.

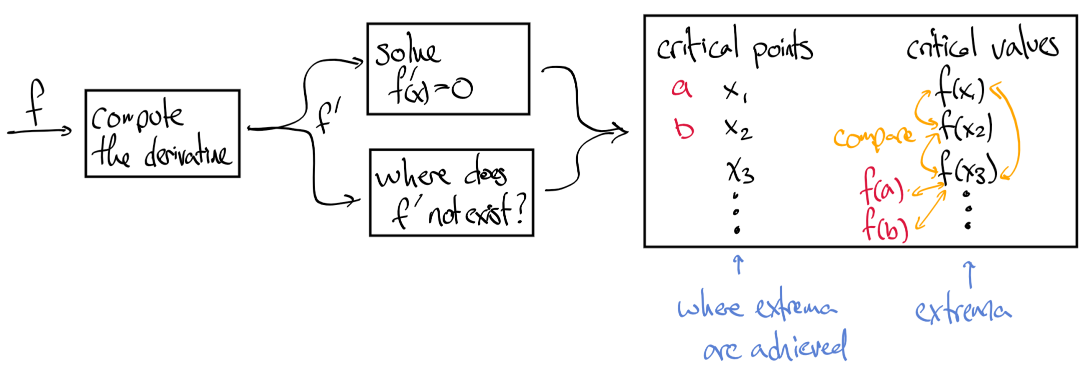


Figure 7.2.7. Theorem 7.2.4 as a procedure.

Example 7.2.8. Find the maximum of $f(x) := 5x - x^2$ on the interval $[1, 3]$; see the figure at the right. Computing $f'(x) = 5 - 2x$ and setting it equal to zero we see that $f'(x) = 0$ precisely when $x = 2\frac{1}{2}$. There are no points where f is undefined, so our list consists of just the one point plus the two endpoints: $\{1, 2\frac{1}{2}, 3\}$. Checking the values of f there produces $4, 6\frac{1}{4}, 6$. The maximum is the greatest of these, occurring at $x = 2\frac{1}{2}$.

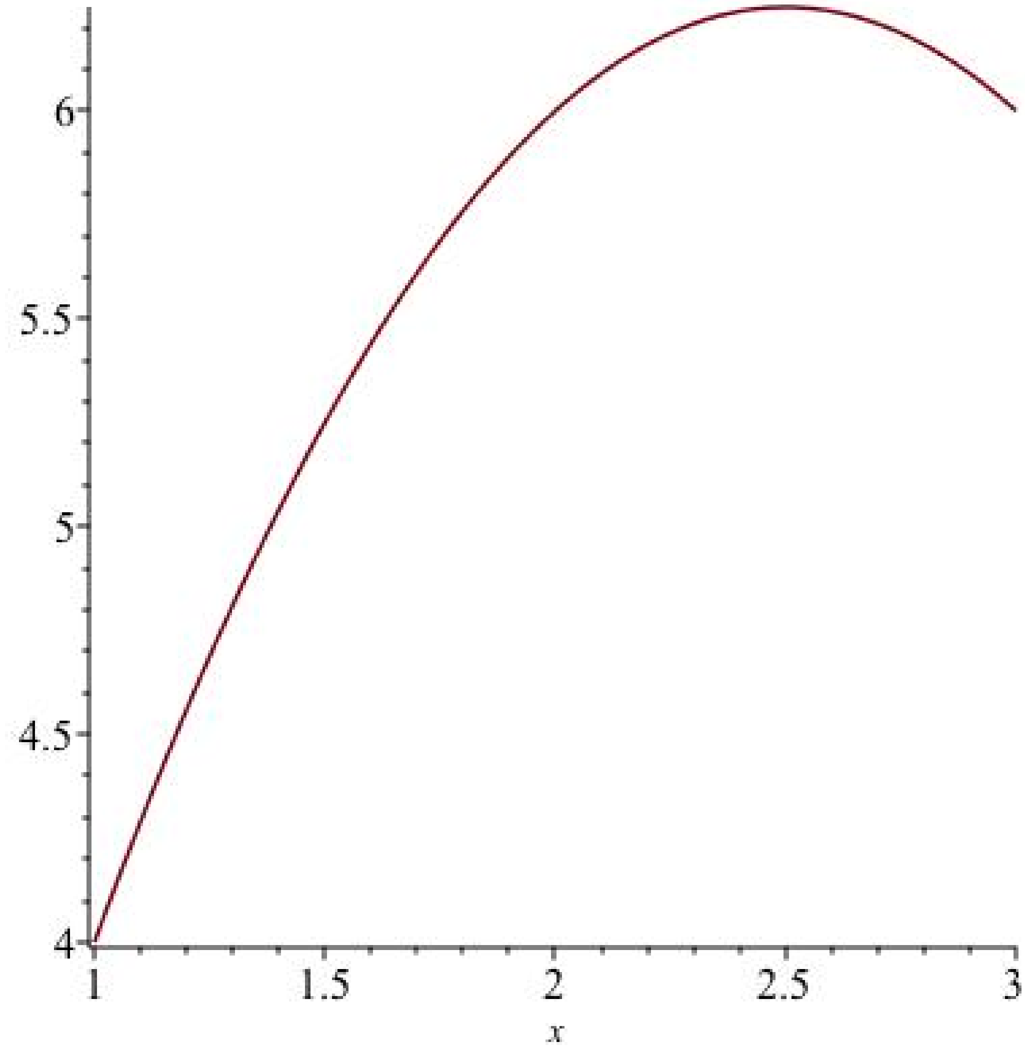
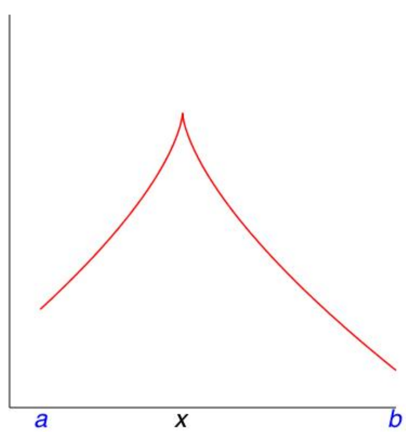


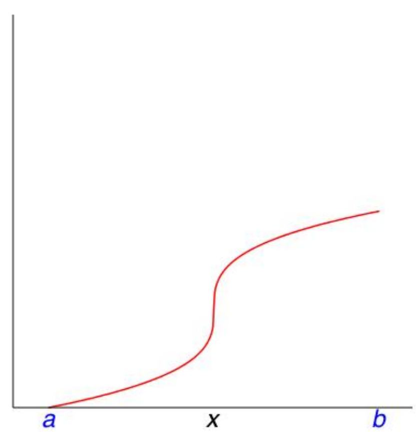
Figure 7.2.9. optimizing $f(x) = 5x - x^2$ on the interval $[1, 3]$

Checkpoint 105.

1.



2.



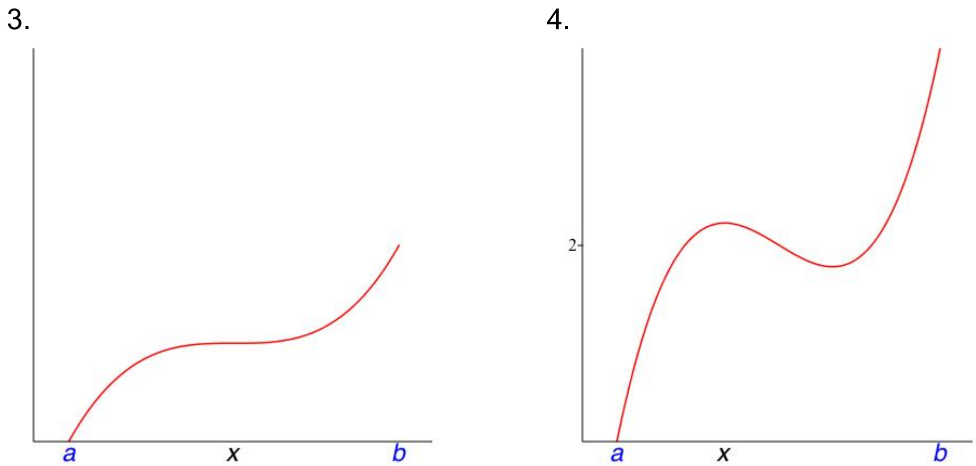


Figure 7.2.10. four functions for Checkpoint 106

Checkpoint 106.

Example 7.2.11. If the interval is not closed, the function might have no minimum. Let $f(x) = x$ and consider the half-closed interval $(0, 1]$. In this case we have a continuous function but not a closed interval.

This example represents a scenario where you make a donation in bitcoin to enter a virtual tourist attraction and you want to spend as little as possible. You have 1 bitcoin, so that's the maximum you can donate; donations can be any positive real number but zero is not allowed.

The minimum of x on $(0, 1]$ does not exist: there is no smallest positive real number.

The interpretation is clear: no matter how little you donate, you could have donated less. Mathematically, this clarifies the need for a closed interval in Theorem 7.1.4.

Checkpoint 107.

Second derivatives and extrema. Recall that wherever f has a second derivative, if $f'' \neq 0$ then the sign of f'' determines the concavity of f. If $f''(x) > 0$ then f is concave upward and if $f''(x) < 0$ then f is concave downward. At a point where $f' = 0$, if we know the concavity, we know whether f has a local maximum or local minimum.

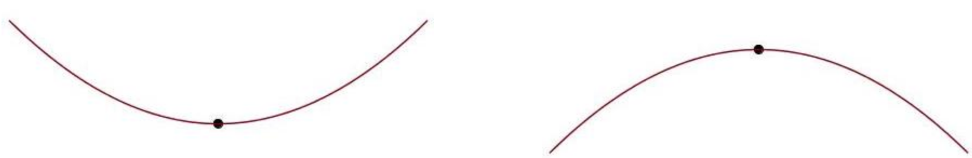


Figure 7.2.12. a critical point where $f'' < 0$ (left) and where $f'' > 0$ (right)

Example 7.2.13. What are the extrema of the function $f(x) := x^2 + 1/x$ on the interval $(0, 2)$? The only critical point is where $f'(x) = 2x - 1/x^2 = 0$, hence $x = \sqrt[3]{1/2}$. Here, $f''(x) = 2 + 2/x^3 > 0$ therefore this is a local minimum. There are not any local maxima. This means f has no global maximum on $(0, 2)$. It may have a global minimum, and indeed, Figure 7.2.14 shows that $x = \sqrt[3]{2}$ is a local minimum. In your homework you will get some more tools for arguing whether a local extremum on a non-closed interval is a global extremum.

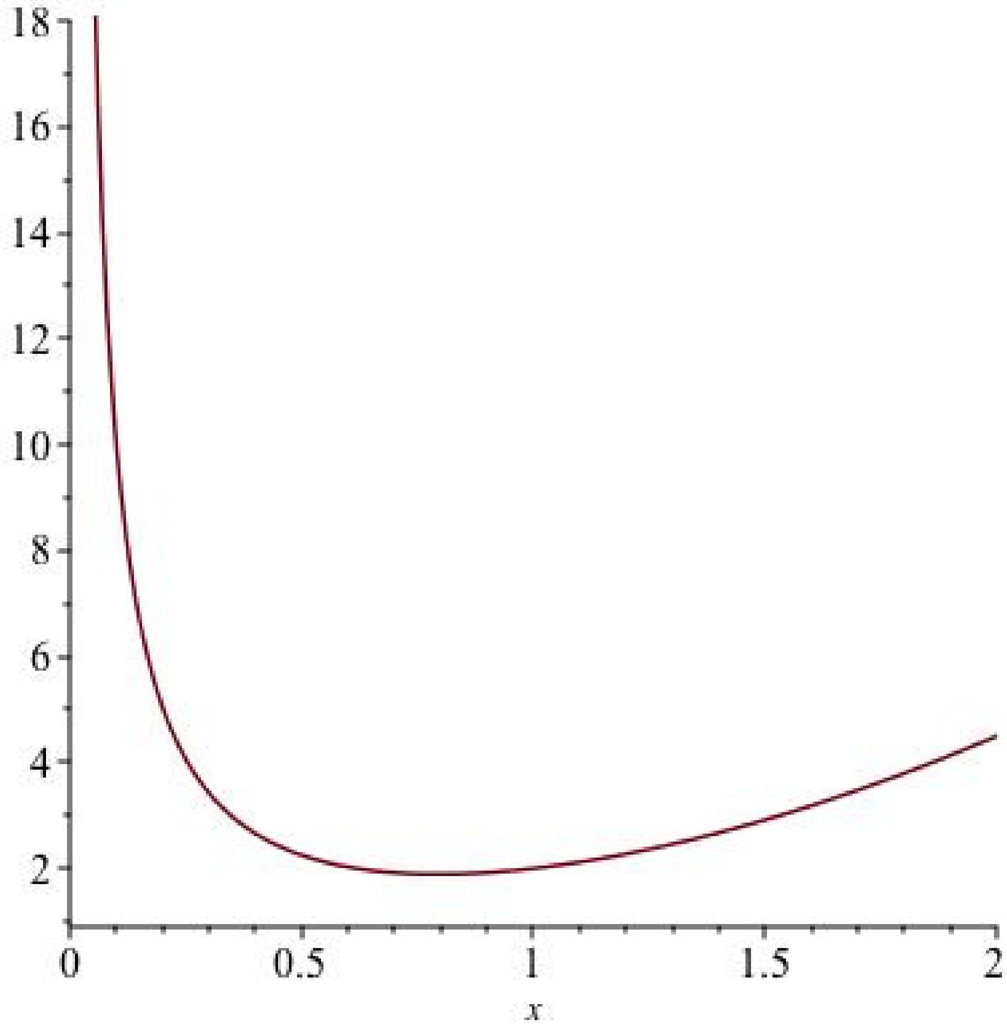


Figure 7.2.14. the function $x^2 + 1/x$ on the interval $(0, 2)$

Remark 7.2.15. If the second derivative vanishes along with the first, you won't know any more than you did already.

7.3 Some example applications

Finding extrema is part of a subject called ***optimization***. The idea is that you control a parameter x and are would like to maximize some ***objective function*** $f(x)$, which is perhaps how large you can build something, or perhaps revenue minus cost, or the efficiency of extraction of a natural resource.

Example 7.3.1. The logistic equation models growth rate per unit time, call it R, of a population as $R(x) = Cx(A - x)$. Here C is a constant of proportionality, x is the present population, and A is a theoretical limit on the population size supported by the habitat. At what size is the population growing the fastest?

We need to find the maximum of $R(x) := Cx(A - x)$ on $[0, A]$. The reason for restricting to this interval is that we are told the population size is constrained to be at most A, and of course it has to be nonnegative. Computing $R'(x) = C(A - 2x)$, we find $R' = 0$ for a single value, $x = A/2$. Checking the endpoints, we find R is zero at both. Therefore the maximum value occurs at $x = A/2$.

Checkpoint 108. n.edu/~ancoop/1070/sec-max-min.html#project-108)

Example 7.3.2. Suppose the cost of supplying a station is proportional to the distance from the station to the nearest port, and the cost of the land for the station is inversely proportional to the distance to the nearest port. Adding together these costs, what is the least expensive distance at which to put the station?

Letting x be the distance to the nearest port and $f(x)$ be the cost, we are told that $f(x) = ax + b/x$ where a and b are unspecified constants. The value of $f(x)$ is defined for every positive x and f is continuous on $(0, \infty)$. We seek the global minimum of f on $(0, \infty)$. We are not guaranteed there is a minimum. When we solve for $f'(x) = 0$ we find

$$0 = f'(x) = a - \frac{b}{x^2} \qquad \text{hence} \qquad x = \sqrt{\frac{b}{a}}\,.$$

At this value, $f(x) = a\sqrt{\frac{b}{a}} + b/\sqrt{\frac{b}{a}} = 2\sqrt{ab}$. Checking what happens near~0 and ∞, we find $\lim_{x\to 0} f(x) = \infty$ and $\lim_{x\to\infty} f(x) = \infty$. Therefore, there is a minimum value, which we have determined to be $\sqrt{ab}$ occuring at $x = \sqrt{\frac{a}{b}}$.

Example 7.3.3. The functions $x^\gamma e^{-x}$, for $x \geq 0$, arise in probability modeling. They are called Gamma densities. We will return to these in Unit 13 (ss-prob.html). For now, we would like to understand the shape of these functions. An example with $m = 5$ is shown in Figure 7.3.4.

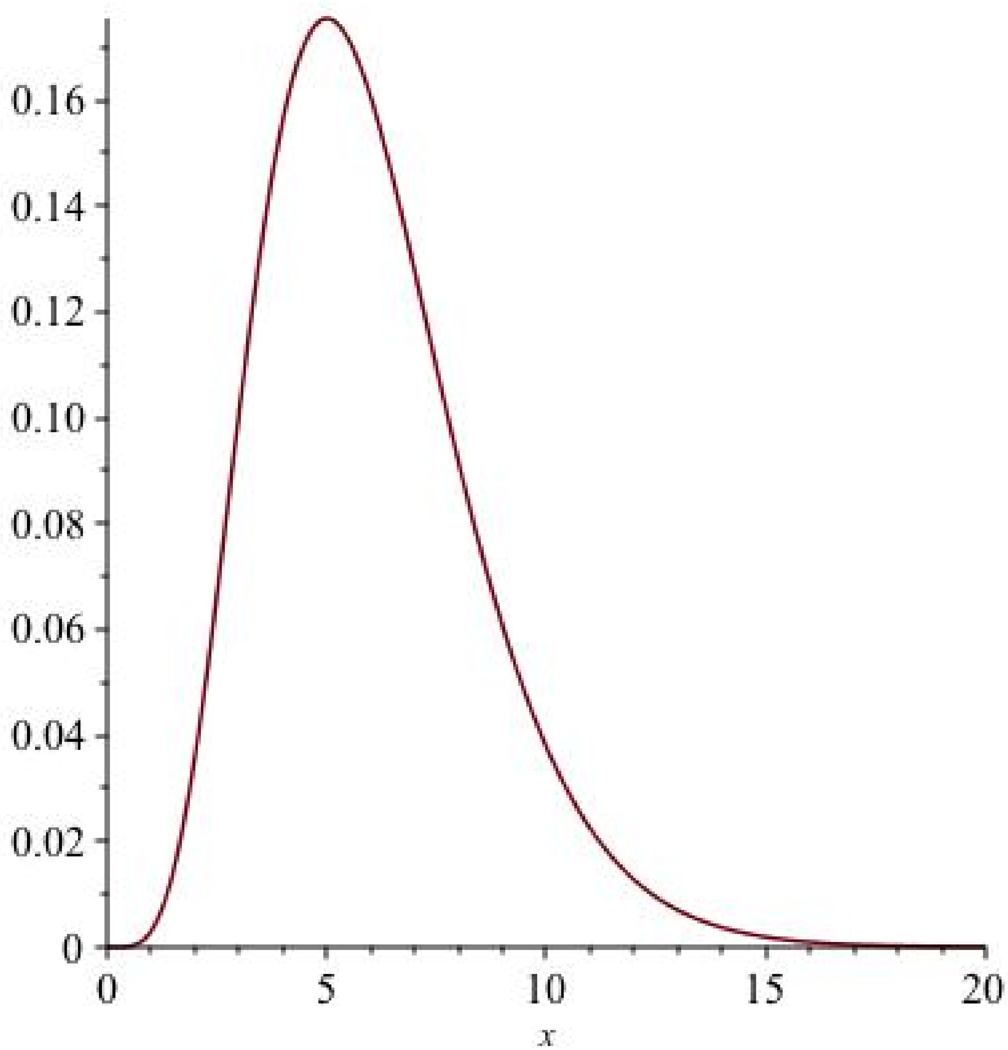


Figure 7.3.4. Gamma-5 density

The place where one is mostly to find the random variable is where the maximum of the density occurs. Where does the maximum of $f(x) := x^5 e^{-x}$ occur? We know that the value is zero at $x = 0$ and positive everywhere else. We also know $\lim_{x\to\infty} f(x) = 0$. This means there must be a maximum at some positive finite x. The function f is differentiable for all positive x, therefore the maximum can only occur where $f' = 0$. Solving

$$0 = f'(x) = 5x^4 e^{-x} - x^5 e^{-x} .$$

Factoring out $x^4 e^{-x}$, we see that $x = 5$. Therefore, the maximum occurs at $x = 5$.

Checkpoint 109.

Example 7.3.5. Let h be the height of a member of a carnivore species. In this simple model, the food gathering capability of an individual is given by kh^2 while its daily food needs are given by ch^3.

Why? We can only make educated guesses about the reason the equations in the model have this form. If an animal's speed is proportional to its height then the model stipulates territory is proportional to the square of this. Perhaps territory is the area that can be reached in a given amount of time such as an hour or a day. As to why food needs would be proportional to volume, one might imagine that sustaining and nourishing tissue requires nutrients proportional to volume.

What are the units of c and k? Units of c are food per length^3 and units of k are food per length^2. For example, if food is measured in kilograms and length in meters, then food per length^3 would be kg/m^3; however one might measure food in other ways such as calories, or numbers of a particular animal of prey, etc.

To maximize food gathering ability minus food needs, how tall should members of this species be? The objective function we want to maximize is $kh^2 - ch^3$. Having been told no limitations on size, we assume h can be any positive real number, though we may have to retract that if the optimum turns out to have unrealistic scale. Differentiating $f(h) := kh^2 - ch^3$ with respect to h yields $2kh - 3ch^2$ and setting equal to zero gives the two solutions 0 and $x_* := (2k)/(3c)$. This indeed has units of length. Clearly $f(0) = 0$. The value of the objective function at x_* is $4k^3/(27c^2)$, which is positive. Therefore the maximum of f on $[0, \infty)$ is either $4k^3/(27c^2)$ achieved at $h = (2k)/(3c)$ or there is no maximum because the function can get arbitrarily large as $h \to \infty$. At infinity, $f(h) \sim -ch^3$ because $kh^2 \ll ch^3$ as $h \to \infty$. Therefore, h has a maximum at a positive location, whose value is $4k^3/(27c^2)$

Checkpoint 110.

∑ () MATH 1070: Mathematics of Change, part I (article-1.html)

Calculus Group

7 Applying the optimization procedure

()Theorem 7.2.4 gives us a procedure for finding extrema of functions on closed intervals. Now we're going to apply that procedure to help us find the best, cheapest, most effective, etc.

As with any application problem, the hardest part is setting up the mathematical model that captures the situation we want to apply ()Theorem 7.2.4 to. Since our optimization tool applies to functions, the task boils down to writing a single function. Luckily for us, we discussed a lot of what we'll need here back in Unit 2 (sec-modeling.html).

8.1 Optimization in geometry

We'll start with a few geometric problems, where the objective function is a little more obvious.

Example 8.1.1. We're going to build a window in the shape of a rectangle topped by an equilateral triangle. We want to make a window which lets in the most light -- that is, with the greatest possible area. In order to build the window, we have to use wood trim. We have 16 feet of wood trim to build the window with.

Such a window has two dimensions: the width w and the height h of the rectangle. The rectangular portion has area wh and the triangular portion has area $\frac{1}{2}w^2$. So the total area is

$$A(w,h) = wh + \frac{1}{2}w^2\,.$$

We also need to record the fact that our supplies are limited. A little geometry shows that to build the window requires two pieces of trim with length h and four of length w.

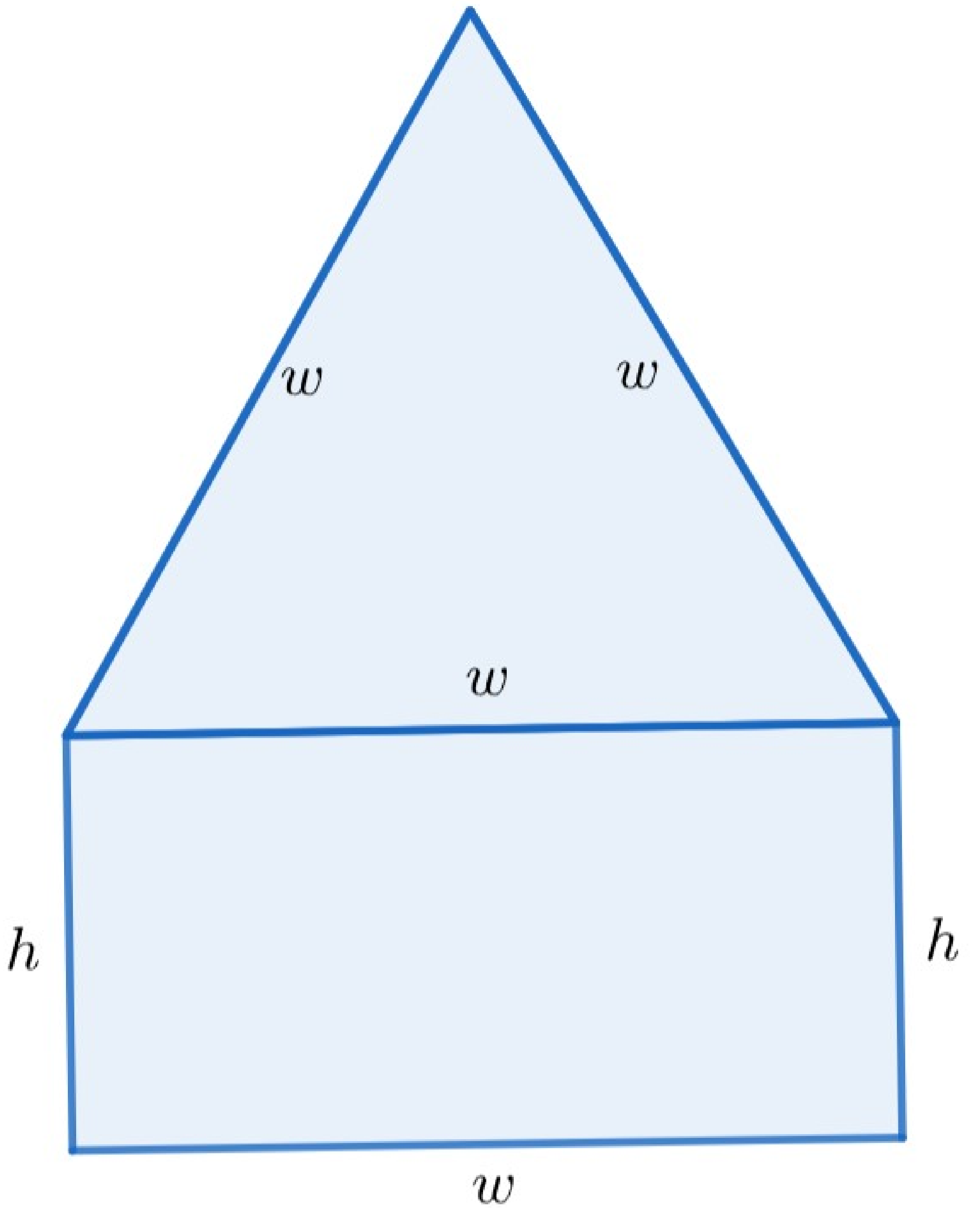


Figure 8.1.2. A window made from a rectangle and a triangle.

Technically we don't have to use *all* the trim, but if we had some left over, we could have used it to build a bigger window. So let's assume we use all 16 feet; that is, we assume

$$16 = 2h + 4w\ .$$

We can solve this equation for either w or h. Let's solve for h:

$$h = 8 - 2w$$

and substitute that into the formula for area:

$$A(w, h(w)) = w(8 - 2w) + \frac{1}{2}w^2$$

Now we've got a function which we can optimize. We want to have a sensible result, so we know that w can't be less than 0, and can be at most 4. So we want to optimize on the interval $[0, 4]$.

Differentiating, we get $\frac{dA}{dw} = 8 - 3w$. So there is a single critical point at $w = \frac{8}{3}$. We have

$$A(0, h(0)) = 0$$
$$A(4, h(4)) = A(4, 0) = \frac{1}{2}4^2 = 8$$
$$A\left(\frac{8}{3}, h(\frac{8}{3})\right) = A\left(\frac{8}{3}, \frac{8}{3}\right) = \left(\frac{8}{3}\right)^2 + \frac{1}{2}\left(\frac{8}{3}\right)^2 = \frac{32}{3}$$

Since $\frac{32}{3}$ is greater than either 0 or 8, we see that the maximal area occurs when we choose width and height both equal to $\frac{8}{3}$.

This example shows a few things. First, notice that many optimization problems come equipped with a ***constraint***. Here that was the fact that we only had so much trim. If you think about it, constrained optimization is the kind you usually deal with -- our world is full of scarcity.

Second, we can interpret each of the values we compared in the context of the problem. At $w = 0$, the window has width zero. At $w = 4$, $h = 0$ so we only have a triangular section of window. $w = \frac{8}{3}$ is somewhere in between.

Third, the optimal dimensions happened to be equal to one another. This is typical -- optimizers are often symmetric (in this case, the symmetry is that w and h are the same.

Here are some other geometric optimization problems.

Checkpoint 111.

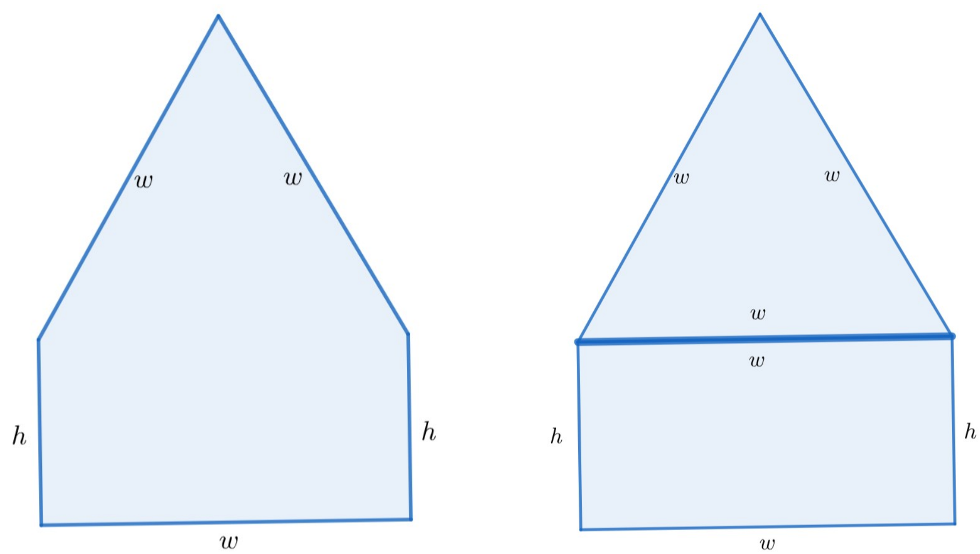


Figure 8.1.3. Windows made from a rectangle and a triangle.

Checkpoint 112.

8.2 Optimization in economics and business

Let's think about production. A standard model, called the ***Cobb-Douglas production function***, says that the productivity of a firm is proportional to both a power of the labor inputs L and a power of the capital inputs K; and that the powers add to 1. That is,

$$P = kK^{\alpha}L^{\beta}$$

where k is the constant of proportionality and $\alpha + \beta = 1$. We may as well write $\beta = 1 - \alpha$, so that our formula reads

$$P = kK^{\alpha}L^{1-\alpha}\,.$$

Clearly, if we could increase K and L without constraint, we could increase the firm's output arbitrarily. But we have to operate subject to a budget. Spending more on labor means we have less to spend on capital, and vice versa. We model this by

$$B = K + L$$

that is, the total budget is the sum of the captial costs and the labor costs.

A natural question to ask is: given a budget, how do we maximize output?

Just as with the window example, we manipulate our constraint to express one variable in terms of the other: $L = B - K$. Then we substitute this into the objective function:

$$P(K, L(K)) = kK^{\alpha}(B - K)^{1-\alpha}\,.$$

That's a function we can optimize, on the interval $[0, B]$.

Checkpoint 113.

Checkpoint 114.

maximizing profit. Consider the problem of a firm producing and selling a single good (say, pairs of sneakers). The goal of the firm is to make the most money possible.

Before we get into a precise model, let's set the ground rules.

production level	We'll call the number of pairs of sneakers we produce t.
production costs	We'll write $C(t)$ for the total cost of producing t pairs of sneakers. $C(t)$ is an increasing function.
markets clear	We'll assume that we sell every pair of sneakers we make.
revenue	We'll write $R(t)$ for the total revenue that selling t pairs of sneakers brings in. $R(t)$ is also an increasing function.
profit	The profit we make is $P(t) = R(t) - C(t)$. Our goal is to maximize $P(t)$.

In finance and economics, the adjective ***marginal*** is used to denote a derivative. So we say ***marginal revenue*** to mean $R'(t)$ and ***marginal costs*** to mean $C'(t)$.

Checkpoint 115.

This use of the word $marginal$ comes from the fact that, using the tangent line approximation to $C(t)$,

$$C(t+1) \sim C(t) + C'(t)\,((t+1) - t)$$
$$C(t+1) \sim C(t) + C'(t)$$

In other words, $C'(t)$ is approximately the additional cost added by the additional pair of sneakers that took us from production level t to production level $t+1$.

In fact, thinking about derivatives this way can be very useful to understanding the situation of maximizing profit.

Checkpoint 116. n.edu/~ancoop/1070/section-8.html#project-116)

Formally, we're trying to optimize

$$P(t) = R(t) - C(t)$$

so our first step ought to be differentiating:

$$P'(t) = R'(t) - C'(t)\ .$$

We want to find critical points of P; that is, we need to solve

$$\begin{aligned} P'(t) &= 0 \\ R'(t) - C'(t) &= 0 \\ R'(t) &= C'(t) \end{aligned}$$

That is, ***we're looking for the production level where marginal cost and marginal revenue are equal***.

Checkpoint 117. n.edu/~ancoop/1070/section-8.html#project-117)

modeling revenue. Because profit is the difference of revenue and costs, understand how to solve $P'(t) = 0$ amounts to understanding the revenue function $R(t)$ and the cost function $C(t)$.

Revenue seems straightforward. If we produce and sell t pairs of sneakers for a price of p dollars per sneaker, then

$$R(t) = p \cdot t$$

But! the more sneakers we produce and sell, the less unique an individual wearing those sneakers is. The more sneakers we produce and sell, the fewer people go unshod. That tends to drive the price down. So the price p isn't a constant; it's a function $p(t)$ of the number of pairs of sneakers we've sold.

Let's say that we've done some market research, and we've found that the market price of a pair of sneakers seems to obey

$$p(t) = 300 - .05t.$$

Checkpoint 118. n.edu/~ancoop/1070/section-8.html#project-118)

Notice that our model predicts that for high enough levels of production, $p(t)$ is negative. That is, if we completely saturate the market, we'd have to start paying people to take our sneakers (instead of them paying us). Is this prediction reasonable?

Checkpoint 119. n.edu/~ancoop/1070/section-8.html#project-119)

modeling costs. What about costs? A standard model of costs is linear:

$$C(t) = C_0 + mt$$

where C_0 is called the ***fixed cost*** and represents the costs we have to spend no matter what: capital outlay for the factory, bribes for local politicans, etc.; and m is the marginal cost (materials and labor to produce a single pair of sneakers).

Checkpoint 120.

Let's say we build a factory for $10,000 and our marginal cost is $2 per pair of sneakers. Then we have

$$C(t) = 10000 + 2t$$

and

$$P(t) = t \cdot (300 - .05t) - (10000 + 2t) \ .$$

How do we acheive the maximum profit? When

$$0 = P'(t) = 302 - .1t$$

which is $t = 3020$. The profit we actually make at that production level is $P(3020) = 446,020$. Not too bad.

Checkpoint 121.

But costs are not always linear. We say that a cost function obeys ***economies of scale*** if the marginal cost gets smaller as we increase the production level. For example, a worker producing their first pair of sneakers might take a lot of time, but by the time they get to their 30^{th} pair of sneakers, the same worker can probably do so much more quickly (which means the labor cost for that pair of sneakers will be lower.

Checkpoint 122.

Just as we did with revenue, we can bootstrap our linear model for costs $C(t) = C_0 + mt$ into a more model by replacing m with $m(t)$. If we think of our workers in the sneaker factory as gaining skill over time, then we could write something like

$$m(t) = 1 + \left(\frac{1}{2}\right)^{t/10}$$

which is to say: the marginal cost starts at $2 per pair of sneakers, has a long-run limit of $1 per pair of sneakers, and every 10 pairs of sneakers produced moves the marginal cost halfway to the long-run limit.

How do we achieve the maximum profit in this case? We have

$$\begin{aligned} P(t) &= R(t) - C(t) \\ &= t \cdot (300 - .05t) - \left(10000 + \left(1 + \left(\frac{1}{2}\right)^{t/10}\right)t\right) \end{aligned}$$

which means that we need to solve

$$0 = P'(t) = 299 - .1t - \left(\frac{1}{2}\right)^{t/10} - \frac{1}{10} t \ln\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^{t/10} \ .$$

Unfortunately there's not an algebraically clean solution to this equation, but a graphing utility says that $t = 2990$ is very close.

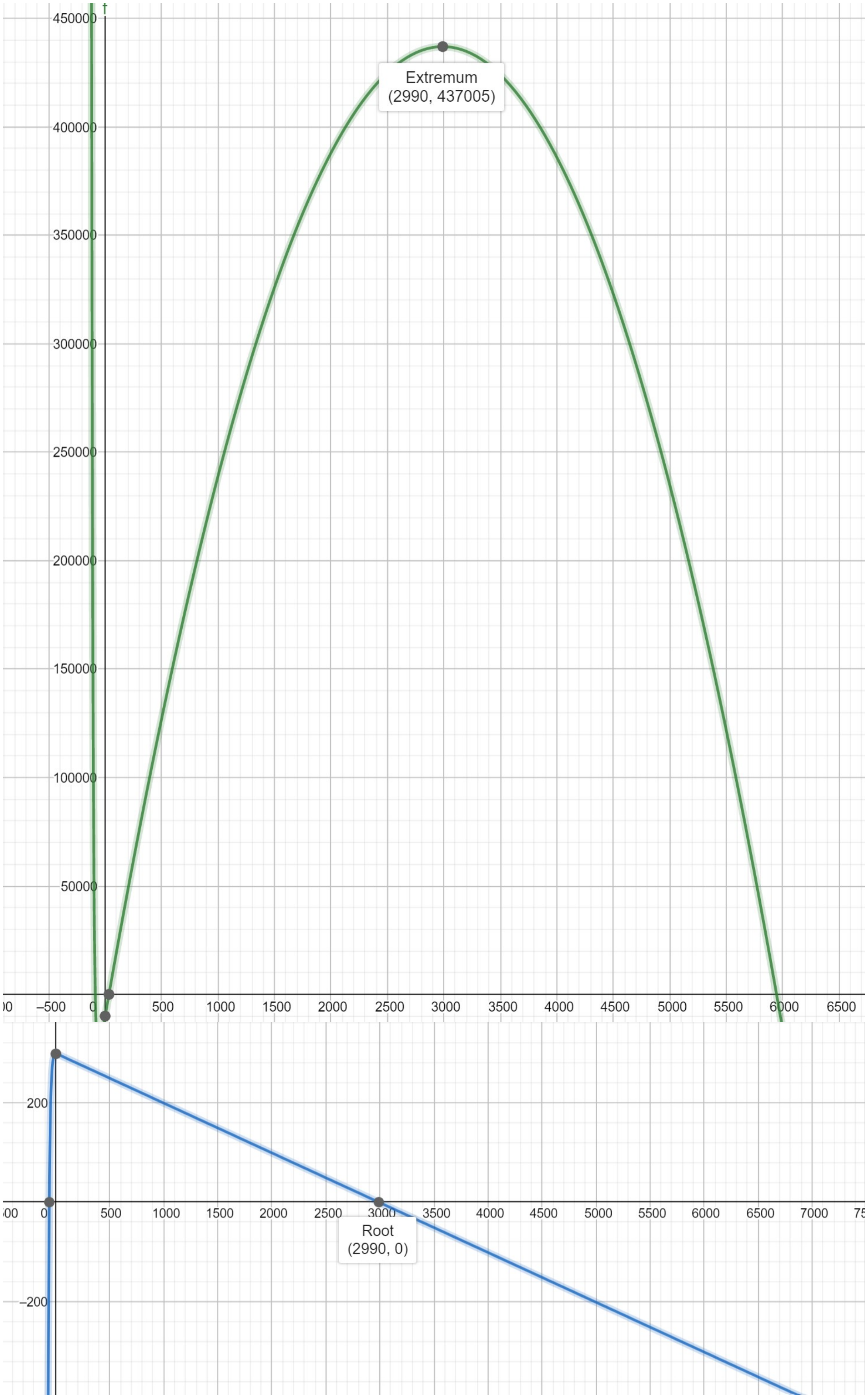


Figure 8.2.1. Graphs of $P(t)$ and $P'(t)$, showing that $t = 2990$ is approximately a critical point.

Σ () MATH 1070: Mathematics of Change, part I (article-1.html)

Calculus Group

8 Summation (…edu/~ ncoop/1070/section-9.html#section-9)

9.1 Sequences (…edu/~ancoop/1070/section-9.html#ss-sequences)

We already briefly discussed sequences. When working with sums and the "Sigma" notation for summations, you need to be able to write formulas for sequences you understand intuitively. For example, if you want to write the sequence $7, 9, 11, 13, \ldots$ in the notation $\{b_n : n \geq 1\}$, so that $b_1 = 7$,$b_2 = 9$ and so on, one choice would be to say,

> *Let* $\{b_n : n \geq 1\}$ *be the sequence defined by* $b_n := 5 + 2n$.

The subscript n is called the ***index*** (plural: ***indices***). Indexing can begin at any natural number. In this case, as is most common, we began at $n = 1$. Defining $\{b_n : n \geq 3\}$ by $b_n := 1 + 2n$ yields the same sequence: $7, 9, 11, 13, \ldots$.

Secondly, the informal notation $7, 9, 11, 13, \ldots$ is not mathematically precise because it assumes we all agree exactly what the pattern is. Producing a formula for the n^{th} term removes any ambiguity. A formula is often necessary if you want to sum the sequence or to use it to define other sequences. This section considers some common types of sequences and gives you some practice writing a formula for the general term.

Checkpoint 123. (…n.edu/~ancoop/1070/section-9.html#project-123)

Definition 9.1.1. A sequence is called ***arithmetic*** if the difference between successive terms is constant.

Our example sequence $7, 9, 11, 13, \ldots$ is an arithmetic sequence with common difference 2. It is particularly easy to write a formula for the general term of an arithmetic sequence if you start indexing at zero. The n^{th} term is the zeroth term plus n copies of the common difference. In notation, if the common difference is d and the sequence is $\{a_k : k \geq 0\}$,this means $a_k = a_0 + kd$. Setting $a_0 = 7$ and $d = 2$ gives $a_k = 7 + 2k$ for the sequence $7, 9, 11, 13, \ldots$.

Checkpoint 124. (…n.edu/~ancoop/1070/section-9.html#project-124)

Definition 9.1.2. A sequence is called ***geometric*** if the ratio between successive terms is constant. In other words, if the sequence is $\{u_j\}$, then the ratio u_{j+1}/u_j has some common value r for all j.

$10, 20, 40, 80, 160, \ldots$

Checkpoint 125. (…n.edu/~ancoop/1070/section-9.html#project-125)

Sequences with alternating signs appear often enough that it's a good idea to know a way to write their general term. The key to being able to write such sequences is to notice that $(-1)^n$ bounces back and forth between $+1$ and -1. The odd terms are negative, so starting with $n = 1$ (or 3 or 5, etc.) starts with -1 whereas starting with 0 (or 2, -2, etc.) starts with $+1$. You can incorporate this in a sum as a multiplicative factor and it will change the sign of every second term. Thus for example, to write the sequence $1, -2, 3, -4\ldots$ you can write $(-1)^{n+1} \cdot n$. Note that we used $(-1)^{n+1}$ rather than $(-1)^n$ so that the term corresponding to $n = 1$ was positive rather than negative.

When the sum has a pattern that takes a couple of steps to repeat, the greatest integer function can be useful. For example, $1, 1, 1, 2, 2, 2, 3, 3, 3, \ldots$ can be written as $a_n := \left\lfloor \frac{n+2}{3} \right\rfloor$ for $n \geq 1$. Actually, it comes out a little more simply if you index starting from zero: $a_n := \left\lfloor \frac{n}{3} \right\rfloor + 1$ for $n \geq 0$.

Definitions by cases work for sequences just the way they do for functions. Suppose you want to define a sequence with an opposite sign on every third term, such as $-1, -1, 1, -1 - 1, 1, \ldots$. You can do this by cases as follows.

$$a_n := \begin{cases} -1 & n \text{ is not a multiple of } 3 \\ 1 & n \text{ is a multiple of } 3 \end{cases}\ldots$$

Plenty of sequences don't fit any of these molds. Writing a formula for the general term is a matter of trying an expression, seeing if it works, then if not, tinkering to get it right.

9.2 Finite series

Let's talk for a minute about a notation you have likely seen before. It is called the "Sigma" notation because that's the name of the Greek letter Σ. The notation involves an indexing variable which runs between a lower limit and an upper limit. The lower and upper limits are required to be integers\footnote{This is unlike computer science, where a loop counter can increment by any number}. If the indexing variable is n,the lower limit is L,the upper limit is U and the general term is b_n,the summation looks like $\sum_{n=L}^{U} b_n$. What this means is to add together all the values of b_n starting with $n = L$ and ending with $n = U$.

Example 9.2.1. $\sum_{n=1}^{5} 2^{-n}$ represents the sum

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32}.$$

The summand, as you can see, is usually a function of the indexing variable; otherwise, the summand would not change from term to term. There may be other variables, for example $\sum_{k=3}^{6} kx$ evaluates to $3x + 4x + 5x + 6x$,which is equal to $18x$. Note that this other variable x persists when the sum is evaluated. It is a free variable. On the other hand, the index of summation, k in this case, is a bound variable. It runs over a set of values (in this case 3 to 6) and does not appear in the final value.

Checkpoint 126.

When a sequence is summed, the result is called a ***series*** (the plural is also *series*).

Example 9.2.2. The sum $\sum_{n=5}^{19} \frac{3}{n-2}$ represents a series with 15 terms because there are 15 integers in the range from 5 to 19. Informally, we might write this sum by writing the first few terms and the last term, with dots in between (traditionally the dots are centered for series, as opposed to at the bottom of the line for sequences). Thus we would write $\frac{3}{3} + \frac{3}{4} + \cdots + \frac{3}{17}$, assuming this conveyed enough information for the reader to understand the precise sum.

Of course there is no reason why the index should go from 5 to 19. There have to be fifteen terms, but why not write the sum with the index going from 1 to 15? Then it would look like

$$\sum_{n=1}^{15} \frac{3}{n+2} \, .$$

Another natural choice is to let the index run from 0 to 14:

$$\sum_{n=0}^{14} \frac{3}{n+3} \, .$$

All three of these formulas represent the exact same sum.

Checkpoint 127.

9.3 Some series you can explicitly sum

The series in Example 9.2.2 sums to a rational number. According to Excel it is equal to $23763863/4084080$. There isn't any really nice formula for this sum in terms of the values 5 and 19 and the function $n \mapsto 3/(n-2)$. In fact most series don't have nice summation formulas. Arithmetic and geometric series are exceptions. Because they are common and the formulas are simple and useful, we include them in this course.

Arithmetic series. Here's an example of how to sum an arithmetic series, which generalizes easily to summing any arithmetic series. This particular example is a well known piece of mathematical folklore (google "Gauss child sum").

Example 9.3.1.

Problem Sum the numbers from 1 to 100.

Solution Pair the numbers starting from both ends: 1 pairs with 100, 2 pairs with 99, and so forth, ending at 50 paired with 51. There are 50 pairs each summing to 101, so the sum is $50 \times 101 = 5050$.

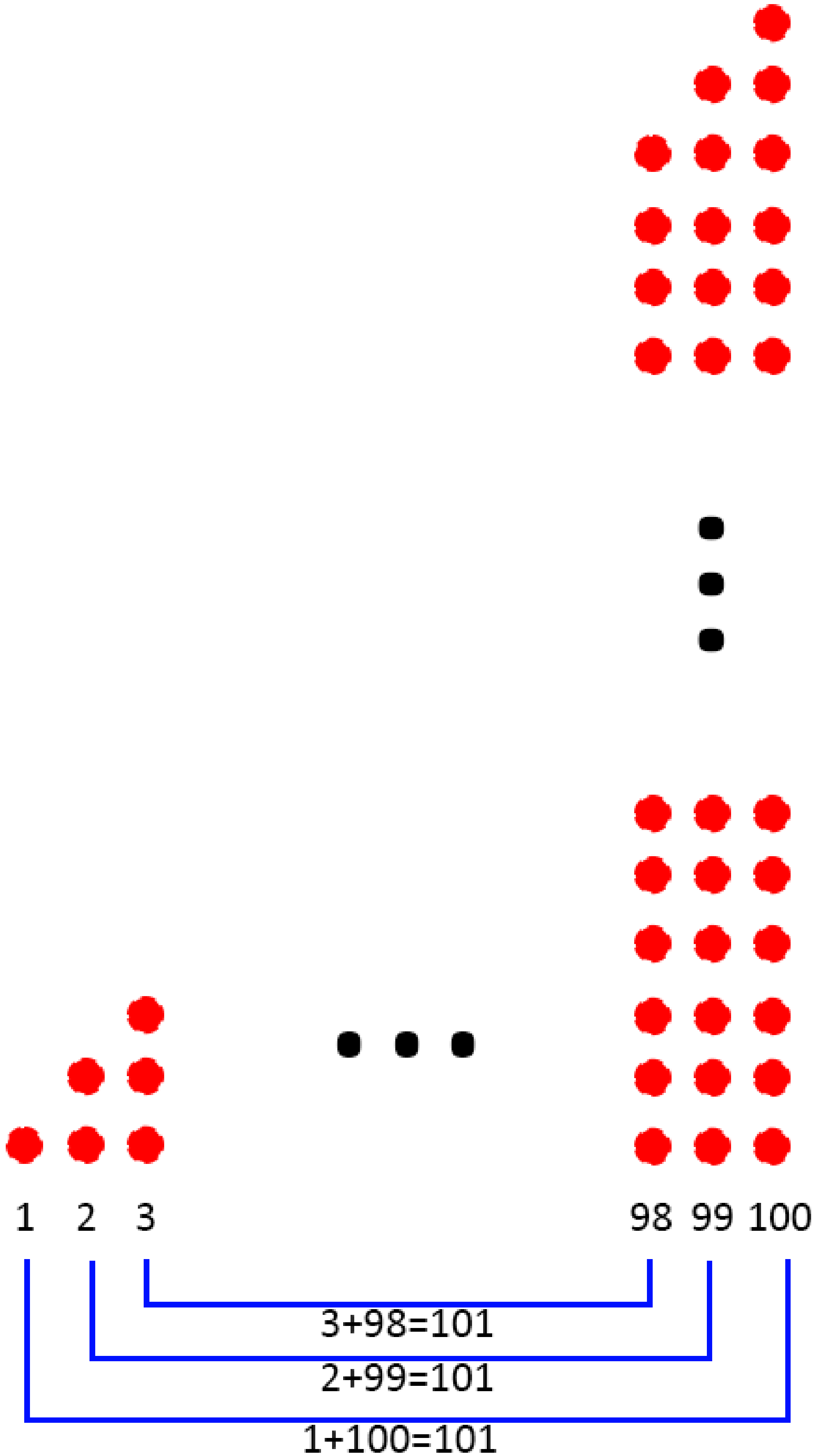


Figure 9.3.2. A diagram of the solution described in Example 9.3.1 .

Another way to get the same formula is only slightly different.

Example 9.3.3. Evaluate $\sum_{n=13}^{29} n$. There are 17 terms and the average is 21, which can be computed by averaging the first and last terms: $(13+29)/2 = 21$. Therefore, the sum is equal to $17 \times 21 = 357$.

Checkpoint 128.

Geometric series. The standard trick for summing geometric series is to notice that the sum and r times the sum are very similar. It is easiest to explain with an example.

Example 9.3.4. Evaluate $\sum_{n=1}^{10} 7 \cdot 4^{n-1}$. To do this we let S denote the value of the sum. We then evaluate $S - 4S$ (because $r = 4$). I have written this out so you can see the cancellation better.

$$\begin{aligned} S - 4S = & \quad 7 + 28 + 112 + \cdots + 7 \cdot 4^9 \\ & - \left(28 + 112 + \cdots + 7 \cdot 4^9 + 7 \cdot 4^{10}\right). \end{aligned}$$

Thus,

$$(1-4)S = \quad 7 - 7 \cdot 4^{10}.$$

From this we easily get $S = (7 - 7 \cdot 4^{10})/(1-4) = 7(4^{10}-1)/3 = 2446675$.

Checkpoint 129.

Now let's evaluate the general form $\sum_{n=1}^{M} A \cdot r^{n-1}$.

Letting S denote the sum we have $S - rS = A - Ar^M$ and therefore

$$S = A \, \frac{1 - r^M}{1 - r}.$$

When A and r are positive, all the terms are positive, hence the sum is positive as well. When $r < 1$ this is very evident from the formula. When $r > 1$ it is true as well, but easier to see multiplying top and bottom by -1 so as to get $A(r^M - 1)/(r-1)$. When $r = 1$ this quotient is undefined, however the sum is very easy: M copies of A sum to $A \cdot M$.

9.4 Infinite series

No discussion of series would be satisfied if it didn't answer the question, "Is $0.9999\ldots$ (repeating) actually equal to 1?" As you can probably guess, it is a matter of definition. However, there is a standard definition, and therefore we can in fact supply an answer (see below).

Definition 9.4.1.

$$\sum_{n=L}^{\infty} b_n := \lim_{M \to \infty} \sum_{n=L}^{M} b_n .$$

This definition might require a bit of unpacking. First of all, the colon-equal is right: the symbol $\sum_{n=L}^{\infty} b_n$ on the left is not already defined, and we are *defining* it to be the value on the right. So what we are saying is that the sum of an infinite series is the limit of a certain sequence, called the ***sequence of partial sums***.

Example 9.4.2. How does this definition apply to the so-called ***harmonic series***, $\sum_{n=1}^{\infty} 1/n$? It says that this infinite sum is equal to the limit of the sequence $\{H_M\}$,where H_M is the ***harmonic number*** $\sum_{n=1}^{M} 1/n$. The harmonic numbers H_M are said to be the ***partial sums*** of the harmonic series. Interpreting the infinite sum in this way doesn't tell us whether the limit is defined, or if so, what it is, it just tells us that if we can evaluate the limit $\lim_{M\to\infty} H_M$, this is *by definition* the sum of the harmonic series. If the limit is undefined, then the sum of the harmonic series is undefined.

Checkpoint 130. n.edu/~ancoop/1070/section-9.html#project-130)

Example 9.4.3. enn.edu/~ancoop/1070/section-9.html#example-48)

Problem evaluate $1 + 1/2 + 1/4 + 1/8 \cdots$.

Solution this is the infinite sequence $\sum_{n=0}^{\infty} (1/2)^n$. The value is the limit of the partial sums $S_M := \sum_{n=0}^{M} (1/2)^n$. Evaluating these finite sums gives

$$S_M = \frac{1 - (1/2)^{M+1}}{1 - 1/2} = 2 - \frac{1}{2^M}\,.$$

The infinite sum is then $\lim_{M\to\infty} 2 - (1/2)^M$ which is clearly equal to 2.

Checkpoint 131. n.edu/~ancoop/1070/section-9.html#project-131)

9.5 Financial applications

When writing equations for balance sheets of loans, annuities, endowments and other investment schemes, it helps to be able to convert quickly between continuous growth and discrete-time growth such as monthly, quarterly or yearly. Recall from Example 6.1.6 that an interest rate (written as a real number, so you need to divide by 100 if it is quoted as a percent) of r corresponds to a growth factor of e^r over the course of a year. The increase in an initial amount M over a year is $e^r M - M = (e^r - 1)M$. This increase is proportional to M of course, and the constant of proportionality $e^r - 1$ is called the ***annual yield***. Denoting this by y,we can state either in terms of the other:

$$y = e^r - 1 \qquad ; \qquad r = \ln(1 + y)\,. \tag{9.5.1}$$

For those committed to stating things in terms of percentages, let $Y = 100y$ denote the annual percentage yield (APY) and $R = 100r$ denote the percentage interest rate, this becomes

$$Y = 100(e^{R/100} - 1) \qquad ; \qquad R = 100 \ln(1 + Y/100) \, .$$

Checkpoint 132.

Consider a mortgage loan (loan for a house) or car loan. Typically payments on these are made monthly, which we will take to be every $1/12$ of a year. In this case the factor by which your debt grows each month is $e^{r/12}$,where r is the (annual) interest rate. That's only if you don't pay off the loan. Actually, these loans are typically configured so you pay a fixed amount every month until the loan is paid off in an integer number of months (usually, in fact, an integer number of years). To agree on some notation, let r be the annual interest rate, P be the ***principal***, that is the initial debt, and let M be the monthly payment.

In order to deal successfully with used car sales people, it's helpful to understand how these determine your balance over the successive months. The key relation is to understand what happens from one month to the next. We will discuss this, then leave the rest of the balance sheet computation for in-class discussion and homework. To determine your debt after a month, just take your initial debt P, multiply by the factor $e^{r/12}$ for the growth of the debt over the first month, and subtract the amount you just paid off, namely M. We can write this as $P_1 = e^{r/12} P_0 - M$. It holds equally from any month to the next: $P_{n+1} = e^{r/12} P_n - M$, where P_n is your debt after n months.

How about your retirement account? Say you put M dollars every month into an interest bearing account. How much do you have after n months? It's the same formula, with an opposite sign because you're adding to your balance, not subtracting.

Checkpoint 133.

A guaranteed rate annuity works similarly. By the time you retire you have put P dollars into an account. (How did this happen? See Checkpoint 133) You hand this over to a company who guarantees you a certain APY every year, call it Y. Each year you also withdraw a fixed amount to live on, call it M.

Checkpoint 134.

The University of Pennsylvania's endowment works something like this. The balance increases by roughly 5% each year due to the growth of the investments and new donations. Meanwhile, during the year, the university spends roughly 3.4% of the present endowment. Unlike the formula for growth of a retirement fund or reduction of debt, this one is only approximate because the actual return varies. Nevertheless, it is useful for forecasting. Let E_n denote the size of the endowment after n years.

Checkpoint 135.

MATH 1070: Mathematics of Change, part I (article-1.html)

Calculus Group

9 Integrals

10.1 Area

Integrals compute many things, the most fundamental of these being area. The definition of area is more subtle than one might think. Most people's understanding of area is based on a physical concept of how much two-dimensional space is taken up. For example, if you have to paint an irregular flat shape, how much paint does it take?

Looking back at the treatment of area in the pre-college math curriculum, you can see the steps toward a mathematical definition. First, for rectangles with integer sides a and b, you can count the number of 1×1 squares needed to make the rectangle, leading to the area formula $A = a \times b$. From the physical point of view this is a formula, but from the mathematical point of view it is a definition, extended later to non-integer side lengths. Areas of triangles are not studied until much later. For right triangles with sides a and b and hypotenuse c, the area is shown to be equal $ab/2$ by showing that two of these fit together to make an $a \times b$ rectangle. This invokes a new principle: areas of congruent figures are equal. To compute the area of a parallelogram or trapezoid, the ***dissection principle*** is invoked: cutting up and rearranging the pieces of a figure preserves the area. These principles, all of which make intuitive and physical sense, are illustrated in Figure 10.1.1.

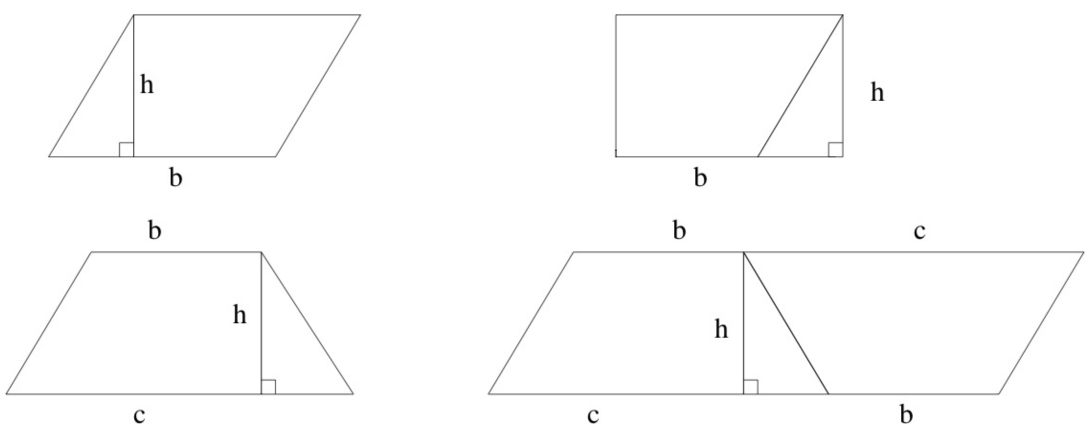


Figure 10.1.1. identifying congruent pieces of a dissection to evaluate areas of parallelograms and trapezoids

Checkpoint 136.

The area of a circle is introduced, usually without much explanation. Do you know why the area of a circle of radius r is equal to πr^2? One common explanation is that areas of similar figures are related by a scaling principle. Recalling that area has units of squared length, it makes sense that scaling a figure by λ should scale the area by λ^2. All circles are similar; if follows that the area of a circle should be Kr^2 for some constant K. We can name this constant π but that leaves a nagging question unanswered. Scaling also shows that the circumference of a circle should be proportional to the radius, therefore $C = K'r$ for some other constant K'. This turns out to be 2π. But why should K' be double K? An argument involving dissections and limits is shown in Figure 10.1.2.

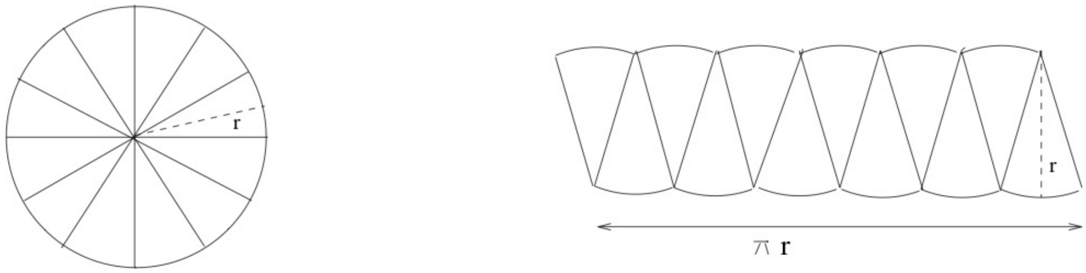


Figure 10.1.2. a limit of dissections relates the constants for circumference and area

Checkpoint 137.

Once limits are brought into the discussion, there is a way to define areas of much more general shapes. The idea is this: put as many non-overlapping squares of some small side length ε as you can inside the shape. These cover an area less than the area of the shape, but if ε is small, it seems credible that the area is getting close to the area of the shape. If the limit as $\varepsilon \to 0^+$ exists, this should be the area of the shape. Similarly, you could completely cover the shape with squares of side ε if you are willing to cover a slightly too big region. When ε is small, you don't cover too much extra area. The limit as $\varepsilon \to 0^+$ should also be the area of the shape. To make a long story short (you can hear the full story in Math 360), there are many shapes for which it is possible to prove that these two limits exist and are equal. For these shapes we can *define* area to be this common limiting value. This mathematical definition captures our existing physical intuition and is also consistent with the principles we already adpoted: congrunce, scaling and dissection.

10.2 Riemann sums and the definite integral

With this build up, we will give a mathematical definition for the area for a certain restricted class of shapes. These are rectangular on three sides but whose top is described by a continuous function. More precisely, let $a < b$ be real numbers and let f be positive and continuous on the closed interval $[a, b]$. We will define the area of the region R bounded on the left by the vertical line $x = a$, on the right by the vertical line $x = b$, on the bottom by the x-axis (the line $y = 0$), and on the top by the graph of f (the curve $y = f(x)$). This region is shown in Figure 10.2.1.

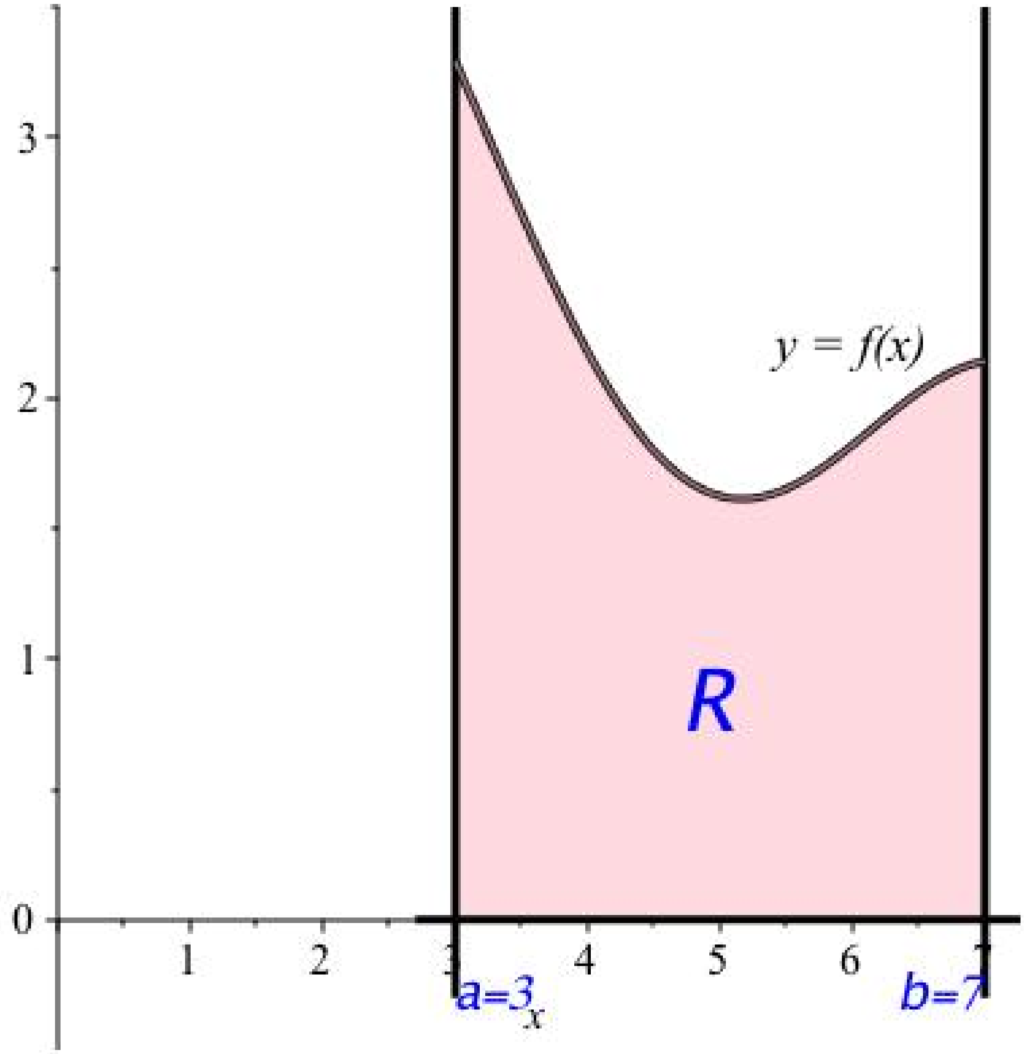


Figure 10.2.1. region between the x-axis and the graph of a function

We now define the lower and upper Riemann sums with n rectangles for a function f on an interval $[a, b]$. If you prefer a picture, refer to Figure 10.2.2.

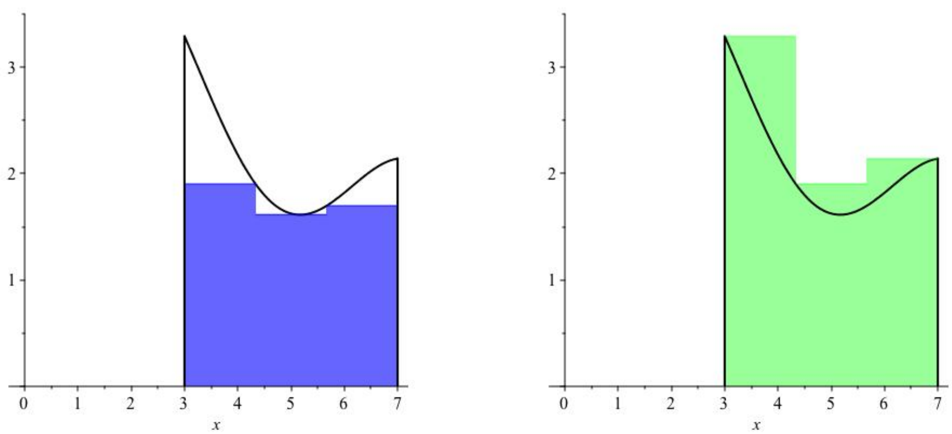


Figure 10.2.2. lower and upper Riemann sums

Checkpoint 138.

Definition 10.2.3. Let f be a nonnegative continuous function on an interval $[a, b]$ and let n be a positive integer. Let $I_1, \ldots, I_n$ denote the intervals you get when you divide $[a, b]$ into n equally sized intervals. For each interval I_k, let y_k be the minimum value of f on I_k and let R_k be the rectangle with base I_k on the x-axis and height y_k. The ***lower Riemann sum*** for f on $[a, b]$ with n rectangles is the sum of the areas of the rectangles R_k, for $1 \leq k \leq n$. The ***upper Riemman sum*** is defined similarly, with the maximum value instead of the minimum value on each interval.

Checkpoint 139.

Example 10.2.4. We are not given precise values for the function f in Figure 10.2.2, but we can estimate from the graph. The rectangles each have width $4/3$. The respective heights for the lower Riemann sum appear to be roughly $1.9, 1.6$ and 1.7, making the lower Riemann sum equal to $(4/3)1.9 + (4/3)1.6 + (4/3)1.7 = (4/3)5.2 \approx 6.93$. The upper Rieman sum is computed from rectangles with approximate heights $3.3, 1.9$ and 2.15, leading to a total area of $(4/3)7.35 = 9.8$.

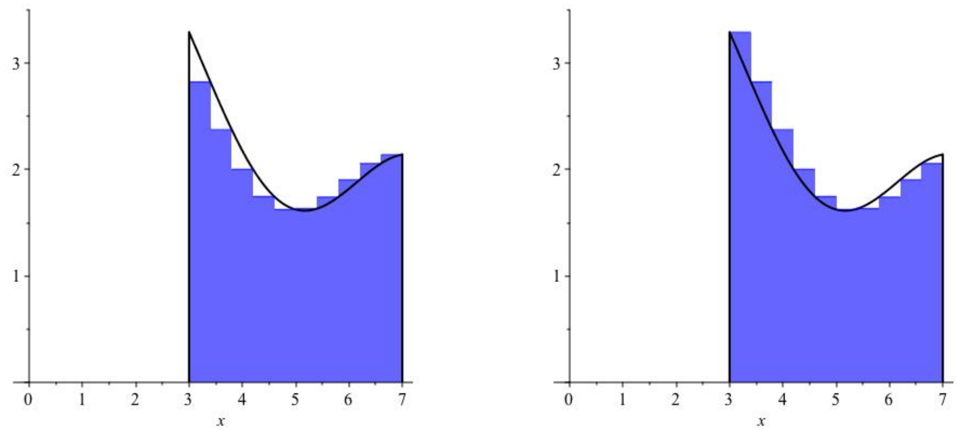


Figure 10.2.5. left and right Riemann sums

The ***left Riemann sums*** and ***right Riemann sums*** are defined similarly, except that instead of using the minimum or maximum values of the function on each sub-interval the left Riemann sums uses the value at the left endpoint of each interval I_k, while the right Riemann sum uses the value at the right endpoint of each sub-interval I_k. Examples are shown in Figure 10.2.5.

Checkpoint 140.

The upper and lower Riemann sums give upper and lower bounds on the area of the figure. The left and right Riemann sums are neither upper nor lower bounds for the area, but they are sandwiched in between the lower and upper Riemann sums, so they also converge to the area. They are useful because always choosing the left endpoint (or always choosing the right endpoint) leads to a simpler formula.

Checkpoint 141.

The values of the lower and upper Riemann sums in Figure 10.2.2 are approximately 6.9 and 9.8. These are not very close to each other, leaving considerable uncertainty about the true area. Replacing by the left (say) Riemann sums, we can program the sum into a computing device and compute for much greater values of n. If we increase n from 3 to 10, as in Figure 10.2.5, we find the Riemann sums come out to approximately 8.48 and 8.02 -- somewhat better. These are not necessarily bounds: the true value could be greater than both, or less than both, or in between. Replacing n by 50 gives 8.28. This is again not a bound, however the following theorem guarantees that as $n \to \infty$, this will converge to the area.

Theorem 10.2.6. The upper Riemann sums for any continuous function f on any closed interval $[a, b]$ converge as $n \to \infty$. The lower Riemann sums converge to the same value. It follows that you can let $y_k = f(x_k)$ for ***any***$x_k \in I_k$ and the sums of rectangle areas will still converge to this common limit.

Definition 10.2.7. The common limit in Theorem 10.2.6 is called the ***definite integral of*** f ***from*** a ***to*** b and is denoted $\int_a^b f(x)\,dx$.

Checkpoint 142. n.edu/~ancoop/1070/section-10.html#project-142)

Remark 10.2.8. The variable x is a bound variable; the notation $\int_a^b f(u)\,du$ would represent the same thing. Also, as in the notation for derivatives, you shouldn't try to interpret what the symbol du means on its own. It evokes the width of an infinitesimal rectangle, but you can't always count on it to behave nicely in equations.

Checkpoint 143. n.edu/~ancoop/1070/section-10.html#ex-properties-of-int)

10.3 Trapezoidal approximation

Sometimes it can be frustrating using Riemann sums because a lot of calculation doesn't get you all that good an approximation. You can see a lot of "white space" between the function f and the horzontal lines at the top of the rectangles that make up the upper or lower Riemann sum. If instead you let the rectangle become a right trapezoid, with both its top-left and top-right corner on the graph $y = f(x)$, then you get what is known as the *trapezoidal approximation*. The figure shows a trapezoidal approximation of an integral $\int_0^4 f(x)\,dx$ with five trapezoids. Note that the first and last trapezoid are degenerate, that is, one of the vertical sides has length zero and the trapezoid is actually a right triangle. It is perfectly legitimate for one or more of the trapezoids to be degenerate.

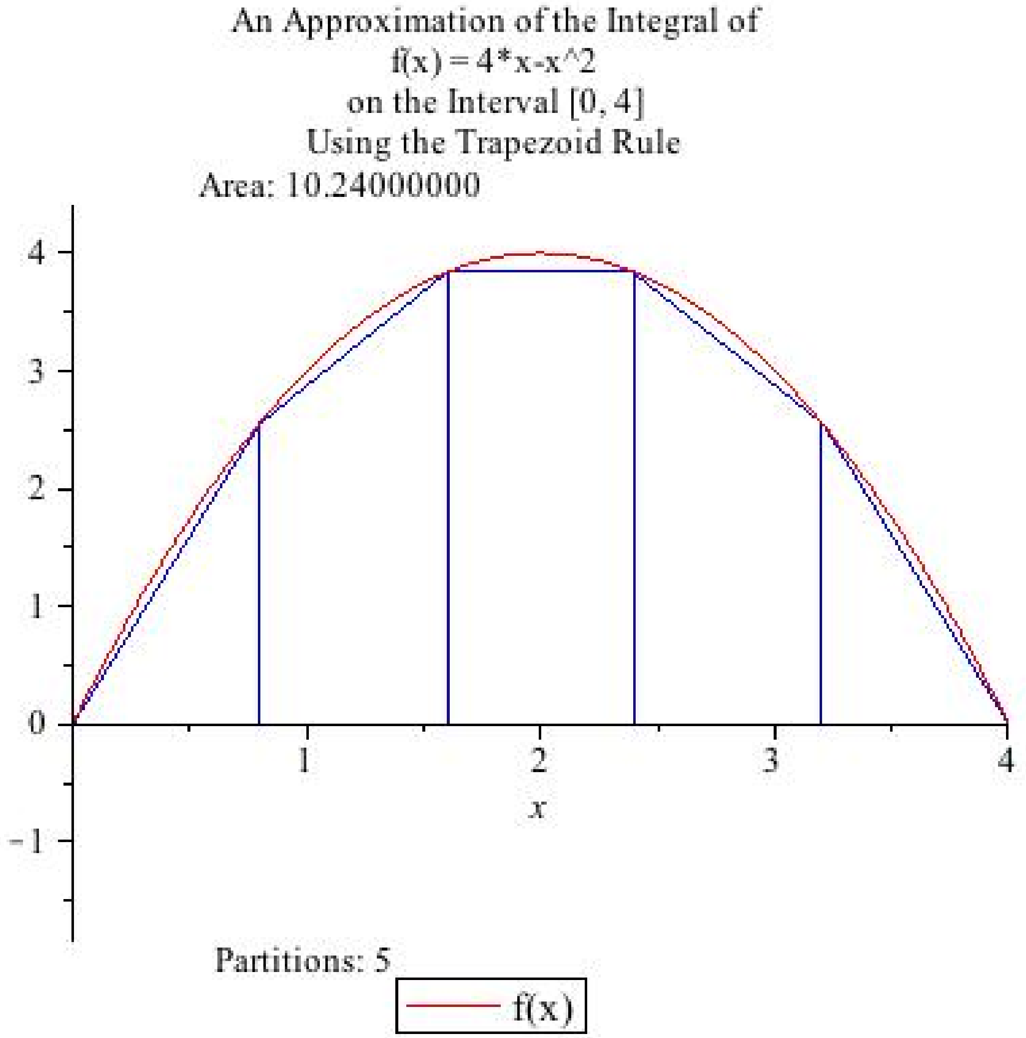


Figure 10.3.1.

Because the tops of the slices are allowed to slant, they remain much closer to the graph $y = f(x)$ than do the Riemann sums. Because the area of a right trapezoid is the average of the areas of the two rectangles whose heights are the value of f at the two endpoints, it is easy to compute the trapezoidal approximation: ***it is just the average of the left-Riemann sum and the right-Riemann sum corresponding to the same partition into vertical strips.***

Example 10.3.2. Let's compute the trapezoidal approximation for $\int_1^2 \frac{1}{1+x^2}$ with 10 trapezoids.

Averaging the left and right Riemann sums always gives a sum containing the $n-1$ common terms plus half the first term for the left Riemann sum and the last term for the right Riemann sum. In this case one gets

$$\frac{1}{2}\frac{f(1)}{10} + \frac{1}{2}\frac{f(0)}{10} + \sum_1^9 \frac{1}{10} f\left(1 + \frac{j}{10}\right).$$

The outcome of trapezoidal approximation in general can be summarized as,

Sum the values of f along a regular grid of x-values, counting endpoints as half, and multiply by the spacing between consecutive points.

The trapezoidal estimate is usually much closer than the upper or lower estimate, though it has the drawback of being neither an upper nor a lower bound. However, if you know the function to be concave upward then the trapezoidal estimate is an upper bound. Similarly if $f'' < 0$ on the interval then the trapezoidal estimate is an lower bound. In the figure, f is concave downward and the trapezoidal estimate is indeed a lower bound.

Example 10.3.3. The function $1/(1+x^3)$ is concave upward on $[1,2]$ (compute and see that the second derivative is a positive quantity divided by $(1+x^3)^3$) so the trapezoidal estimate should be not only very close but an upper bound. Indeed, the trapezoidal estimate is the average of the upper and lower previously computed and is equal to $0.25485\ldots$ which is indeed just slightly higher than the true value of $0.25425\ldots$.

Just as Riemann sums estimate by strips with constant height (degree zero) and trapezoids estimate by strips whose height is a linear function, you could imagine using higher degree polynomials (because you can still compute their

10.4 Interpretations of the integral

Area is the most visually obvious interpretation but there are many others. If material (or charge, or mass, etc.) is spread out unevenly over an interval, the ***density*** at any point is the amount of material per length near that point. It has units of material divided by length. The total amount of material in the interval is gotten by summing how the amount of material over small intervals. When the interval is small enough, we can estimate the amount of material as $f(x)$ times the length of the interval where x is any point in the interval. This is not exact because f generally will still vary over the interval, but not by much when x is small. The limit as the interval lengths go to zero will be $\int_a^b f(x)\,dx$ and will represent the total material.

Example 10.4.1. A 3-inch blade of grass is covered in mold. The amount of mold decreases up the blade because it is killed by sunlight. The density of mold per inch is $1000e^{-x/3}$ spores per inch at height x inches from the ground. The total number of spores on the blade of grass is given by $\int_0^3 1000e^{-x/3}\,dx$.

Checkpoint 144. (n.edu/~ancoop/1070/section-10.html#project-144)

Definition 10.4.2. average over an interval. The average of a quantity varying over an interval $[a,b]$ according to a function f is defined to be $\frac{1}{b-a}\int_a^b f(x)\,dx$.

Example 10.4.3. Suppose the temperature over a day is $f(t)$ degrees Celsius t hours after midnight. The average temperature over the day is then $\frac{1}{24}\int_0^{24} f(t)\,dt$.

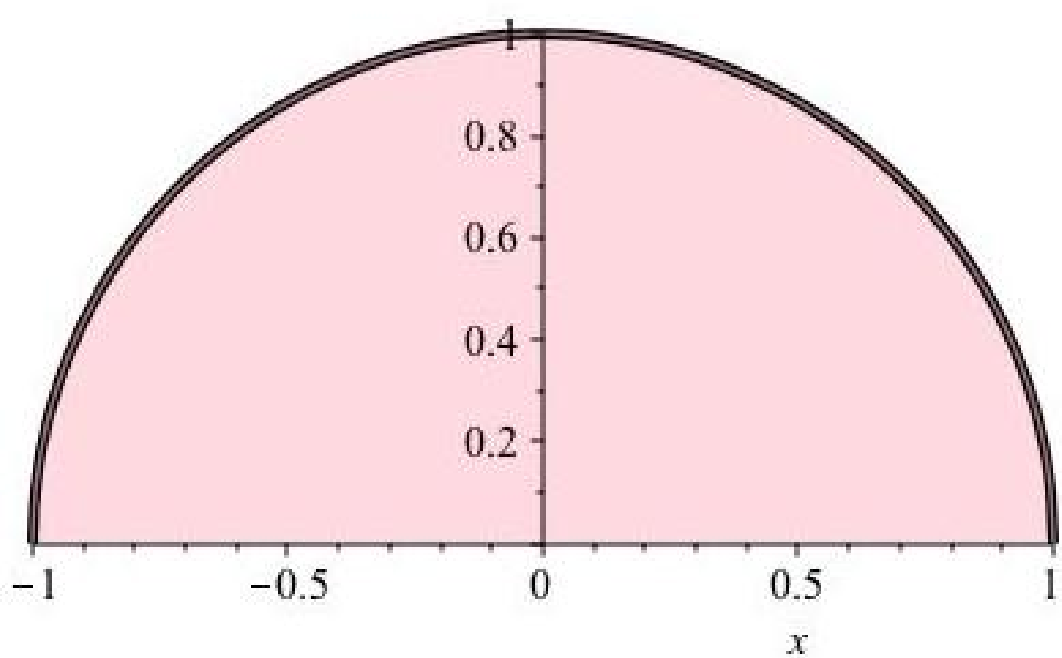
Checkpoint 145. (n.edu/~ancoop/1070/section-10.html#project-145)

An integral is a limit of a sum of rectangles' areas. The units are therefore the same units as the rectangles' areas. The rectangles live on a graph where the x-axis has units of the argument variable and the y-axis has units of the function. Therefore the rectangle units, hence the integral units, are units of the argument times units of the function. In the grass example, the function was density (spores per inch) and the argument was inches, therefore the integral had units of spores. It is a good thing that this agrees with our interpretation of the integral as the total number of spores. In the temperature example, f has units of temperature and t is in units of time, so the integral of f has units of temperature times time. This sounds like a strange unit but it's not unheard of. Severity of cold spells is measured, for example, in heating degree-days. The average is the integral divided by the time, so it is in units of temperature. Of course: the average temperature should be a temperature!

In physics there are countless things represented by integrals. One is the ***moment***. Suppose mass is spread out along $[a,b]$ with density f (you know what that means now, right?). Integrate f and you get the total mass. If instead you compute $\int_a^b x\, f(x)\, dx$ you get the ***moment of inertia***, which tells you how much the weight counts when balancing (imagine a teeter-totter pivoting on the origin), or how much torque is needed to produce a given angular acceleration.

In probability theory, random quantities can be discrete or continuous. If the random quantity X is discrete it means that there is a sset of values $x_1, x_2, \ldots$ such that probabilities for $X = x_k$ sum to 1. This could be a finite sum or the sum of an infinite sequence (you now know the definition of an infinite sum, right?). For continuous quantities, you need integrals. The probabilities for finding X to take various values are spread continuously over an interval (possibly an infinite interval such as the whole real line). There will be a ***probability density function*** f such that the probability of finding X in a given interval $[a,b]$ will be $\int_a^b f(x)\, dx$. We will say more about this in Unit 12 (ss-improper.html), after we have defined integrals where one or both of the limits of integration can be infinite.

Going back to the area interpretation, you may ask what about more general shapes? It turns out you don't really need straight sides. The vertical walls on the left and right sides of the regions Figure 10.2.1 and Figure 10.2.2 can disappear. For example, letting $f(x) = \sqrt{1-x^2}$ and $[a,b] = [-1,1]$ produces the upper half of a disk.



So far we have required f to be a nonnegative function. What if f is negative?

Let Δ_k denote the width of I_k. The most useful definition turns out to be that the integral is still the limit of sums of the quantities $\sum_k f(x_k) \cdot \Delta_k$ but we must interpret this as a new concept, called *signed area* rather than area. We won't worry too much about signed area; it just means we need to keep track of whether f is positive or negative before we know whether $\int_a^b f(x)\,dx$ computes area or its negative.

Figure 10.4.4 shows a function which is positive on $[0, 0.42]$ and negative on the interval $[0.42, 1]$. The integral $\int_0^1 f(x)\,dx$ will be slightly negative because it adds a positive area A_1 to a negative signed area A_2.

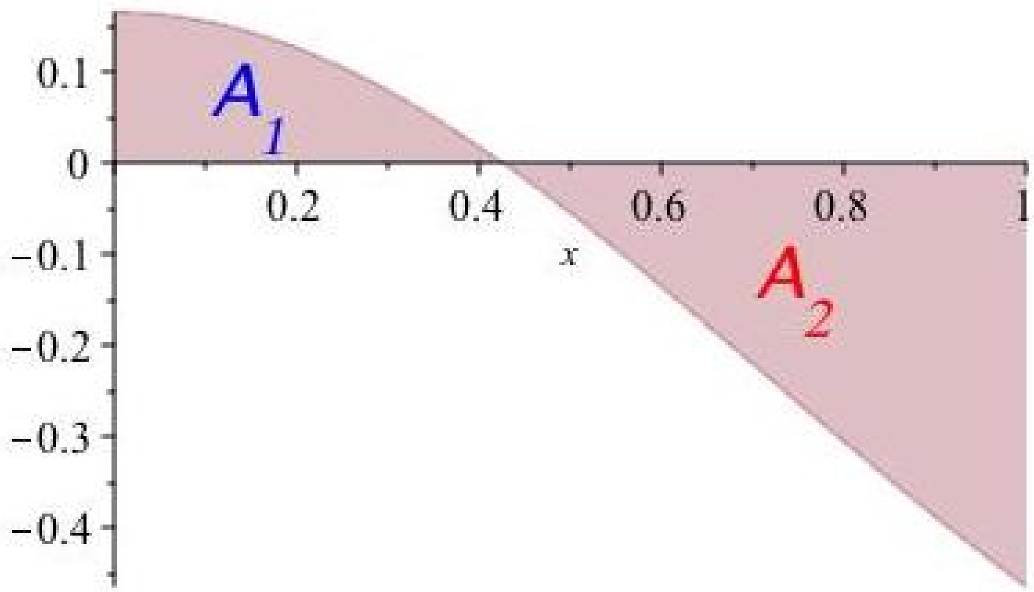


Figure 10.4.4. the signed area A_2 will be negative because it is below the x-axis

Another useful definition along the same lines switches the upper and lower limits.

Definition 10.4.5. If $a < b$ then $\int_b^a f(x)\,dx$ is defined to equal $-\int_a^b f(x)\,dx$.

Suppose f and g are functions such that $f \geq g$ on $[a, b]$. One interpretation of $\int_a^b [f(x) - g(x)]\,dx$ is that it is the area of the shape with upper boundary f and lower boundary g.

We started out computing areas of a very specific set of shapes, looking like three sides of a rectangle and a possibly curved upper boundary. Using the idea of upper and lower boundaries we can use integrals to give the area of a much greater variety of shapes.

Checkpoint 146.

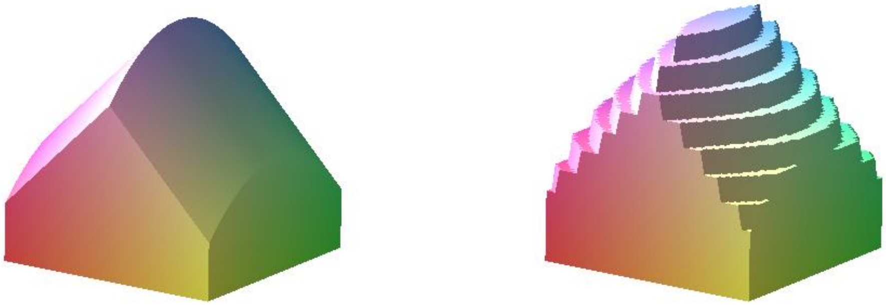


Figure 10.4.6. a solid volume (left) cut into slabs (right)

The examples of densities of quantities spread out along a line is somewhat limited. When quantities spread out, usually they spread over a region in a plane or in three dimensions.

The next course in this sequence covers multivariable integration. Still, there are some higher dimensional things you can do with ordinary integrals. One of these is to compute a volume of an object if you know the area of its cross-sections. Dividing the object into n very thin slabs, the volume of the k^{th} one is roughly the thickness Δ_k times the cross-sectional area of the k^{th} slab, call it A_k; see Figure 10.4.6.

The limit of $\sum_{k=1}^{n} A_k \Delta_k$ should give the volume. Line up the slabs so that the x-axis goes perpendicular to the slabs. This limit looks awfully similar to the limit of $\sum_{k=1}^{n} f(x_k)\Delta_k$ where x_k is any point on the x-axis inside the k^{th} slab and f is the function telling the cross-sectional area at every x-value. Therefore, the volume is computed by $\int_a^b f(x)\,dx$ where a and b are the x-values at the first and last slab respectively.

Example 10.4.7. volume of a pyramid. We write an integral for volume of a pyramid whose base is a square of side length s and whose height is h. It corresponds best to the description above if we orient it so the height is measured along the x-direction with the apex at the origin. See Figure 10.4.8. The cross-section is a square with side increasing linearly from 0 to s as x increases from 0 to h. Thus, the side length is given by $\ell(x) = (s/h)x$, hence the cross-sectional area is given by $f(x) = (s/h)^2 x^2$ between $x = 0$ and $x = h$. The volume is therefore given by $\int_0^h (s/h)^2 x^2\,dx$. When you learn to compute integrals, this will turn out to be a pretty easy one.

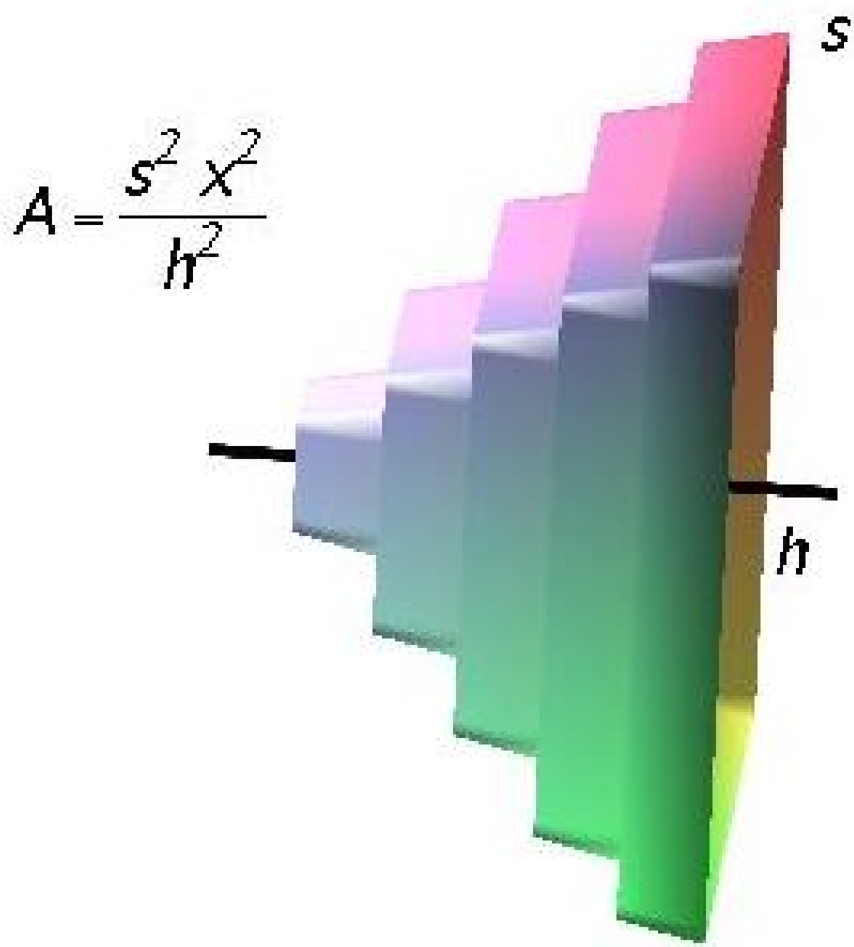


Figure 10.4.8. a pyramid, cut into slabs along the x-direction

10.5 Estimating sums via integrals

$\ln x 1/x . 1/x \ln x.$

Example 10.5.1. harmonic sum estimated by an integral. Problem: estimate the 100^{th} harmonic number $1 + 1/2 + 1/3 + \cdots + 1/100$. To solve this, we may as well estimate $H_n := \sum_{k=1}^{n} 1/k$ for any positive integer n. Summing $1/n$ looks a lot like integrating $1/x$. In fact, suppose we write a Riemann sum for $\int_1^n 1/x\,dx$ that has precisely $n-1$ rectangles. Then the intervals I_k are just the intervals $[1,2], [2,3], \ldots, [n-1, n]$. Even better, we can make areas of the rectangles be exactly the same as in the sum. We just need to use the upper Riemann sum: $1 + 1/2 + \cdots + 1/(n-1)$; see the left-hand side of Figure 10.5.2 for a picture of this when $n = 9$.

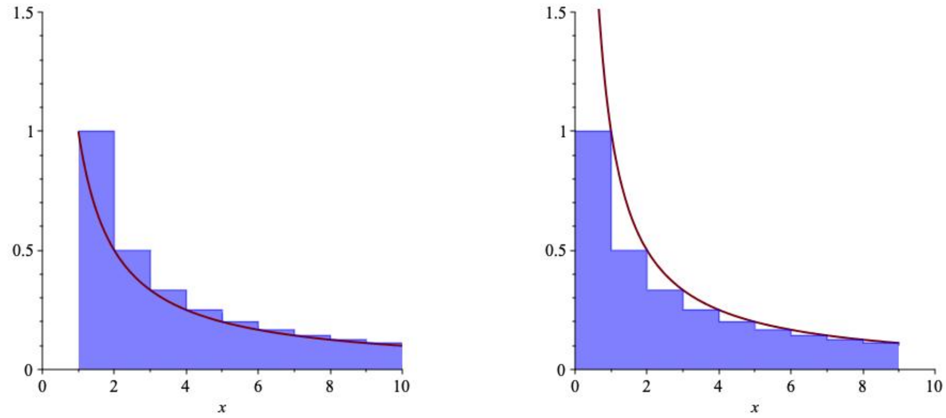


Figure 10.5.2. representing the harmonic sum as upper and lower Riemann sums

Figure 10.5.2 shows that H_{n-1} is an upper Riemann sum for $\int_1^n 1/x\,dx$. It turns out (see Proposition 11.1.6 below) that this integral has a nice formula:

$$\int_1^n \frac{1}{x}\,dx = \ln n - \ln 1 = \ln n\,.$$

Therefore, we have shown the bound $H_{n-1} \geq \ln n$. In particular, choosing $n = 101$, we see that $H_{100} \geq \ln 101 \approx 4.615$.

Is this an upper bound or a lower bound? It depends on your point of view. If we were trying to figure out the integral up to 100, H_{100} would be an upper bound on the value. But in this case we know the integral and are trying to estimate H_{100}. The integral provides a lower bound, in this case 4.615.

What about an upper bound on H_{100}. The obvious thing is to see if we can make the same sum be a lower Riemann sum. Watch what happens when you try to do this. Take the graph, shift all the rectangles one unit left, and voilà! (See the right-hand side of Figure 10.5.2.) This shows that H_{100} is a lower Riemann sum for a slightly different integral, namely $\int_0^{100} \frac{1}{x}\,dx$. Alas, this is not an integral we can do because $1/x$ is not continuous at $x = 0$. In fact, when we study improper integrals, we will see this evaluates to $+\infty$. Sure, we get the upper bound $H_{100} \leq \infty$, but that is hardly useful. All is not lost, however, if we use some common sense. The same picture shows that an upper bound for the harmonic sum starting at 2 instead of 1 is

$$\sum_{k=2}^{100} \frac{1}{k} \leq \int_1^{100} \frac{1}{x}\,dx = \ln 100 \approx 4.605\,.$$

So, adding back the 1, we see that $H_{100} \leq 1 + \ln 100 \approx 5.605$. This is about as good as we can do with the techniques we have so far: $4.615 \leq H_{100} \leq 5.605$. For the record, $H_{100} \approx 5.1874$.

10 Computing integrals

All continuous functions have integrals, but not all of the integrals have clean familiar formulas. For example, the definite integral $\int_3^8 \frac{1}{\ln x}\, dx$ is a well defined quantity; indeed $\int_a^b \frac{1}{\ln x}\, dx$ is well defined for any $b > a > 1$, but the function $b \mapsto \int_a^b (1/\ln x)\, dx$ is not equal to any combination of named functions such as powers, logs, exponentials and trig functions. The same is true of the normal (bell curve) density function e^{-x^2}, or $\sqrt{\sin x}$ or $\sqrt{1-4x^2}/\sqrt{1-x^2}$. The prevalence of functions like this is the reason we need good numeric approximations to integrals (as discussed in Subsection 10.2 (section-10.html#sec-riemannsum) and Subsection 10.3 (section-10.html#sec-trapezoid)). In the remainder of this section we concentrate on anti-derivatives for which reasonably nice exact expressions exist.

11.1 The fundamental theorem of calculus

Somewhat surprisingly, finding clean formulas for integrals involves *derivatives*. To see how this works, we look at the *indefinite* integral. Replacing the upper limit on the integral by a variable yields a function of that variable. To say this in another way, we may consider $\int_a^b f(x)\, dx$ as a function of the free variables a and b (it can't be a function of x because x is a bound variable). Let a remain a constant but consider b to be a variable. We then have a function, $b \mapsto \int_a^b f(x)\, dx$. Denote this function by G, in other words $G(b) := \int_a^b f(x)\, dx$.

Example 11.1.1. Let $f(x) := 3x$ and $a = 0$. Then $G(b) := \int_0^b 3x\, dx$. Definite integrals compute area, hence $G(b)$ is the area of the triangle with vertices at the origin, $(b, 0)$ and $(b, 3b)$. The triangle area formula gives $G(b) = (3/2)b^2$.

For fun (we have a warped sense of fun), compute G'. That's an easy one: $G'(b) = 3b$. Note that this is the integrand of the original integral, with the free variable b in place of the bound variable x. This is not a coincidence, as the following theorem asserts.

Theorem 11.1.2. Fundamental Theorem of Calculus. Let f be a continuous function on an interval $[a, c]$. For $b \in (a, c)$, let $G(b) := \int_a^b f(x)\, dx$. Then $G'(b) = f(b)$.

Sketch of proof. The derivative from the derivative from the right is given by

$$G'(b^+) = \lim_{h \to 0^+} \frac{G(b+h) - G(b)}{h} \,.$$

When h is small, the value of $G(b+h)$ is very well approximated by $G(b) + hf(b)$; in the picture at the right, $G(b)$ is the blue area and $G(b+h)$ is the blue area plus the shaded black and white area. Plugging this in gives $\frac{G(b)+hf(b)-G(b)}{h} = f(b)$. To turn this into a proof, you need to use continuity of f to show that the error replacing $G(b+h)$ by $G(b) + hf(b)$ is $\ll h$, so the approximation does not affect the limit. You already know enough to understand the argument, but in the interest of time, the details are left to a course in mathematical analysis.

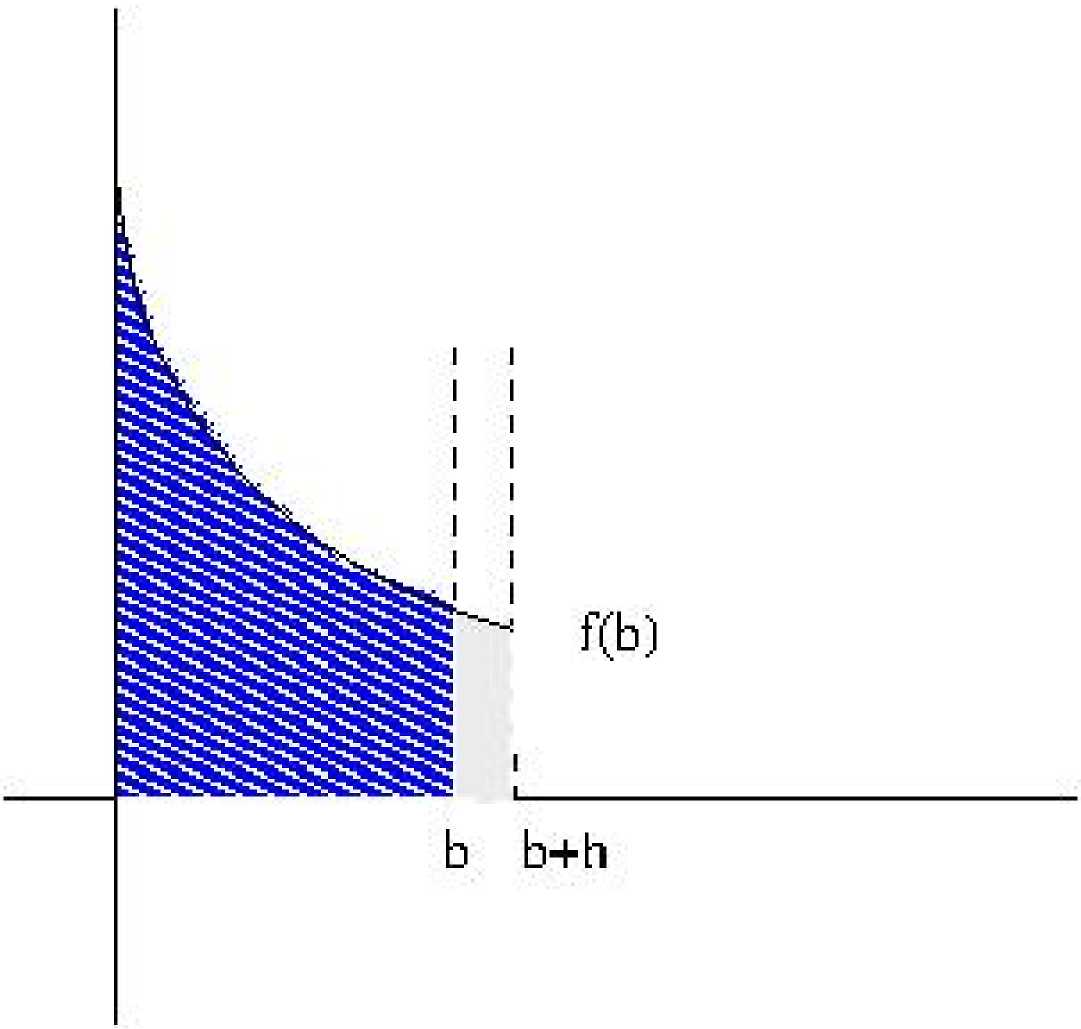


Figure 11.1.3.

anti-derivatives. The Fundamental Theorem of Calculus says we can evaluate integrals of f if we know a function G whose derivative is f. That motivates the next definition.

Definition 11.1.4. An ***anti-derivative*** of a function f is any function G such that $G' = f$.

How do we find anti-derivatives? The next chapter is entirely about computing these. Like rules for differentiation, rules for anti-differentiation start from a collection known results. For derivatives, we obtained these from the definition by computing limits. For anti-derivatives, we will get these simply by remembering some basic derivatives. The simple rule yielding the derivative of a polynomial may be run backwards. So for example the monomial ax^m has anti-derivative $\frac{a}{m+1}x^{m+1}$. We can sum these, obtaining the anti-derivative of any polynomial: an anti-derivative of $\sum_{k=0}^{m} a_k x^k$ is given by $\sum_{k=0}^{m} \frac{a_k}{k+1} x^{k+1}$. In fact this works for negative or fracdtional powers, as long as the power is not -1.

Checkpoint 147.

We say "***an*** anti-derivative" rather than "***the*** anti-derivative" because for any given f, is always more than one anti-derivative. The functions G and $G + c$, where c is a constant, have the same derivative, so one is an anti-derivative of f if the other is. This is the only way anti-derivatives can differ.

Once you know the value of the anti-derivative at any point, it is easy to reconstruct the correct anti-derivative as an integral, as in the following example.

Example 11.1.5. Suppose G is an anti-derivative of f and $G(3) = 7$. We will look for an anti-derivative of the form $G(b) = c + \int_a^b f(x)\,dx$. To write G as an integral with a variable upper limit, begin by choosing the constant for the lower limit. The most convenient choice is 3, because we are supposed to know the value of G at 3. The function $b \mapsto \int_3^b f(x)\,dx$ is zero at 3, so we will need to add 7. We therefore choose

$$G(b) = 7 + \int_3^b f(x)\,dx\,.$$

For concreteness, let's see how this works with the example from above: $f(x) = 3x$. Then $G(b) = 7 + \int_3^b 3x\,dx$. We already computed $\int_0^b 3x\,dx = \frac{3}{2}b^2$ and similarly $\int_0^3 3x\,dx = (3/2)3^2 = 27/2$. Subtracting, $\int_3^b 3x\,dx = (3/2)b^2 - 27/2$. Thus the anti-derivative we are looking for is $7 + (3/2)b^2 - 27/2 = (3/2)b^2 - 13/2$.

Checkpoint 148.

This example shows a general principle, which we record as a proposition.

Proposition 11.1.6. computing definite integrals with anti-derivatives. The definite integral $\int_a^b f(x)\,dx$ is equal to $G(b) - G(a)$, also denoted $G|_a^b$, when G is any anti-derivative of f.

Remark 11.1.7. This implies that $H(b) - H(a) = G(b) - G(a)$ when H is any other anti-derivative of f. In other words, differences of an anti-derivative at a specified pair of points do not depend on which particular anti-derivative was chosen.

Checkpoint 149.

11.2 Remembering and guessing

Computing derivatives, as you saw in Unit 5 (sec-comp-der.html), rests on combination rules and working out some basic cases. For anti-derivatives the same is true, with" working out" replaced by "remembering". In other words, if you remember what the derivative of f is, then you know how to compute an anti-derivative of f'. This is how we computed anti-derivatives for polynomials, for example. The strategy is then: (1) list the derivatives we already know, organized in a way that allows us to query what function goes with a given derivative; and (2) give combining rules for anti-derivatives. This gives the following proposition. Note that in each case, remembering allows us to identify just one of the antiderivatives; we trust you can compute the others from that.

we use an integral sign without upper and lower limits to denote the antiderivative: e.g., $\int (3x^2+1)\,dx$ is equal to x^3+x, plus any constant. We usually write this as x^3+x+C. By custom, we don't change the variable. In previous sections, for example, we were careful to write $\int_0^b (3x^2+1)\,dx$ as a function of b, namely b^3+b. But when writing the indefinite integral we tend to write $\int (3x^2+1)\,dx = x^3+x+C$, not b^3+b+C. This is because it's shorthand for

> *The indefinite integral of the function $x \mapsto 3x^2+1$ is any function $x \mapsto x^3+x+C$.*

The variable x is bound, so the choice of letter does not affect the meaning.

Proposition 11.2.1. The following basic anti-derivatives are computed by reversing Proposition 5.1.8.

1. $\int x^m = \frac{1}{m}x^{m-1} + C$ *as long as* $m \neq 0$.
2. $\int \frac{1}{x}\,dx = \ln x + C$
3. $\int \cos x\,dx = \sin x + C$
4. $\int \sin x\,dx = -\cos x + C$
5. $\int \sec^2 x = \tan x + C$
6. $\int e^x\,dx = e^x + C$
7. $\int \frac{1}{\sqrt{1-x^2}}\,dx = \arcsin x + C$
8. $\int \frac{1}{1+x^2}\,dx = \arctan x + C$

Checkpoint 150.

The derivative of a sum or difference is the sum or difference of the derivatives. The derivative of $c \cdot f$ is c times the derivative of f for any real constant c. This leads immediately to the following proposition.

Proposition 11.2.2. linearity of the anti-derivative.

$$\int [f(x) + g(x)]\, dx = \int f(x)\, dx + \int g(x)\, dx\,;$$
$$\int c \cdot f(x)\, dx = c \cdot \int f(x)\, dx\,.$$

Proof of the first statement. Let F be an anti-derivative of f and G be an anti-derivative of G. Thens $(F+G)' = F' + G' = f + g$ therefore $(F+G)$ is an antiderivative of $f+g$.

Checkpoint 151. n.edu/~ancoop/1070/section-11.html#project-151)

The word "anti-derivative" is a mouthful and so is the verb form "anti-differntiate". Because computing integrals comes down to anti-differentiation, common practice is to use the verb ***integrate*** in place of "anti-differentiate". We also call an anti-derivative an "integral". Proposition 11.2.1 and Proposition 11.2.2 allow us to compute some more integrals.

Example 11.2.3. Let's compute the integral of $\frac{a\cos x + b/\cos x}{\cos x}$. Simplifying,

$$\frac{a\cos x + b/\cos x}{\cos x} = a + b\sec^2 x\,.$$

Therefore

$$\begin{aligned}\int \frac{a\cos x + b/\cos x}{\cos x}\, dx \\ &= \int [a + b\sec^2 x]\, dx \\ &= \int a\, dx + b \int \sec^2(x)\, dx \\ &= ax + \tan x + C\end{aligned}$$

Checkpoint 152. n.edu/~ancoop/1070/section-11.html#project-152)

Example 11.2.3 should worry you. Does it seem a bit contrived? The expression $\frac{a\cos x + b/\cos x}{\cos x}$ just happens to simplify into two expressions covered by the list of cases in Proposition 11.2.1. If that seems like a piece of luck, it is. With only Proposition 11.2.1 and Proposition 11.2.2 you won't get very far. The next two sections give two rules for combining integrands that will greatly increase your ability to integrate. Keep in mind though, that in some sense you are still lucky whenever you can compute an analytic expression for an anti-derivative: many anti-derivatives have no nice formula.

11.3 Integration by parts

The sum rule for derivatives is simple enough that it leads directly to the first statement of Proposition 11.2.2, which is an identical rule for anti-derivatives. There is also a product rule, but it does not lead directly to an identical rule for anti-derivatives. That's because the product rule for derivatives is not symmetric. The derivative of fg is not $f'g'$ but rather $f'g + g'f$. When we run the product rule backwards, we get

$$\int \left[f'(x)g(x) + g'(x)f(x)\right] dx = f(x)g(x) + C\,. \tag{11.3.1}$$

The problem is, this doesn't tell us how to integrate a product such as $f'g'$! (11.3.1) is great if someone asks us to compute the anti-derivative of $f'g + g'f$, but this is rare, harder to spot, and does not answer the question as to the anti-derivative of a product.

The best we can do is to exploit (11.3.1) as much as we can. This leads to the following proposition.

Proposition 11.3.1. integration by parts. Let u and v be differentiable functions. Suppose $u'v$ is known to have anti-derivative G. Then $v'u$ has anti-derivative $uv - G$. In a single equation,

$$\int v'u\,dx = uv - \int u'v\,dx\,. \tag{11.3.2}$$

Proof. This is just the product rule for derivatives run in reverse: $(uv)' = u'v + v'u$, therefore

$$(uv - G)' = u'v + v'u - G' = v'u\,.$$

The way this works in practice is that when integrating an expression, you try to identify the expression as $v'u$ for some functions u and v. Then you check whether you already know the anti-derivative to $u'v$. If so, you subtract this from uv and you are done. Sometimes there are several possible ways to do this, in which case you may have to try them all until you find one that works.

Example 11.3.2. Use integration by parts to integrate xe^x. Oviously this decomposes as a product of x and e^x. One of these should be v' and the other should be u. Let's try setting

$$\begin{aligned} v' &= x\,; \\ u &= e^x \end{aligned}$$

At first this goes smoothly: the expression we chose for v has a known anti-derivative and the one we chose for u has a known derivative, therefore we can find v and u':

$$\begin{aligned} v &= \frac{x^2}{2} + C\,; \\ u' &= e^x \end{aligned}$$

Unfortunately the next step doesn't work: $u'v = e^x(x^2/2 + C)$, which is not something whose anti-derivative we recognize no matter what choice we make for the constant C.

Back to the drawing board. Let's try the other possible assignment of roles:

$$v' = e^x\,;$$
$$u = x$$

Again it goes smoothly at first: the expression we chose for v has a known integral e^x and the one we chose for u has a known derivative 1, therefore

$$v = e^x + C\,;$$
$$u' = 1$$

Now we're in better shape. Choose $C = 0$ (usually this works if anything does). Then $u'v = e^x$, for which an integral is known, namely e^x. Therefore,

$$\int xe^x\,dx = \int uv'\,dx = uv - \int u'v\,dx = xe^x - \int e^x\,dx = xe^x - e^x + C\,.$$

We did a long-winded example to show you the process of trial and error and to show how each step works. What would have happened if we chose a different value of C? It turns out it always works exactly as well.

Checkpoint 153. n.edu/~ancoop/1070/section-11.html#project-153)

It usually takes several worked examples and a lot of practice before integration by parts feels natural. Because "a lot of practice" means different things to different people, we include only a few mandatory self-check and howework problems, putting a greater number online for those who want to practice.

Example 11.3.3. Compute the definite integral $\int_0^{2\pi} x\sin(x)\,dx$. We start with the indefinite integral, which we compute by parts. Based on what happened with xe^x, let's decide to start with the choice $u = x, v' = \sin x$. Then $v = -\cos x$ and $u' = 1$, which yields

$$\int x\sin(x)\,dx = -x\cos x - \int(-\cos x)\,dx = -x\cos x - (-\sin x) = \sin x - x\cos x + C\,.$$

Evaluating the definite integral (notice we chose $C = 0$),

$$\begin{aligned}\int_0^{2\pi} x\sin(x)\,dx &= [\sin x - x\cos(x)]_{x=2\pi} - [\sin x - x\cos(x)]_{x=0}\\ &= [\sin(2\pi) - 2\pi\cos(2\pi)] - [\sin(0) - 0\cdot\cos(0)]\\ &= -2\pi\end{aligned}$$

Checkpoint 154. n.edu/~ancoop/1070/section-11.html#project-154)

Here are a few more tips to help you use integration by parts. Also, you should see a notational variation that is common in textbooks and on the web. Instead of $\int v'u\,dx = uv - \int u'v\,dx$, people sometimes write

$$\int u\,dv = uv - \int v\,du\,.$$

Because u and v are functions of x, you can think of $du := u'(x)\,dx$ and $dv := v'(x)\,dx$, whereby this form of the identity comes out to exactly the same thing as (11.3.2)

Repeated integration by parts. Sometimes integration by parts doesn't quite get you to an expression $u'v$ that you know how to evaluate, but it gets you closer, so that repeating the integration by parts solves the problem.

Example 11.3.4. We'll compute $\int x^2 e^x$. Letting $v' = e^x$ and $u = x^2$ gives

$$\int x^2 e^x\,dx = x^2 e^x - \int 2xe^x\,dx\,.$$

That last expression isn't covered by Proposition 11.2.1 but we just saw (take out the constant factor 2) that it can be done by parts and integrates to $2(xe^x - e^x) = 2(x-1)e^x$. Therefore,

$$\int x^2 e^x\,dx = x^2 e^x - 2(x-1)e^x = (x^2 - 2x + 2)e^x\,.$$

It should be apparent you can integrate $p(x)e^x$ this way for any polynomial p. Some textbooks have a separate algorithm for this called tabular integration. We won't be teaching that, but you can google it if you ever need the anti-derivative of $p(x)e^x$ where $p(x)$ has degree more than, say, 3 (doing it by hand gets longer and more complicated as the degree of p grows). To see how this will go, try the following exercise, which is about as much as we would ever ask you to do by hand.

Checkpoint 155.

Don't forget v' could be 1. You can always decompose any expression as itself times 1. In the langauge of $v\,du$ and $u\,dv$, that says $\int f(x)\,dx$ can always be thought of as $u\,dv$ where $u(x) = f(x)$ and $dv = dx$, that is, $v' = 1$. This only sometimes works but it's good to know.

Example 11.3.5. Compute $\int \ln(x)\,dx$. There's only one term to decompose so we pretty much have to use the $dv = dx$ trick. Setting $u(x) = \ln x$ and $dv = dx$, gives (recalling that the derivatve of $\ln x$ is $1/x$),

$$\int \ln(x)\,dx = (\ln x)(x) - \int x \cdot \frac{1}{x}\,dx = x\ln x - \int 1 \cdot dx = x\ln x - x + C\,.$$

This is a good one to memorize - it's very useful to recall quickly how to integrate the natural log.

11.4 Substitution

(ancoop/1070/section-11.html#ss-substitution)

Integration by parts is what you get from reversing the product rule. Reversing the chain rule is called ***substitution***. You can probably guess what it says. The chain rule says $(d/dx)f(g(x)) = f'(g(x))g'(x)$. Therefore, we need a rule to tell us that $\int f'(g(x))g'(x)\,dx = f(g(x))$. This gives the simplest form of the substitution method.

Theorem 11.4.1. Suppose g is differentiable on an interval (a,b) and let I (which will also be a closed interval) be the range of g. Suppose h is differentiable on I. Then

$$\int h'(g(x))g'(x)\,dx = h(g(x)) + C\,.$$

Writing f for h', this becomes

$$\int f(g(x))g'(x)\,dx = \left(\int f\right) \circ g \tag{11.4.1}$$

where the identity $f = h'$ allows us to write the indefinite integral $\int f$ in place of h on the right. This second form is sometimes clearer because we often arrive at the form $f(g(x))g'(x)$ before we have identified the antiderivative of f, hence it makes sense for the right-hand side to leave $\int f$ unevaluated.

Example 11.4.2. We compute the integral of $\frac{(\ln x)^2}{x}$. The numerator $(\ln x)^2$ looks like a composition $f(g(x))$ where $f(x) = x^2$ and $g(x) = \ln x$. We are in luck because $g'(x) = 1/x$ so there is alread a g' sitting there. The expression to be integrated looks like $f(g(x))g'(x)$, so applying (11.4.1),

$$\int \frac{(\ln x)^2}{x}\,dx = \left(\int x^2\right) \circ \ln\,.$$

The indefinite integral of x^2 is $x^3/3$, so the final answer is that the indefinite integral of $(\ln x)^2/x$ is $(\ln x)^3/3 + C$.

Checkpoint 156. (n.edu/~ancoop/1070/section-11.html#project-156)

The substitution rule is very often stated in the language of science, with a variable u, thought of as a physical quantity related to the variable x via $u = g(x)$. Instead of a theorem, this version is usually described as a procedure.

1. Change variables from x to u (hence the common name "u-substitution")
2. Keep track of the relation between dx and du
3. If you chose correctly you can now do the u-integral
4. When you're done, substitute back for x

Again, we let g be the function relating u to x via $u = g(x)$, and again you need hypotheses, namely the ones stated in Theorem 11.4.1). Then $du = g'(x)\, dx$. Usually you don't do this kind of substitution unless there will be an $g'(x)\, dx$ term waiting which you can then turn into du. Also, you don't do this unless the rest of the occurences of x can also be turned into u. If g has an inverse function, you can do this by substituting $g^{-1}(u)$ for x everywhere. Now when you reach the fourth step, it's easier because you can just plug in $u = g(x)$ to get things back in terms of x.

This notation gives a particularly nice simplification when $u = x + c$ for some constant c. Replacing x by $x + c$ is called a ***translation***. In the first unit of the course, we discussed what this does to the graph. It is a very natural change of variables, corresponding to a different starting point for a parametrization.

Example 11.4.3. translation. Compute the indefinite integral of $\sqrt{x+6}$. Let $u = x + 6$. Then $du = dx$. Integrating the $1/2$ power (one of the basic facts in Proposition 11.2.1),

$$\int \sqrt{x+6}\, dx = \int \sqrt{u}\, du = \frac{2}{3} u^{3/2} = \frac{2}{3}(x+1)^{3/2} + C\,.$$

The moral of this story is that you can "read off" integrals of translations. For example, knowing $\displaystyle\int \cos x\, dx = \sin x + C$ allows you to read off $\displaystyle\int \cos(x - \pi/4)\, dx = \sin(x - \pi/4) + C$. Don't let this example fool you into thinking it works this way for functions other than translations. Thinking that $\displaystyle\int \cos(\sqrt{x})\, dx = \sin(\sqrt{x}) + C$ is wrong; it is the calculus equivalent of the algebra mistake $(a+b)^2 = a^2 + b^2$.

Here's an example of u-substitution with something other than a translation.

Example 11.4.4. Let's compute $\displaystyle\int \sin^n x \cos x\, dx$.

Solution: substitute $u = \sin x$ and $du = \cos x\, dx$. This turns the integral into $\displaystyle\int u^n\, du$ which is easily valuated as $u^{n+1}/(n+1) + C$. Now plug back in $u = \sin x$ and you get the answer

$$\frac{\sin^{n+1} x}{n+1} + C\,.$$

You might think to worry whether the substitution had the right domain and range, was one to one, etc., but you don't need to. When computing an indefinite integral you are computing an anti-derivative and the proof of correctness is whether the derivative is what you started with. You can easily check that the derivative of $\sin^{n+1} x/(n+1)$ is $\sin^n x \cos x$.

After a translation, the next simplest substitution is a ***dilation***, where $u(x) = cx$ for some nonzero real number c. This is the other case in which substitution always succeeds: if you can integrate $f(x)$ you can always integrate $f(cx)$. We leave it to you to work this out, first in an example, then in the general case.

Checkpoint 157.

When evaluating a definite integral you can compute the indefinite integral as above and then evaluate. A second option is to change variables, including the limits of integration, and then never change back.

Example 11.4.5. Let's compute $\int_1^2 \frac{x}{x^2+1}\,dx$.

If we let $u = x^2 + 1$ then $du = 2x\,dx$, so the integrand becomes $(1/2)\,du/u$. If x goes from 1 to 2 then u goes from 2 to 5, thus the integral becomes

$$\int_2^5 \frac{1}{2}\frac{du}{u} = \frac{1}{2}(\ln 5 - \ln 2)\,.$$

Of course you can get the same answer in the usual way: the indefinite integral is $(1/2)\ln u$; we substitute back and get $(1/2)\ln(x^2+1)$. Now we evaluate at 2 and 1 instead of 5 and 2, but the result is the same: $(1/2)(\ln 5 - \ln 2)$.

Backwards substitution. There are times when the best substitution is of the form $x = g(u)$ rather than $u = g(x)$. No matter what f and g are, the substitution $x = g(u),\ dx = g'(u)\,du$ always leads to a new integral, it's just hard to choose g in a way that makes the new integral simpler than the old one. It turns out there are some integrals, not apparently involving trig functions, where substituting $x = g(u)$ for some trig function g will magically unlock a dead end. Knowing tricks for dealing with a wide class of anti-derivative extractions is not the aim of this course, therefore we will not be featuring this method in the text. If you're interested in seeing one of these, try googling "integrate sqrt(1-x^2)".

Looking it up. Math is about understanding relations of a precise nature, about abstraction, and about making models of physical phenomena. It is also about building a library of computational tricks, but that's only a small part of math, and it's somewhat time-consuming. We have taught you what we think it is reasonable for you to know and remember -- to have in your quick-access library. For all the other integrals currently known to mankind, there are lookup tables. The following integral table is stolen from a popular calculus book. Use it as a reference as needed.

1. $\int k\, dx = kx + C$ (any number k)

2. $\int x^n\, dx = \dfrac{x^{n+1}}{n+1} + C$ $(n \neq -1)$

3. $\int \dfrac{dx}{x} = \ln|x| + C$

4. $\int e^x\, dx = e^x + C$

5. $\int a^x\, dx = \dfrac{a^x}{\ln a} + C$ $(a > 0,\ a \neq 1)$

6. $\int \sin x\, dx = -\cos x + C$

7. $\int \cos x\, dx = \sin x + C$

8. $\int \sec^2 x\, dx = \tan x + C$

9. $\int \csc^2 x\, dx = -\cot x + C$

10. $\int \sec x \tan x\, dx = \sec x + C$

11. $\int \csc x \cot x\, dx = -\csc x + C$

12. $\int \tan x\, dx = \ln|\sec x| + C$

13. $\int \cot x\, dx = \ln|\sin x| + C$

14. $\int \sec x\, dx = \ln|\sec x + \tan x| + C$

15. $\int \csc x\, dx = -\ln|\csc x + \cot x| + C$

16. $\int \sinh x\, dx = \cosh x + C$

17. $\int \cosh x\, dx = \sinh x + C$

18. $\int \dfrac{dx}{\sqrt{a^2 - x^2}} = \sin^{-1}\left(\dfrac{x}{a}\right) + C$

19. $\int \dfrac{dx}{a^2 + x^2} = \dfrac{1}{a}\tan^{-1}\left(\dfrac{x}{a}\right) + C$

20. $\int \dfrac{dx}{x\sqrt{x^2 - a^2}} = \dfrac{1}{a}\sec^{-1}\left|\dfrac{x}{a}\right| + C$

21. $\int \dfrac{dx}{\sqrt{a^2 + x^2}} = \sinh^{-1}\left(\dfrac{x}{a}\right) + C$ $(a > 0)$

22. $\int \dfrac{dx}{\sqrt{x^2 - a^2}} = \cosh^{-1}\left(\dfrac{x}{a}\right) + C$ $(x > a > 0)$

11 Integrals over the whole real line

Plenty of applications of integration make sense if we try to integrate over intervals like $[3, \infty)$ or $(-\infty, 7]$. A couple of examples:

Example 12.0.1. If we assume (which is a questionable assumption!) that the current economic situation will continue as it is now, and $R(t)$ represents the annual revenue in year t (measured in dollars per year), then

$$\int_{10}^{100} R(t)\, dt$$

would represent the total revenue collected in years 10 through 100;

$$\int_{10}^{100000} R(t)\, dt$$

would represent the total revenue collected from year 10 to year 100,000. We could ask: *What is the revenue collected from year 10 on?*. That is, we want to integrate $R(t)$ on the interval $[10, \infty)$.

Example 12.0.2. The density $D(h)$ of particles in the atmosphere, as a function of the altitude h above the surface of the earth, decays approximately exponentially: $d(h) = Ce^{-kh}$. If we integrate $D(h)$ over a closed interval like $[0, 10]$, we get the total number of particles between altitudes $h = 0$ and $h = 10$; if we keep increasing the upper limit of the integral, we pick up more and more particles. If we want to know how many particles there are *altogether*, we need to keep letting h get bigger and bigger. The natural symbol to represent this would be

$$\int_{0}^{\infty} Ce^{-kh}\, dh$$

Now, many times the answer to the question "How much revenue will we take in from now until eternity?" is something we'd expect to be "infinitely much". But, as we'll see, this is not always the case.

The situation when integrating out to infinity is similar to the situation with infinite sums. Because there is no already assigned situation with infinite sums. Because there is no already assigned meaning for summing infinitely many things, we defined this as a limit, which in each case needs to be evaluated:

$$\sum_{k=1}^{\infty} a_k := \lim_{M\to\infty} \sum_{k=1}^{M} a_k\,.$$

It is the same when one tries to integrate over the whole real line. We define such integrals by integrating over a bigger and bigger piece and taking the limit. In fact the definition is even pickier than that. We only let one of the limits of integration go to zero at a time. Consider first an integral over a half-line $[a, \infty)$.

12.1 Definitions

We'd better start with a precise definition.

Definition 12.1.1. one-sided integral to infinity. Let a be a real number and let f be a continuous function on the infinite interval $[a, \infty)$. We define

$$\int_a^\infty f(x)\,dx := \lim_{M \to \infty} \int_a^M f(x)\,dx\,. \tag{12.1.1}$$

$(-\infty, b]$

$$\int_{-\infty}^b f(x)\,dx := \lim_{M \to -\infty} \int_M^b f(x)\,dx\,.$$

Checkpoint 158.

$\infty \int_1^\infty dx/x^2 = (-1/x)|_1^\infty = 0 - (1) = 1. -1/(\infty)$

If we want both limits to be infinite then we require the two parts to be defined separately.

Definition 12.1.2. two-sided integral to infinity. Let a be a real number and Let f be a continuous function on the whole real line. Pick a real number c and define

$$\int_{-\infty}^\infty f(x)\,dx := \int_{-\infty}^c f(x)\,dx + \int_c^\infty f(x)\,dx\,. \tag{12.1.2}$$

If either of these two limits is undefined, the whole integral is said not to exist.

Example 12.1.3. What is $\displaystyle\int_{-\infty}^\infty \frac{x}{x^2+1}$? Choosing $c = 0$, we see it is the sum of two one-sided infinite integrals $\int_0^\infty x/(x^2+1)\,dx + \int_{-\infty}^0 x/(x^2+1)\,dx$. Going back to the definition replaces each one-sided infinite integral by a limit:

$$\lim_{M \to \infty} \int_0^M \frac{x}{x^2+1}\,dx + \lim_{M \to \infty} \int_{-M}^0 \frac{x}{x^2+1}\,dx\,.$$

It looks as if this limit is coing to come out to be zero because $x/(x^2+1)$ is an odd function. Integrating from $-M$ to M will produce exactly zero, therefore

$$\lim_{M \to \infty} \int_{-M}^M \frac{x}{x^2+1}\,dx = \lim_{M \to \infty} 0 = 0\,. \tag{12.1.3}$$

Be careful! The definition says not to evaluate (12.1.3) but rather to evaluate the two one-sided integrals separately and sum them. We will come back to finish this example later.

At this point you should be bothered by three questions.

What is c? Does it matter? How do you pick it?

The answer to the first question is, pick c to be anything, you'll always get the same answer. This is important because otherwise, what we wrote isn't really a definition. The reason the integral does not depend on c is that if one changes c from, say, 3 to 4, then the first of the two integrals loses a piece: $\int_3^4 f(x)\,dx$. But the second integral gains this same piece, so the sum is unchanged. This is true even if one or both pieces is infinite. Adding or subtracting the finite quantity $\int_3^4 f(x)\,dx$ won't change that.

If we get $-\infty+\infty$, shouldn't that possibly be something other than "undefined"?

The answer to the second question is yes, sometimes you can be more specific. The one-sided integral to infinity is a limit. Cases where a finite limit does not exist can be resolved into limits of ∞ or $-\infty$, along with the remaining cases where no limit exists even allowing for infinite limits. Because integrals over the whole real line are sums of one-sided (possibly infinite) limits, the rules for infinity from Sections~\ref{ss:variations} and~\ref{ss:LH} can be applied. In other words, integrals over the whole real line are the sum of two one-sided limits; we can add real numbers and $\pm\infty$ according to the rules in Definition 6.0.1: $\infty+\infty=\infty$ (and analogously with $-\infty$), $\infty+a=\infty$ when a is real (and analogously with $-\infty$), $\infty-\infty=UND$, $UND+\text{anything}=UND$, and so on.

Why do we have to split it up in the first place?

The third question is also a matter of definition. The reason we make the choice to do it this way is illustrated by the integral of the sign function

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

On one hand, $\int_{-M}^{M} f(x)\,dx$ is always zero, because the postive and negative parts exactly cancel. On the other hand, $\int_M^\infty f(x)\,dx$ and $\int_{-\infty}^{M} f(x)\,dx$ are always undefined. Do we want the answer for the whole integral $\int_{-\infty}^{\infty} f(x)\,dx$ to be undefined or zero? There is no intrinsically correct choice here but it is a lot safer to have it undefined. If it has a value, one could make a case for values other than zero by centering the integral somewhere else, as in the following exercise.

Checkpoint 159.

Example 12.1.4. The function $\sin(x)/x$ is not defined at $x=0$ but you might recall it does have a limit at 0, namely $\lim_{x\to 0}\sin(x)/x=1$. Therefore the function

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

is a continuous function on the whole real line. Its graph is shown in Figure 12.1.5.

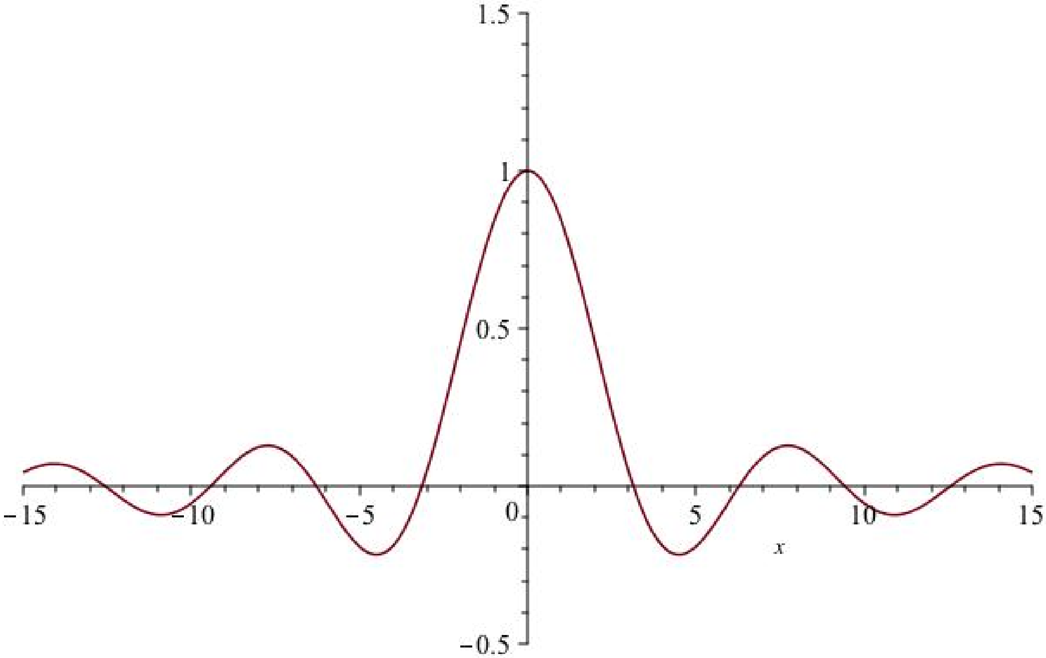


Figure 12.1.5. graph of the function sinc

To write down a limit that defines this integral, we first choose any c. Choosing $c = 0$ makes things symmetric. The integral is then defined as the sum of two integrals, $\int_{-\infty}^{0} \operatorname{sinc}(x)\, dx + \int_0^\infty \operatorname{sinc}(x)\, dx$. Going back to the definition of one-sided integrals as limits, this sum of integrals is equal to

$$\lim_{M\to-\infty} \int_M^0 \operatorname{sinc}(x)\, dx + \lim_{M\to\infty} \int_0^M \operatorname{sinc}(x)\, dx\,.$$

It is not obvious whether these limits exist. One thing is easy to discern: because sinc is an even function, the two limits have the same value (whether finite or not). We can safely say:

$$\int_{-\infty}^{\infty} \operatorname{sinc}(x)\, dx = 2 \cdot \lim_{M\to\infty} \int_0^M \operatorname{sinc}(x)\, dx\,.$$

Checkpoint 160.

12.2 Convergence

The central question of this section is: how do we tell whether a limit such as $\int_b^\infty f(x)\, dx$ exists? If the limit exists, we would like to evaluate it if possible, or estimate it otherwise. When discussing convergence you should realize that $\int_a^\infty f(x)\, dx$ either diverges for all values of a or converges for all values of a as long as f is defined and continuous on $[a, \infty)$. For this reason, we use the notation $\int^\infty f(x)\, dx$ or, to be really blunt, $\displaystyle\int_{\text{whocares}}^{\infty} f(x)\, dx$.

Checkpoint 161.

Case 1: you know how to compute the definite integral. Suppose $\int_b^M f(x)\, dx$ is something for which you know how to compute an explicit formula. The formula will have M in it. You have to evaluate the limit as $M \to \infty$. How do you do that? There is no one way, but that's why we studied limits before. Apply what you know. What about b, do you have to take a limit in b as well? I hope you already knew the answer to that. In this definition, b is any fixed number. You don't take a limit.

These special cases will become theorems once you have worked them out.

Table 12.2.1. Convergence tests. Fill out this table to use as a reference.

Name of test	Type of integral	Condition for convergence
	$\int_b^\infty e^{kx}\,dx$	
power test	$\int_b^\infty x^p\,dx$	
	$\int_b^\infty \frac{(\ln x)^q}{x}\,dx$	

You will work out these cases in class: write each as a limit, evaluate the limit, state whether it converges, which will depend on the value of the parameter, k, p or q. Go ahead and pencil them in once you've done this. The second of these especially, is worth remembering because it is not obvious until you do the computation where the break should be between convergence and not.

Checkpoint 162. n.edu/~ancoop/1070/ss-improper.html#project-162)

Case 2: you don't know how to compute the integral. In this case you can't even get to the point of having a difficult limit to evaluate. So probably you can't evaluate the improper integral. But you can and should still try to answer whether the integral has a finite value versus being undefined. This is where comparison tests come in. You build up a library of cases where you do know the answer and then, for the rest of functions, you try to compare them to functions in your library.

Sometimes a comparison is informative, sometimes it isn't. Suppose that f and g are positive functions and $f(x) \le g(x)$ for all x. Consider several pieces of information you might have about these functions.

Table 12.2.2. Comparison tests. Fill out this table as a reference.

	what you know	conclusion
(a)	$\int_b^\infty f(x)\,dx$ converges to a finite value L.	
(b)	$\int_b^\infty f(x)\,dx$ does not converge.	
(c)	$\int_b^\infty g(x)\,dx$ converges to a finite value L.	
(d)	$\int_b^\infty g(x)\,dx$ does not converge.	

In which cases can you conclude something about the other function? We are doing this in class. Once you have the answer, either by working it out yourself or from the class discussion, please pencil it in here so you'll have it for later reference.

Checkpoint 163. n.edu/~ancoop/1070/ss-improper.html#project-163)

Asymptotic comparison tests. Here are two key ideas that help your comparison tests work more of the time, based on the fact that the question "convergent or not?" is not sensitive to certain differences between integrands.

Multiplying by a constant does not change whether an integral converges.

That's because if $\lim_{M\to\infty}\int_b^M f(x)\,dx$ converges to the finite constant L then $\lim_{M\to\infty}\int_b^M Kf(x)\,dx$ converges to the finite constant KL.

Checkpoint 164. n.edu/~ancoop/1070/ss-improper.html#project-164)

It doesn't matter if $f(x) \le g(x)$ for every single x as long as the inequality is true for sufficiently large x. If $f(x) \le g(x)$ once $x \ge 100$, then you can apply the comparison test to compare $\int_b^\infty f(x)\,dx$ to $\int_b^\infty g(x)\,dx$ as long as $b \ge 100$. But even if not, once you compare $\int_{100}^\infty f(x)\,dx$ to $\int_{100}^\infty g(x)\,dx$, then adding the finite quantity $\int_b^{100} f(x)\,dx$ or $\int_b^{100} g(s)\,dx$ will not change whether either of these converges.

Putting these two ideas together leads to the conclusion that if $f(x) \le Kg(x)$ from some point onward and $\int_b^\infty g(x)\,dx$ converges, then so does $\int_b^\infty f(x)\,dx$. The theorem we just proved is:

Theorem 12.2.3. asymptotic comparison. If f and g are positive functions on some interval (b,∞) and if there are some constants M and K such that

$$f(x) \le Kg(x) \text{ for all } x \ge M \tag{12.2.1}$$

then convergence of the integral $\int_b^\infty g(x)\,dx$ implies convergence of the integral $\int_b^\infty f(x)\,dx$.

In particular, if $f(x) \ll g(x)$ as $x\to\infty$ then \mathbb{E}qref{eq:O} holds, hence convergence of the integral $\int_b^\infty g(x)\,dx$ implies convergence of the integral $\int_b^\infty f(x)\,dx$.

Checkpoint 165. n.edu/~ancoop/1070/ss-improper.html#ex-asym)

Example 12.2.4. power times negative exponential. Does $\int_1^\infty x^8 e^{-x}\,dx$ converge? One way to do this is by computing the integral exactly. This takes eight integrations by parts, and is probably too messy unless you figured out how to do "tabular" integration (optional when you learned integration by parts). In any case, there's an easier way if you only want to know whether it converges, but not to what.

We claim that $x^8e^{-x} \ll e^{-(1/2)x}$ (you could use $e^{-\beta x}$ in this argument for any $\beta\in(0,1)$). It follows from the asymptotic comparison test that convergence of $\int_1^\infty e^{-(1/2)x}$ implies convergence of $\int x^8e^{-x}\,dx$. We check the claim by evaluating

$$\lim_{x\to\infty}\frac{x^8e^{-x}}{e^{-(1/2)x}} = \lim_{x\to\infty}\frac{x^8}{e^{(1/2)x}} = 0$$

because we know the power x^8 is much less than the exponential $e^{(1/2)x}$.

Checkpoint 166. n.edu/~ancoop/1070/ss-improper.html#project-166)

A particular case of Theorem 12.2.3 is when $f(x)\sim g(x)$. When two functions are asymptotically equivalent, then each can be upper bounded by a constant multiple of the other, hence we have the following proposition.

Proposition 12.2.5. If f and g are positive functions and $f \sim g$ then $\int^{\infty} f(x)\,dx$ converges if and only if $\int^{\infty} g(x)\,dx$ converges.

Example 12.2.6. n.edu/~ancoop/1070/ss-improper.html#example-72)

Does $\displaystyle\int_1^{\infty} \frac{dx}{x^2+3x}$ converge?

We can use comparison test (c) here: $\dfrac{1}{x_2+3x} \le \dfrac{1}{x^2}$ and we know $\displaystyle\int_1^{\infty} \frac{dx}{x^2}$ converges, hence so does $\displaystyle\int_1^{\infty} \frac{dx}{x^2+3x}$.

Does $\displaystyle\int_4^{\infty} \frac{dx}{x^2-3x}$ converge?

Now the inequatlity goes the other way, so we are in case (c) of the comparison test and we cannot conclude anything from direct comparison. However, we also know $\dfrac{1}{x^2-3x} \sim \dfrac{1}{x^2}$ as $x \to \infty$, therefore we can conclude convergence again by Proposition 12.2.5.

Did you wonder about the lower limit of 4 in the second part? That wasn't just randomly added so you'd be more flexible about the lower limits of integrals to infinity. It was put there to ensure that f was continuous; note the discontinuity at $x = 3$.

Checkpoint 167. n.edu/~ancoop/1070/ss-improper.html#project-167)

12 Integration and Probability

Maybe you've taken a courses in probability. Maybe you saw a little probability theory in high school. Maybe you've never studied anything to do with probability.

The first thing students usually learn is ***discrete probability***, where the random variables take values in a finite set, with given probabilities for each outcome. That's because this can be studied with middle school mathematics. For example, rolling two 6-sided dice leads to 36 possible outcomes, each equally likely; this in turn leads to 11 possible outcomes for the sum of the two dice, with probabilities ranging from $1/36$ for 2 and 12 to $6/36$ for 7. All questions about rolls of finitely many dice can be answered with careful thinking and basic arithmetic.

Random variables whose values are spread over all real numbers, or a real interval, require calculus even to define (much less to study). These are called ***continuous*** random variables, and are the topic of this section.

Philosophically, a real-valued random variable X is a quantity that has a value equal to some real number, but will have a different value each time some kind of experiment is run. It is unpredictable, therefore we cannot answer the question "What is the value of X?" but only "What is the probability that the value of X lies in the set A?"

For example, suppose we throw a dart at a 12 foot wide wall, from a long enough distance and with poor enough aim that it is as likely to hit any region as any other (if we miss completely, we get another try). Say the random variable X is the distance (in feet) from the left edge of the wall. We can ask for the probability that $X \leq 2$, that is that the dart lands within two feet of the left edge.

Checkpoint 168.

13.1 Probability densities

For discrete random variables you answer this type of question by summing the probability that X is equal to y for every y in the set A. For continuous random variables, the probability of being equal to any one real number is zero. In the example with the dart, the probability that it lands exactly $\sqrt{3}$ feet from the left edge (or 1 foot, or $1/3$ of a foot, or any other real number of feet) is zero. The only way to get a nonzero probability is to consider an entire interval of values. Thus the most basic questions we ask about X are: what is the probability that $X \in [a, b]$, where $a < b$ are fixed real numbers. These probabilities will be governed by a ***probability density***, which is a nonnegative function telling how likely it is for X to be in an interval centered at any given real number.

De nition 13.1.1. probabilitiy densities.

probability density A probability density is a nonnegative function f such that $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

random variable A random variable X is said to have probability density f if the probability of finding X in any interval $[a,b]$ is equal to $\int_a^b f(t)\,dt$.

probability We denote the probability of finding X in $[a,b]$ by $\mathbb{P}(X \in [a,b])$.

Checkpoint 169.

Sometimes f is defined only on an interval $[a,b]$ and not on the whole real line. The interpretation is that the random variable X takes values only in $[a,b]$. Probabilities for X are then defined by integrating in sub-intervals of $[a,b]$. Often one extends the definition of f to all real numbers by making it zero off of $[a,b]$. This may result in f being discontinuous but its definite integrals are still defined.

Example 13.1.2. The standard exponential random variable has density e^{-x} on $[0,\infty)$. If X has this density, what is $\mathbb{P}(X \in [-1,1])$? This is the same as $\mathbb{P}(X \in [0,1])$, because by assumption X cannot be negative. We compute it by $\displaystyle\int_0^1 e^{-x}\,dx = e^{-x}\Big|_0^1 = 1 - e^{-1}$. As a quick reality check we observe that the quantity $1 - \frac{1}{e}$ is indeed between zero and one, therefore it makes sense for this to be a probability.

Checkpoint 170.

Often the model dictates the form of the function f but not a multiplicative constant.

Example 13.1.3. For example, if we know that $f(x)$ should be of the form Cx^{-3} on $[1,\infty)$ then we would need to find the right constant C to make this a probability density. The function f has to integrate to 1, meaning we have to solve

$$\int_1^{\infty} Cx^{-3}\,dx = 1$$

for C. Solving this results in $C = 2$, therefore the density of f is $2/x^3$ on $[1,\infty)$.

Checkpoint 171.

13.2 Summary statistics

Several important quantities associated with a probability distribution are the mean, the variance, the standard deviation and the median. We call these summary statistics because they tell us about the probability density *as a whole*. Again, a couple of paragraphs don't do justice to these ideas, but we hope they explain the concepts at least a little and make the math seem more motivated and relevant.

Probably the intuitively-simplest summary statistic is the median. This is the 50^{th} percentile of the distribution -- the value of the measured variable which splits the distribution into two equal pieces.

Definition 13.2.1. The ***median*** of a random variable X having probability density f is the real number m such that

$$\mathbb{P}(X > m) = \mathbb{P}(X < m) = \frac{1}{2}\,. \tag{13.2.1}$$

Checkpoint 172. n.edu/~ancoop/1070/ss-prob.html#project-172)

Definition 13.2.2. n.edu/~ancoop/1070/ss-prob.html#definition-30)

mean
: If X has probability density f, the ***mean*** or ***expectation*** of X (the two terms are synonyms) is the quantity

$$\mathbb{E}X := \int_{-\infty}^{\infty} x\, f(x)\, dx. \tag{13.2.2}$$

A variable commonly used for the mean of a distribution is μ.

variance
: If X has probability density f and mean μ, the ***variance*** of X is the quantity

$$\mathrm{Var}(X) := \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, dx\,.$$

standard deviation
: The ***standard deviation*** of X is just the square root of the variance

$$\sigma := \sqrt{\mathrm{Var}(X)}\,.$$

Checkpoint 173. n.edu/~ancoop/1070/ss-prob.html#project-173)

Each of these definitions looks like the corresponding formula from discrete probability. By way of example, consider the formula for expectation. You might recall what happens when rolling a die. Each of the six numbers comes up about $1/6$ of the time, so in a large number N of dice rolls you will get (approximately) $N/6$ of each of the six outcomes. The average will therefore be (approximately)

$$\frac{1}{N}\left[(N/6)\cdot 1 + (N/6)\cdot 2 + (N/6)\cdot 3 + (N/6)\cdot 4 + (N/6)\cdot 5 + (N/6)\cdot 6\right].$$

We can write this in summation notation as

$$\sum_{j=1}^{6} j \cdot \mathbb{P}(X = j)\,.$$

If we think of this sum as a Riemann sum for something, and do our Greek-to-Latin trick, we get back the formula (13.2.2).

Checkpoint 174. n.edu/~ancoop/1070/ss-prob.html#project-174)

When instead there are infinitely many possible outcomes spread over an interval, the sum is replaced by an integral

$$\int_{-\infty}^{\infty} x \cdot f(x)\, dx\,. \qquad (13.2.3)$$

A famous theorem in probability theory, called the Strong Law of Large Numbers, says that the formula (13.2.3) still computes the long term average: the long term average of independent draws from a distribution with probability density function f will converge to $\int x \cdot f(x)\, dx$.

Checkpoint 175.

It is more difficult to understand why the variance has the precise definition it does, but it is easy to see that the formula produces bigger values when the random variable X tends to be farther from its mean value μ. The standard deviation is another measure of dispersion. To see why it might be more physically relevant, consider the units.

Probabilities such as $\mathbb{P}(X \in [a,b])$ can be considered to be unitless because they represent ratios of like things: frequency of occurrences within the interval $[a,b]$ divided by frequency of all occurrences. Probability densities, integrated against the variable x (which may have units of length, time, etc.) give probabilities. Therefore, probability densities have units of "probability per unit x-value", or in other words, inverse units to the independent variable.

The units of the mean are units of $\int x f(x)\, dx$, which is units of f times x^2; but f has units of inverse x, so the mean has units of x. This makes sense because the mean represents a point on the x-axis. Similarly, the variance has units of x^2. It is hard to see what the variance represents physically. The standard deviation, however, has units of x. Therefore, it is a measure of dispersion having the same units as the mean. It represents a distance on the x-axis which is some kind of average discrepancy from the mean.

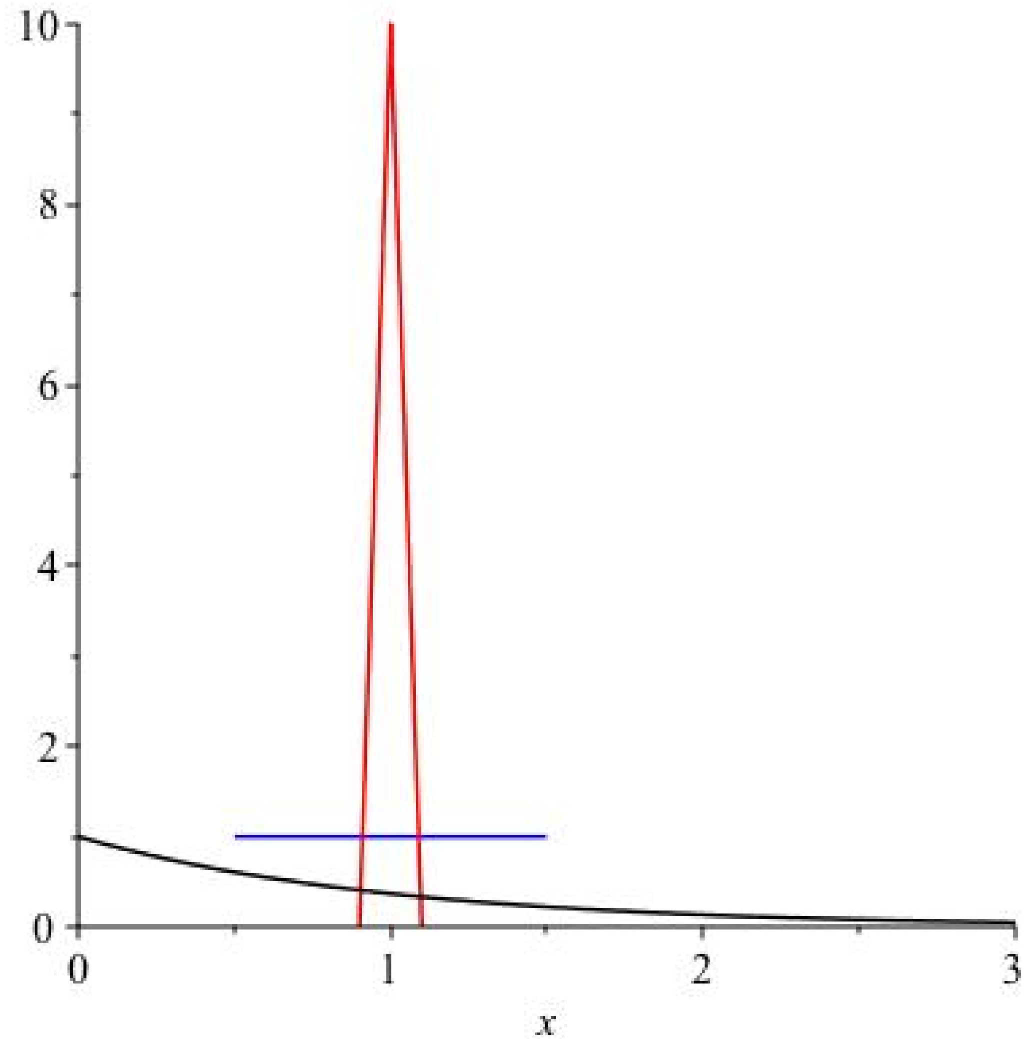


Figure 13.2.3. Three densities with mean 1.

Checkpoint 176.

13.3 Some common probability densities

There are a zillion different functions commonly used for probability densities. Three of the most common are named in this section: the exponential, the uniform, and the normal. These are common in probability for reasons analogous to why exponential behavior is common in evolving systems. Each of these is constructed to have a particular useful property.

The uniform distribution, as the name applies, arises when a random quantity is uniformly likely to be anywhere in an interval. It is often used as an "uninformed" model when all you know is that a quantity has to be somewhere in a fixed interval. The normal arises when many small independent contributions are summed. It is often used to model observational error. The exponential is the so-called memoryless distribution. It arises when the probability of finding X in the next small interval, given that you haven't already found it, is always constant.

All three of these are parametrized families of distributions. Once values are picked for the parameters you get a particular distribution. This section concludes by giving definitions of eac h and discuss typical applications.

The exponential distribution. The exponential distribution has a parameter μ which can be any positive real number. Its density is $(1/\mu)e^{-x/\mu}$ on the positive half-line $[0, \infty)$. This is obviously the same as the density Ce^{-Cx} (just take $C = 1/\mu$) but we use the parameter μ rather than C because a quick computation shows that μ is a more natural property of the distribution.

Checkpoint 177. (https://www2.math.upenn.edu/~ancoop/1070/ss-prob.html#project-177)

The exponential distribution has a very important "memoryless" property. If X has an exponential density with any parameter and is interpreted as a waiting time, then once you know it didn't happen by a certain time t, the amount of further time it will take to happen has the same distribution as X had originally. It doesn't get any more or any less likely to happen in the the interval $[t, t+1]$ than it was originally to happen in the interval $[0, 1]$.

The median of the exponential distribution with mean μ is also easy to compute. Solving $\int_0^M \mu^{-1} e^{-x/\mu}\, dx = 1/2$ gives $M = \mu \cdot \ln 2$. When X is a random waiting time, the interpretation is that it is equally likely to occur before $\ln 2$ times its mean as after. Because $\ln 2 \approx 0.7$, the median is significantly less than the mean. When modeling with exponentials, it is good to remember it produces values that are unbounded but always positive.

Any of you who have studied radioactive decay know that each atom acts randomly and independently of the others, decaying at a random time with an exponential distribution. The fraction remaining after time t is the same as the probability that each individual remains undecayed at time t, namely $e^{-t/\mu}$, so another interpretation for the median is the ***half-life***: the time at which only half the original substance remains. Other examples are the life span of an organism that faces environmental hazards but does not age, or time for an electronic component to fail (they don't seem to age either).

The uniform distribution. The uniform distribution on the interval $[a, b]$ is the probability density whose density is a constant on this interval: the constant will be $1/(b-a)$. This is often thought of the least informative distribution if you know that the the quantity must be between the values a and b. The mean and median are both $(a+b)/2$.

Checkpoint 178. (https://www2.math.upenn.edu/~ancoop/1070/ss-prob.html#project-178)

Example 13.3.1. In your orienteering class you are taken to a far away location and spun around blindfolded when you arrive. When the blindfold comes off, you are facing at a random compass angle (usually measured clockwise from due north). It would be reasonable to model this as a uniform random variable from the interval $[0, 360]$ in units of degrees.

Checkpoint 179. (https://www2.math.upenn.edu/~ancoop/1070/ss-prob.html#project-179)

The normal distribution. The normal density with mean μ and standard deviation σ is the density (https://www2.math.upenn.edu/~ancoop/1070/ss-prob.html#p-1637)

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}\,.$$

The ***standard normal*** is the one with $\mu = 0$ and $\sigma = 1$. There is a very cool mathematical reason for this formula, which we will not go into. When a random variable is the result of summing a bunch of smaller random variables all acting independently, the result is usually well approximated by a normal. It is possible (though very tricky) to show that the definite integral of this density over the whole real line is in fact 1 (in other words, that we have chosen the right constant to make it a probability density).

Annoyingly, there is no nice antiderivative, so no way in general of computing the probability of finding a normal between specified values a and b. Because the normal is so important in statistical applications, they made up a notation for the indefinite integral in the case $\mu = 0, \sigma = 1$, using the capital Greek letter Φ ("phi", pronounced "fee" or "fie"):

$$\Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx \, .$$

So now you can say that the probability of finding a standard normal between a and b is exactly $\Phi(b) - \Phi(a)$. In the old, pre-computer days, they published tables of values of Φ. It was reasonably efficient to do this because you can get the antiderivative F of any other normal from the one for the standard normal by a linear substition: $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$.

13 Taylor approximations

We started the course (back in Preface (sec-zero.html)) with the idea of approximating a function $f(x)$ by its tangent line. The problem of actually finding the tangent line is what motivated us to think hard about limits and how they work, then to apply that idea to secant lines.

Here's another definition of the tangent line (in addition to the ones we saw in Subsection 4.2 (section-4.html#ss-def-tangent-line)):

Definition 14.0.1. alternative definition of the tangent line. The tangent line to the graph of $f(x)$ at $x = c$ is the graph of the linear function $L(x) = mx + b$ which is as good an approximation to $f(x)$ nearby to $x = c$ as any linear function can be.

This definition might make you a bit uncomfortable. How would you use this definition to actually find a tangent line?

We'll use the notion of asymptotic equivalence from Subsection 6.3 (section-6.html#ss-asyeq). What we'd like, first of all, is that $f(x) \sim L(x)$ as $x \to c$. In other words, we want

$$\lim_{x\to c} \frac{f(x)}{L(x)} = 1 \ .$$

Now as long as $f(x)$ and $L(x)$ are both continuous, and $L(c) \neq 0$, the rules of limits tell us that $\lim_{x\to c} \frac{f(x)}{L(x)} = \frac{f(c)}{L(c)}$. So we need $\frac{f(c)}{L(c)} = 1$, which is just the requirement that $L(c) = f(c)$. For reference later, we'll say that the value $f(c)$ *approximates the function $f(x)$ to zeroth order.*

A line is determined by two pieces of information:

- a slope m and an intercept b
- a slope m and a point (x_0, y_0) on the line
- two points (x_0, y_0) and (x_1, y_1) on the line

We just learned that the point $(c, f(c))$ has to be on the line we're looking for, so let's work with the second option and determine L by finding its slope.

Imagine a competition among all lines which pass through the point $(c, f(c))$, for which one is "most like" $f(x)$ nearby $x = c$. Using the point-slope form of the equation for a line, we know that any such competitor has to look like $L(x) = f(c) + m(x - c)$. Now:

We want $f(x) \sim f(c) + m(x - c)$ as $x \to c$
That's the same as demanding $f(x) - f(c) \sim m(x - c)$ as $x \to c$

All $f(x) - f(c) \sim m(x-c)$ as $x \to c$ means is that

$$\lim_{x \to c} \frac{f(x) - f(c)}{m(x-c)} = 1$$

which can only happen if $m = \lim_{x \to c} \frac{f(x)-f(c)}{x-c} = f'(c)$.

In other words, this approach (finding an asymptotically-equivalent line) gives the same formula for the slope of the tangent line that we know and love.

You might ask: why all the work just to get the same formula? Because the understanding of the tangent line as "the line that matches $f(x)$ near $x = c$ as well as any line can" will generalize to questions like:

- What parabola matches the graph of $f(x)$ near $x = c$ as well as any parabola can?
- What cubic function approximates $f(x)$ near $x = c$ the best among all cubic functions?
- etc.

These polynomial approximations are known as the *Taylor polynomials* for $f(x)$. They allow you to use your knowledge of polynomials -- which you've no doubt spent many years honing -- to understand how nonpolynomial functions behave.

14.1 Taylor polynomials

Start with a function $f(x)$. Our goal will be to figure out the best approximation of $f(x)$ (near $x = c$) by a polynomial of whatever degree we want (call that degree n).

$n = 1$. When $n = 1$, we want a degree-1 polynomial approximation. That is, we want the tangent line. So we use

$$T_1(x) = f(c) + f'(c)(x-c) \quad .$$

$n = 2$. A degree-2 polynomial looks like $q(x) = Ax^2 + Bx + C$. Since we're interested in what happens when $x \to c$, let's rewrite this form as $q(x) = a_2(x-c)^2 + a_1(x-c) + a_0$. (The exact relationship between A, B, C and a_0, a_1, a_2 isn't important -- we just want to make sure that we're writing everything in terms of c.)

Let's again write down our demands. We want $f(x) \sim q(x)$ as $x \to c$, i.e.,

$$\lim_{x \to c} \frac{f(x)}{q(x)} = 1.$$

As $x \to c$, $x - c \to 0$. So $\lim_{x \to c} q(x) = a_0$. As long as f is continuous, $\lim_{x \to c} f(x) = f(c)$. So we need $a_0 = f(c)$.

As we did in the linear case, let's imagine a competition among all quadratic polynomials of the form $q(x) = a_2(x-c)^2 + a_1(x-c) + f(c)$. We demand:

$$\begin{aligned} f(x) &\sim q(x) \text{ as } x \to c \\ f(x) &\sim a_2(x-c)^2 + a_1(x-c) + f(c) \text{ as } x \to c \\ f(x) - f(c) &\sim a_2(x-c)^2 + a_1(x-c) \text{ as } x \to c \end{aligned}$$

But $f(x) - f(c) \sim a_2(x-c)^2 + a_1(x-c)$ as $x \to c$ just means

$$\lim_{x\to c}\frac{f(x)-f(c)}{a_2(x-c)^2+a_1(x-c)}=1 \quad .$$

We can rearrange the fraction to:

$$\lim_{x\to c}\frac{f(x)-f(c)}{x-c}\frac{1}{a_2(x-c)+a_1}=1 \quad .$$

As $x\to c$, $\frac{1}{a_2(x-c)+a_1}\to\frac{1}{a_1}$. We end up with $a_1=f'(c)$.

So now we know that $q(x)=a_2(x-c)^2+f'(c)(x-c)+f(c)$. Let's rewrite our demands again:

$$\begin{aligned} f(x) &\sim q(x) \text{ as } x\to c \\ f(x) &\sim a_2(x-c)^2+f'(c)(x-c)+f(c) \text{ as } x\to c \\ f(x)-f(c)-f'(c)(x-c) &\sim a_2(x-c)^2 \text{ as } x\to c \end{aligned}$$

This means we need

$$\lim_{x\to c}\frac{f(x)-f(c)-f'(c)(x-c)}{a_2(x-c)^2}=1 \quad .$$

The form of this limit is $\frac{0}{0}$, so we can apply L'Hôpital's Rule. We need to be a bit careful to differentate *with respect to* x. The numerator's derivative is

$$\frac{d}{dx}\big(f(x)-f(c)-f'(c)(x-c)\big)=f'(x)-f'(c)$$

and the denomator's is

$$\frac{d}{dx}\big(a_2(x-c)^2\big)=2a_2(x-c) \quad .$$

Applying L'Hôpital's Rule, we get

$$1=\lim_{x\to c}\frac{f(x)-f(c)-f'(c)(x-c)}{a_2(x-c)^2}=\lim_{x\to c}\frac{f'(x)-f'(c)}{2a_2(x-c)}=\lim_{x\to c}\frac{1}{2a_2}\frac{f'(x)-f'(c)}{x-c} \quad .$$

So $a_2=\frac{1}{2}f''(c)$.

That is, the quadratic which best approximates $f(x)$ near $x=c$ is

$$T_2(x)=\frac{1}{2}f''(c)(x-c)^2+f'(c)(x-c)+f(c) \quad .$$

$n=3$**.** Now let's approximate $f(x)$ by a cubic of the form $q(x)=a_3(x-c)^3+a_2(x-c)^2+a_1(x-c)+a_0$. It probably won't surprise you that $a_0=f(c)$. Let's see about a_1:

$$\begin{aligned} f(x) &\sim q(x) \text{ as } x\to c \\ f(x) &\sim a_3(x-c)^3+a_2(x-c)^2+a_1(x-c)+f(c) \text{ as } x\to c \\ f(x)-f(c) &\sim a_3(x-c)^3+a_2(x-c)^2+a_1(x-c) \text{ as } x\to c \end{aligned}$$

which just means

$$\lim_{x\to c}\frac{f(x)-f(c)}{a_3(x-c)^3+a_2(x-c)^2+a_1(x-c)}=1 \quad ,$$

or in other words

$$\lim_{x \to c} \frac{f(x) - f(c)}{x - c} \frac{1}{a_3(x-c)^2 + a_2(x-c)^+ a_1} = 1 \quad .$$

Because $\frac{1}{a_3(x-c)^2+a_2(x-c)^+a_1} \to \frac{1}{a_1}$ as $x \to c$, we see that $a_1 = f'(c)$.

So now we have

$$\begin{aligned} f(x) &\sim q(x) \text{ as } x \to c \\ f(x) &\sim a_3(x-c)^3 + a_2(x-c)^2 + f'(c)(x-c) + f(c) \text{ as } x \to c \\ f(x) - f(c) - f'(c)(x-c) &\sim a_3(x-c)^3 + a_2(x-c)^2 \text{ as } x \to c \end{aligned}$$

In order words, we need

$$\lim_{x \to c} \frac{f(x) - f(c) - f'(c)(x-c)}{a_3(x-c)^3 + a_2(x-c)^2} = 1 \quad .$$

L'Hôpital tells us that

$$1 = \lim_{x \to c} \frac{f(x) - f(c) - f'(c)(x-c)}{a_3(x-c)^3 + a_2(x-c)^2} = \lim_{x \to c} \frac{f'(x) - f'(c)}{3a_3(x-c)^2 + 2a_2(x-c)} = \lim_{x \to c} \frac{f'(x) - f'(c)}{x - c} \frac{1}{3a_3(x-c) + 2a_2} = \frac{f''(c)}{2a_2} \quad .$$

So $a_2 = \frac{1}{2} f''(c)$.

Let's update our demand:

$$\begin{aligned} f(x) &\sim q(x) \text{ as } x \to c \\ f(x) &\sim a_3(x-c)^3 + \frac{1}{2} f''(c)(x-c)^2 + f'(c)(x-c) + f(c) \text{ as } x \to c \\ f(x) - f(c) - f'(c)(x-c) - \frac{1}{2} f''(c)(x-c)^2 &\sim a_3(x-c)^3 \text{ as } x \to c \end{aligned}$$

So we consider the limit

$$\lim_{x \to c} \frac{f(x) - f(c) - f'(c)(x-c) - \frac{1}{2} f''(c)(x-c)^2}{a_3(x-c)^3} = 1 \quad .$$

Again L'Hôpital's Rule applies here, so we differentiate the numerator and the denominator to get:

$$1 = \lim_{x \to c} \frac{f'(x) - f'(c) - f''(c)(x-c)}{3a_3(x-c)^2} \quad .$$

This form is still indeterminate, so we apply L'Hôpital's Rule again:

$$1 = \lim_{x \to c} \frac{f''(x) - f''(c)}{6a_3(x-c)} = \lim_{x \to c} \frac{1}{6a_3} \frac{f''(x) - f''(c)}{x - c} \quad .$$

Thus, we need to use $a_3 = \frac{1}{6} f'''(c)$. Our approximating cubic is therefore:

$$T_3(x) = \frac{1}{6} f'''(c)(x-c)^3 + \frac{1}{2} f''(c)(x-c)^2 + f'(c)(x-c) + f(c) \quad .$$

Let's write these out so that we can see the pattern.

$$\begin{aligned} T_1(x) &= f(c) + f'(c)(x-c) \\ T_2(x) &= f(c) + f'(c)(x-c) + \frac{1}{2} f''(c)(x-c)^2 \\ T_3(x) &= f(c) + f'(c)(x-c) + \frac{1}{2} f''(c)(x-c)^2 + \frac{1}{6} f'''(c)(x-c)^3 \end{aligned}$$

Can you see the pattern?

Checkpoint 180.

Definition 14.1.1. Taylor polynomials. The degree-n Taylor polynomial for $f(x)$ at $x=c$ is

$$T_n(x) = f(c) + f'(c)(x-c) + \frac{1}{2}f''(c)(x-c)^2 + \cdots + \frac{1}{n!}f^{(n)}(c)(x-c)^n$$

A few observations to help you remember this formula:

- The variable here is x -- not c.
- In the formula for T_n, the derivatives of f are all evaluated *at* $x=c$.
- In each term, there is a derivative, a power of $x-c$, and a factorial. These all share the same index.
- $T_3(x)$ is just $T_2(x)$ plus a term of degree 3; $T_4(x)$ is just $T_3(x)$ plus a term of degree 4; etc.

Example 14.1.2. *Compute the degree-4 Taylor polynomial for* $f(x) = \sin(x) + \cos(x)$ *at* $x = \frac{\pi}{6}$.

First, we'll need to compute some derivatives:

$$\begin{aligned}
f(x) &= \sin(x) + \cos(x) & f\left(\frac{\pi}{6}\right) &= \frac{1}{2} + \frac{\sqrt{3}}{2} = \frac{1+\sqrt{3}}{2}\\
f'(x) &= \cos(x) - \sin(x) & f'\left(\frac{\pi}{6}\right) &= \frac{\sqrt{3}-1}{2}\\
f''(x) &= -\sin(x) - \cos(x) & f''\left(\frac{\pi}{6}\right) &= \frac{-1-\sqrt{3}}{2}\\
f'''(x) &= -\cos(x) + \sin(x) & f'''\left(\frac{\pi}{6}\right) &= \frac{-\sqrt{3}+1}{2}\\
f^{(4)}(x) &= \sin(x) + \cos(x) & f^{(4)}\left(\frac{\pi}{6}\right) &= \frac{1+\sqrt{3}}{2}
\end{aligned}$$

The formula in Definition 14.1.1 then yields:

$$T_4(x) = \frac{1+\sqrt{3}}{2} + \frac{\sqrt{3}-1}{2}(x-c) + \frac{1}{2}\frac{-1-\sqrt{3}}{2}(x-c)^2 + \frac{1}{6}\frac{-\sqrt{3}+1}{2}(x-c)^3 + \frac{1}{24}\frac{1+\sqrt{3}}{2}(x-c)^4$$

which we could simplify (if we wanted to):

$$T_4(x) = \frac{1+\sqrt{3}}{2} + \frac{\sqrt{3}-1}{2}(x-c) + \frac{-1-\sqrt{3}}{4}(x-c)^2 + \frac{-\sqrt{3}+1}{12}(x-c)^3 + \frac{1+\sqrt{3}}{48}(x-c)^4$$

Now you try your hand at it:

Checkpoint 181.

Here's another important property of the Taylor polynomials.

Checkpoint 182.

This exercise shows two things:

- where the factorial comes from in each term of the Taylor polynomial
- that the derivatives of the Taylor polynomial are the same as the derivatives of f, at least at $x = c$

Definition 14.1.3. Maclaurin polynomials. When we use $c = 0$, we call the result the *Maclaurin* polynomials.

Example 14.1.4. Let's compute the degree-5 Maclaurin polynomial for $f(x) = e^x$:

$$\begin{aligned} f(x) &= e^x & f(0) &= 1 \\ f'(x) &= e^x & f'(0) &= 1 \\ f''(x) &= e^x & f''(0) &= 1 \\ f'''(x) &= e^x & f'''(0) &= 1 \\ f^{(4)}(x) &= e^x & f^{(4)}(0) &= 1 \\ f^{(5)}(x) &= e^x & f^{(5)}(0) &= 1 \end{aligned}$$

So the Maclaurin polynomial is

$$1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + \frac{1}{120}x^5 \quad .$$

14.2 Taylor and Maclaurin polynomials in graphing

We claimed that the Taylor polynomials are polynomials that "act like" the function $f(x)$ near $x = c$. What does that mean in terms of the graph of $f(x)$ and its Taylor polynomials?

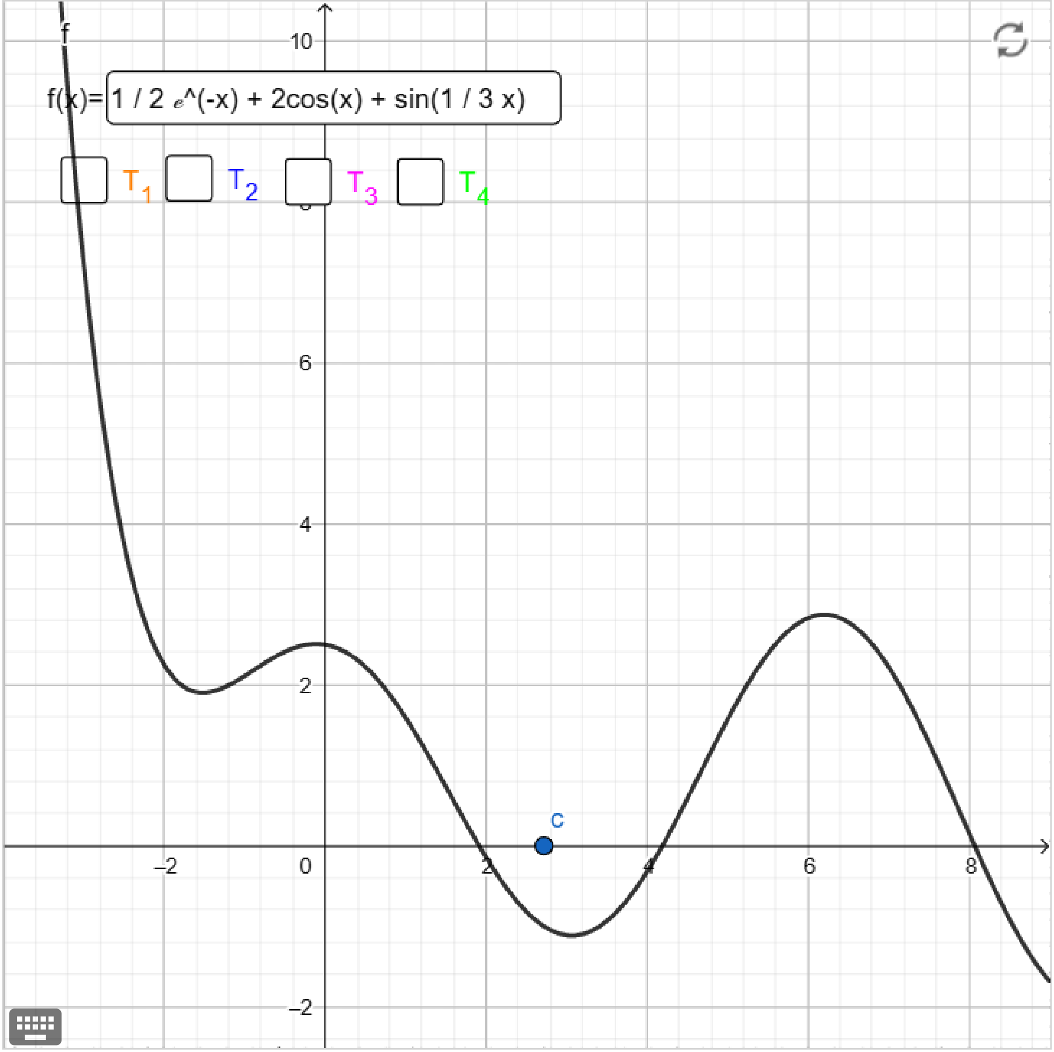


Figure 14.2.1. Use the checkboxes to display the degrees 1, 2, 3, and 4 Taylor approximations to $f(x)$. You can enter your favorite function $f(x)$, and drag the point $x = c$ around. Zoom in and out to see how good the approximation is at different scales.

The fact that the graphs of the Taylor polynomials are similar to those of the original function f gives us a way to explain why the techniques mentioned in Subsection 4.3 (section-4.html#ss-derivs-graphing) work. Let's imagine the situation of a function $f(x)$ with $f(1) = 1$, $f'(1) = -\frac{1}{2}$, and $f''(1) = 3$. This is enough information to compute that the degree-2 Taylor polynomial at $x = 1$ is

$$T_2(x) = 1 - \frac{1}{2}(x-1) + \frac{3}{2}(x-1)^2 \quad .$$

Since the graph of T_2 is close to the graph of f, we can infer properties of f from those of T_2.

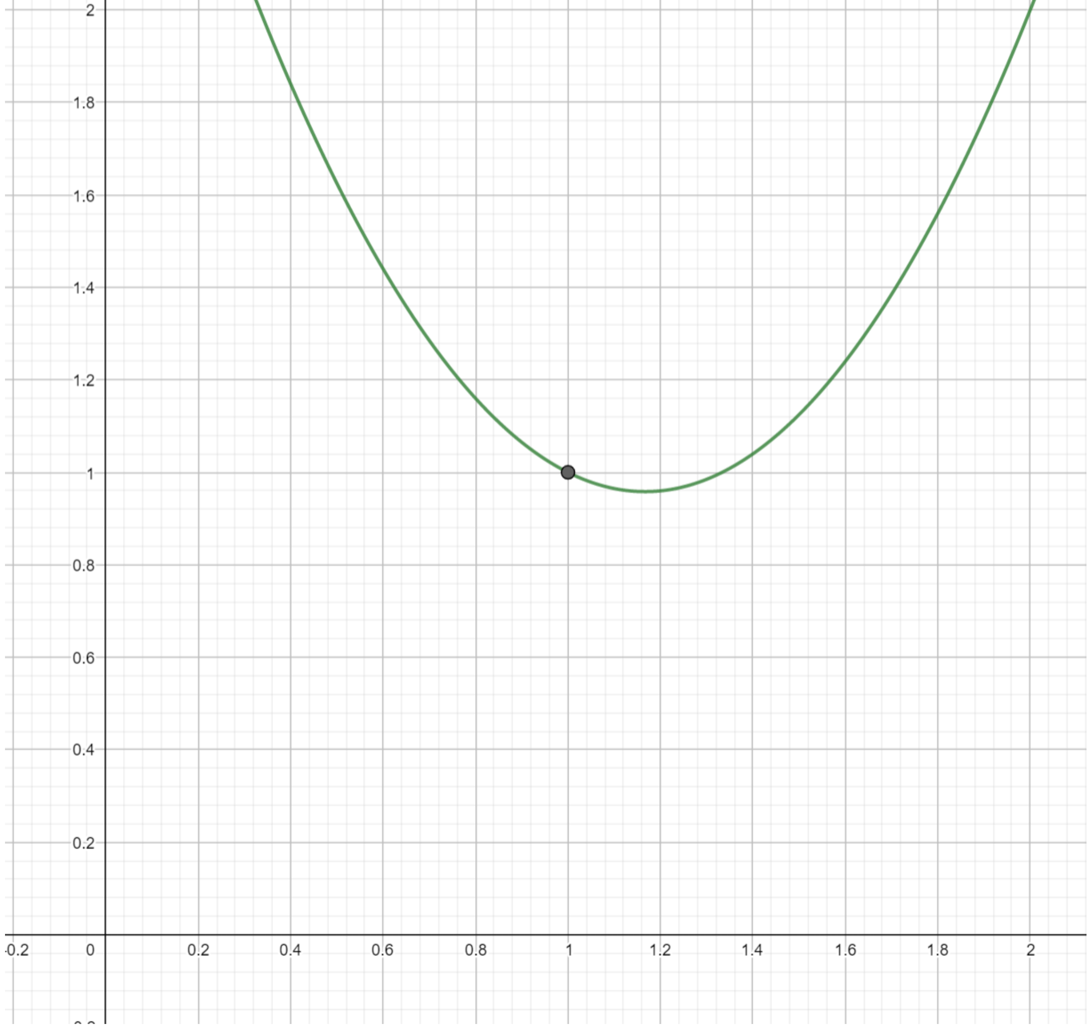


Figure 14.2.2. Even if we only know that $f(1) = 1$, $f'(1) = -\frac{1}{2}$, and $f''(1) = 3$, we can plot the degree-2 Taylor approximation to f and learn about how f behaves near $x = 1$ (indicated by the gray point).

In particular, we know that any quadratic with a positive coefficient of x^2 has a graph which is an upward-facing parabola. We conclude that f is concave-up near $x = 1$.

This idea also tells us how to graphically interpret higher derivatives: a function with $f'''(c) > 0$ acts like a cubic with positive leading coefficient; one with $f^{(4)}(c) < 0$ acts like a quartic with a negative leading coefficient; etc.

Example 14.2.3. We want to hand-draw a precise graph of the cosine function near $x = 0$.

The first approach might be to compute some values of cosine near $x = 0$, say $\cos(-.1)$, $\cos(-.2)$, $\cos(.1)$, etc. But those are hard to do by hand! So instead, we'll use a function we know how to compute the values of by hand: the Maclaurin polynomial of degree 2. That's

$$T_2(x) = 1 - \frac{1}{2}x^2 \quad .$$

The values of T_2 are pretty straightforward to compute:

Table 14.2.4.

x	$T_2(x)$
$-.6$	$1 - \frac{1}{2}(-.6)^2 = .82$

$-.4$	$1-\frac{1}{2}(-.4)^2=.92$
$-.2$	$1-\frac{1}{2}(-.2)^2=.98$
0	1
$.2$	$1-\frac{1}{2}(.2)^2=.98$
$.4$	$1-\frac{1}{2}(.4)^2=.92$
$.6$	$1-\frac{1}{2}(.6)^2=.82$

and that gives us a decent start to plotting.

Checkpoint 183.

14.3 Computing Taylor Polynomials

There's always at least one way to compute a Taylor polynomial: use the formula in Definition 14.1.1. But that usually involves taking a lot of derivatives, which might not be so helpful if the function in question is complicated. So instead, we're going to approach the computation of Taylor polynomials the way we approached computing derivatives and integrals: list out the Taylor polynomials for some basic functions, and then state some rules for how the Taylor polynomial of a complicated function is related to those of its parts.

The basic idea behind all these rules is: Taylor polynomials are polynomials -- so you can do all of the lovely algebra to them that you learned in middle and high school.

Table 14.3.1. Some common Maclaurin polynomials.

$f(x)$	Maclaurin polynomial of degree 4
e^x	$1+x+\frac{1}{2}x^2+\frac{1}{6}x^3+\frac{1}{24}x^4$
$\sin(x)$	$x-\frac{1}{6}x^3$
$\cos(x)$	see Checkpoint 183
$\frac{1}{1-x}$	$1+x+x^2+x^3+x^4$
$\frac{1}{1+x}$	$1-x+x^2-x^3+x^4$
$\ln(1+x)$	$x-\frac{1}{2}x^2+\frac{1}{3}x^3-\frac{1}{4}x^4$

Simple Compositions. Let's say we want to compute the degree-8 Maclaurin polynomial for $f(x)=\sin(x^2)$. We could go ahead and differentiate 8 times. But to save ourselves some work, let's switch variables. Set $t=x^2$. Table 14.3.1 tells us that

$$\sin(t)\approx t-\frac{1}{6}t^3$$

and that this is the best approximation by a polynomial which is degree 4 in t. Notice that being degree 4 in t is the same as being degree 8 in x. If we substitute $t=x^2$ back into our polynomial, we get:

$$\sin(x^2)\approx x^2-\frac{1}{6}x^6$$

which is our answer.

This trick works with powers of x and multiples of x. For example, the degree-5 Maclaurin polynomial for $f(x)=e^{3x}$ is

$$
\begin{aligned}
1+(3x)+\frac{1}{2}(3x)^2+\frac{1}{6}(3x)^3+\frac{1}{24}(3x)^4+\frac{1}{120}(3x)^5 \\
=1+3x+\frac{9}{2}x^2+\frac{27}{6}x^3+\frac{81}{24}x^3+\frac{243}{120}x^5
\end{aligned}
$$

We can also do this with shifts: the degree-3 Maclaurin polynomial for $\ln(2+x)=\ln(1+(x+1))$ is

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

Products. Let's say we want to compute the degree-6 Maclaurin polynomial for the function $h(x)=e^x\sin(x)$. To use the formula in Definition 14.1.1, we would have to take sixth derivatives. Because of the product rule, $h'(x)$ will have two terms; $h''(x)$ will have four terms; etc. So that seems like kind of a mess. On the other hand, Table 14.3.1 tells us that

$$
\begin{aligned}
e^x &\approx 1+x+\frac{1}{2}x^2+\frac{1}{6}x^3+\frac{1}{24}x^4+\frac{1}{120}x^5+\frac{1}{720}x^6 \\
\sin(x) &\approx x-\frac{1}{6}x^3+\frac{1}{120}x^5
\end{aligned}
$$

The Taylor polynomials *are polynomials*, so let's treat them like polynomials and multiply!

Checkpoint 184.

Let's go ahead and do the multiplication. I've color-coded the terms coming from e^x to help with the bookkeeping.

$$
\begin{aligned}
e^x\sin(x) \approx &\left(1+x+\frac{1}{2}x^2+\frac{1}{6}x^3+\frac{1}{24}x^4+\frac{1}{120}x^5+\frac{1}{720}x^6\right)\left(x-\frac{1}{3}x^3+\frac{1}{120}x^5\right) \\
\approx &1\left(x-\frac{1}{6}x^3+\frac{1}{120}x^5\right) \\
&+x\left(x-\frac{1}{6}x^3+\frac{1}{120}x^5\right) \\
&+\frac{1}{2}x^2\left(x-\frac{1}{6}x^3+\frac{1}{120}x^5\right) \\
&+\frac{1}{6}x^3\left(x-\frac{1}{6}x^3+\frac{1}{120}x^5\right) \\
&+\frac{1}{24}x^4\left(x-\frac{1}{6}x^3+\frac{1}{120}x^5\right) \\
&+\frac{1}{120}x^5\left(x-\frac{1}{6}x^3+\frac{1}{120}x^5\right) \\
&+\frac{1}{720}x^6\left(x-\frac{1}{6}x^3+\frac{1}{120}x^5\right) \\
\approx &x-\frac{1}{6}x^3+\frac{1}{120}x^5 \\
&+x^2-\frac{1}{6}x^4+\frac{1}{120}x^6 \\
&+\frac{1}{2}x^3-\frac{1}{12}x^5+\frac{1}{240}x^7 \\
&+\cdots
\end{aligned}
$$

I've left off the remaining rows; you are an expert at manipulating polynomials so you can complete them yourself. But I want to point something out here: there are *lots* of terms with degrees greater than 6. Because we want a polynomial of degree 6, we can just omit these.

Checkpoint 185.

(…n.edu/~ancoop/1070/section-14.html#checkpoint-Maclaurin-product-final)

More Complicated Compositions. We can also use it with more complicated compositions such as $f(x) = \cos(e^x - 1)$. Let's find the degree-3 Maclaurin polynomial for $\cos(e^x - 1)$. We start with the degree-3 Maclaurin polynomial for $\cos(t)$, applied with $t = e^x - 1$:

$$\cos(e^x - 1) \approx 1 - \frac{1}{2}(e^x - 1)^2$$

Now, we remind ourselves that $e^x \approx 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3$, so that $t = e^x - 1 \sim x + \frac{1}{2}x^2 + \frac{1}{6}x^3$, and make that substitution:

$$\cos(e^x - 1) \sim 1 - \frac{1}{2}\left(x + \frac{1}{2}x^2 + \frac{1}{6}x^3\right)^2$$

Like we did with products, now we bite the bullet and do the algebra:

$$\begin{aligned} 1 - \frac{1}{2}\left(x + \frac{1}{2}x^2 + \frac{1}{6}x^3\right)^2 & \\ &= 1 - \frac{1}{2}\left(x^2 + \left(\frac{1}{2}x^2\right)^2 + \left(\frac{1}{6}x^3\right)^2 + 2x\frac{1}{2}x^2 + 2x\frac{1}{6}x^3 + 2\frac{1}{2}x^2\frac{1}{6}x^3\right) \\ &= 1 - \frac{1}{2}x^2 - \frac{1}{3}x^3 - \frac{7}{24}x^4 - \frac{1}{12}x^5 - \frac{1}{72}x^6 \end{aligned}$$

Notice that we actually got a degree 6 polynomial. Since we only needed to approximate f to order 3, we take the cubic part and our answer is:

$$T_3(x) = \frac{1}{2} - \frac{1}{2}x^2 - \frac{1}{6}x^3 \quad .$$

Plotting T_3 together with f confirms that this is a good approximation, at least near $x = 0$:

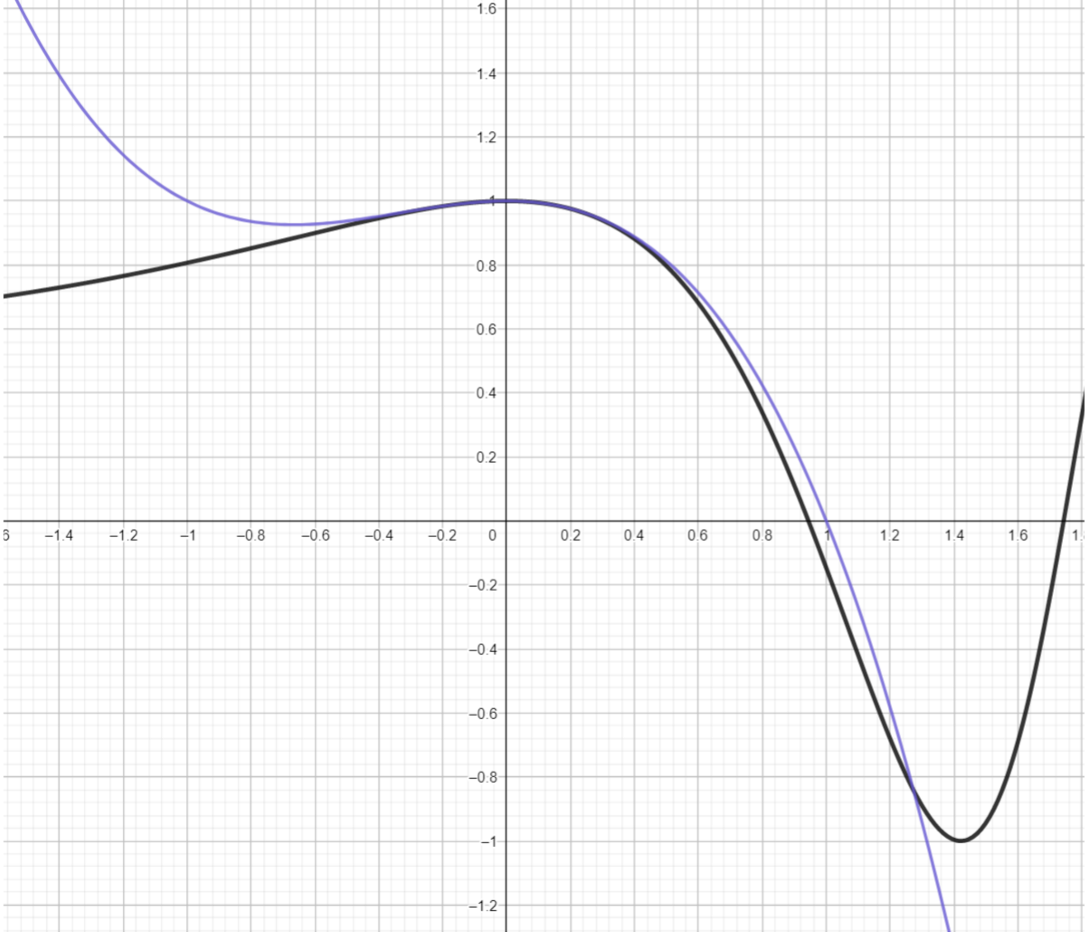
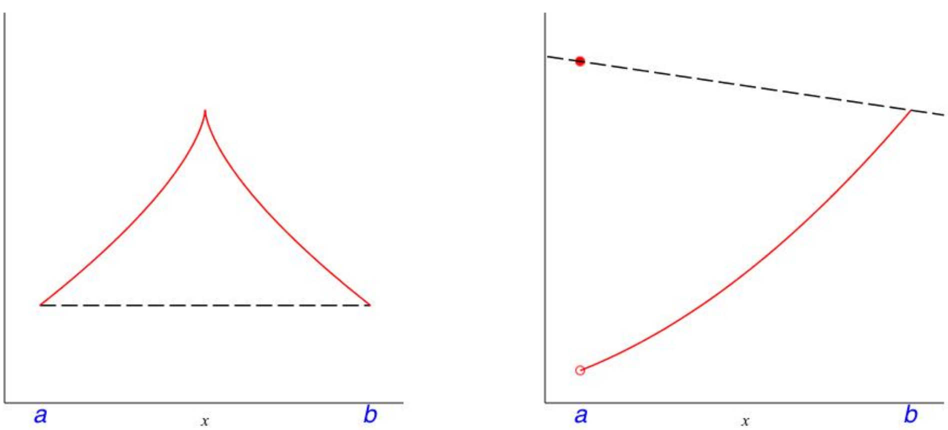


Figure 14.3.2. T_3 (purple) and $f(x) = \cos(e^x - 1)$ (black).

14.4 The Mean Value Theorem and Taylor's Theorem

In class we will discuss the following theorem. Please read it now to see whether it makes intuitive sense to you. The hypotheses will be filled in after the class discussion centered on the counterexamples in Figure 14.4.1.



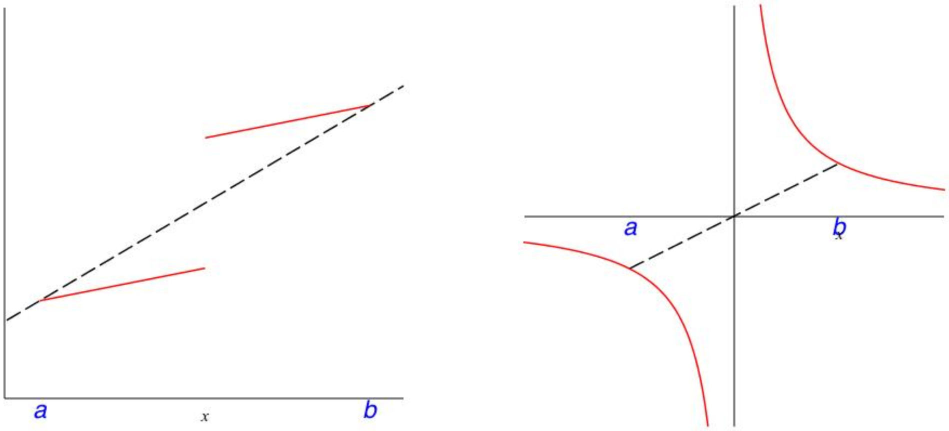


Figure 14.4.1. In each case the dashed line illustrates the average slope $\dfrac{f(b)-f(a)}{b-a}$.

Theorem 14.4.2. Mean value theorem. Let f be a function and $a < b$ be real numbers. Assuming ***SOME HYPOTHESES*** there must be a number $c \in (a, b)$ where the slope of $f$is equal to the average slope over (a, b), that is,

$$f'(c) = \frac{f(b)-f(a)}{b-a}\,. \tag{14.4.1}$$

Checkpoint 186. n.edu/~ancoop/1070/section-14.html#project-186)

Example 14.4.3. Let $f(x)$ be the position (mile marker) of a PA Turnpike driver at time x. Suppose the driver entered the Turnpike at Mile 75 (New Stanton) at 4pm and exited at Mile 328 (Valley Forge) at 7pm. What does the Mean Value Theorem tell you in this case? The average slope of f over interval $[4pm, 7pm]$ is the difference quotient $(f(7)-f(4))/(7-4) = (325-75)/3 = 83\frac{1}{3}$. Thus there is some time c between 4pm and 7pm that $f'(c) = 83\frac{1}{3}$ MPH, in other words, that this driver was traveling at a speed of $83\frac{1}{3}$ MPH.

Checkpoint 187. n.edu/~ancoop/1070/section-14.html#project-187)

Example 14.4.4. Let $f(x) := 1/x$ and let $a < b$ be positive real numbers. What, explicitly in terms of a and b, is the number c guaranteed by the Mean value theorem?

Actually, Example 14.4.4 is a bit beside the point. The Mean Value Theorem only asserts that *there is some number* c; it says nothing about what c actually is.

Why is the Mean Value Theorem in this chapter? We can rewrite (14.4.1) to read:

$$f(b) = f(a) + f'(c)(b-a) \tag{14.4.2}$$

which looks an awful lot like the linear approximation to f at $x = a$:

$$L(x) = f(a) + f'(a)(x-a)$$

except we've plugged in $x = b$ and the place we're evaluating f' is different.

It turns out, this same sort of theorem is true for the higher-order polynomial approximations, too:

Theorem 14.4.5. Taylor's Theorem. Let f be a function and $a < b$ be real numbers. Assuming ***SOME HYPOTHESES***, there must be a number $c \in (a, b)$ where

$$f(b) = f(a) + f'(a)(b-a) + \frac{1}{2}f''(a)(b-a)^2 + \cdots + \frac{1}{n!}f^{(n)}(a)(b-a)^n + \frac{1}{(n+1)!}f^{(n+1)}(c)(b-a)^{n+1}$$

Checkpoint 188.

The last term -- the term involving the mysterious c -- should be thought of as an $error$;Taylor's Theorem says that, if we're willing to accept an amount of error related to the $n+1$st derivative, we can pretend that $f(b)$ is a polynomial function of b.

Checkpoint 189.

Checkpoint 190.