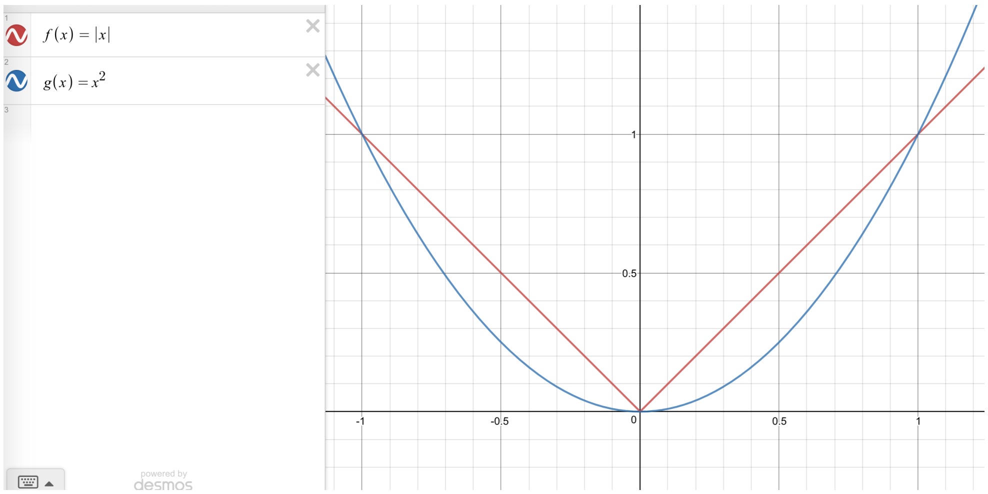
8 The derivative

We think, informally, as continuity as being a measure of "smoothness": if a function f is continuous at a, then small changes in the input to f near a lead only to small changes in the output.

But there are definitely "degrees of smoothness". The functions $f(x) = |x|$ and $g(x) = x^2$ (see figure) are both continuous at 0, and both achieve their minima at 0, but their graphs behave very differently near 0 — g curves gently, while f has a sharp point.



The tool we introduce now, that (among many many other things) distinguishes these two behaviors, is the familiar tool of the *derivative*.

8.1 Two motivating examples

Instantaneous velocity Suppose that a particle is moving along a line, and that its distance from the origin at time t is given by the function $s(t)$.

It's easy to calculate the *average velocity* of the particle over a time interval for, say, time $t = a$ to time $t = b$: it's the total displacement of the particle, $s(b) - s(a)$, divided by the total time, $b - a$.[104]

But what is the *instantaneous velocity* of the particle at a certain time t? To make sense of this, we might do the following: over a small time interval $[t, t + \Delta t]$ (starting

[104] Remember that velocity is a *signed* quantity: if a particle starts 10 units to the right of the origin, and two seconds later is 14 units to the right of the origin, then its average velocity over those two seconds is $(14 - 10)/2 = 2$ units per second, positive because the particle has progressed further from the origin. If, on the other hand, it starts 14 units to the right of the origin, and two seconds later is 10 units to the right of the origin, then its average velocity over those two seconds is $(10 - 14)/2 = -2$ units per second, negative because the particle has progressed *closer* to the origin. In both cases the average *speed* is the same — 2 units per second — speed being the absolute value of velocity.

at time t, ending at time $t + \Delta t$), with $\Delta t > 0$, the average velocity is

$$\frac{\text{displacement}}{\text{time}} = \frac{s(t+\Delta t) - s(t)}{\Delta t}.$$

Similarly over a small time interval $[t + \Delta t, t]$, with $\Delta t < 0$, the average velocity

$$\frac{s(t) - s(t+\Delta t)}{-\Delta t} = \frac{s(t+\Delta t) - s(t)}{\Delta t}.$$

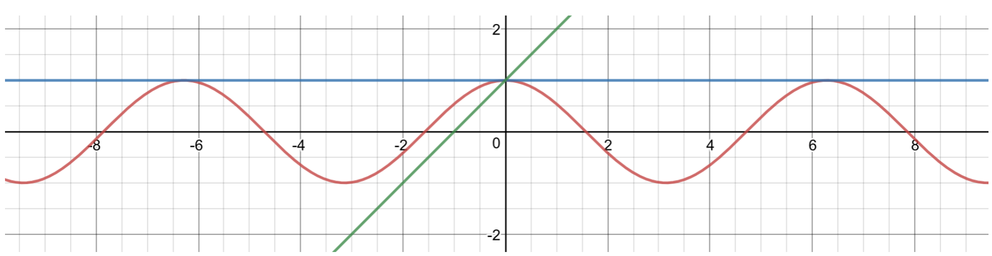
If this common quantity, $(s(t+\Delta t) - s(t))/\Delta t$, is approaching a limit as Δt approaches 0, then it makes sense to *define* instantaneous velocity at time t to be that limit, that is, to be

$$\lim_{\Delta t \to 0} \frac{s(t+\Delta t) - s(t)}{\Delta t}.$$

Tangent line : What is the equation of the tangent line to the graph of function f at some point $(a, f(a))$ on the graph? To answer that, we must answer the more fundamental question, "what do we mean by 'tangent line'?". A preliminary definition might be that

> a tangent line to a graph at a point on the graph is a straight line that touches the graph only at that point.

This is a fairly crude definition, and fairly clearly doesn't work: the line $y = 1$ touches the graph of $y = \cos x$ infinitely many times, at $x = 0, \pm\pi, \pm 2\pi, \ldots$, but clearly should be declared to be a tangent line to $y = \cos x$ at $(0, 1)$; on the other hand, the line $y = 10x$ touches the graph of $y = \cos x$ only once, at $(0, 1)$, but clearly should *not* be declared to be a tangent line to $y = \cos x$ at $(0, 1)$.



What we really want to say, is that a tangent line to a graph at a point on the graph is a straight line that passes through the point, and that just "glances off" the graph at that point, or is "going in the same direction as the graph" at that point, or "has the same slope as the graph does" at that point.

Clearly these phrases in quotes need to be made more precise. What do we mean by "the slope of a graph, at a point"? We can make this precise, in a similar way to the way we made precise the notion of instantaneous velocity.

A *secant line* of a graph is a straight line that connects two points $(x_1, f(x_1))$, $(x_2, f(x_2))$ on the graph. It makes perfect sense to talk of the "slope of a secant line": it is

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

The define the slope at a point $(a, f(a))$, we can consider the slope of the secant line between $(a, f(a))$ and $(a+h, f(a+h))$ for small $h > 0$, or between $(a+h, f(a+h))$ and $(a, f(a))$ for small $h < 0$. In both cases, this is

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

This secant slope seems like it should be a reasonable approximation of the slope of the graph *at* the point $(a, f(a))$; and in particular, if the slopes of the secant lines approach a limit as h approaches 0, then it makes a lot of sense to *define* the slope at $(a, f(a))$ to be that limit:

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

Going back to the original question, once we have found the slope, call it m_a, we can easily figure out the equation of the tangent line to the graph at $(a, f(a))$, since it is the unique straight line that passes through $(a, f(a))$ and has slope m_a:

$$y - f(a) = m_a(x - a) \quad \text{or} \quad y = m_a(x - a) + f(a) \quad \text{or} \quad y = m_a x - m_a a + f(a).$$

The two expressions we have obtained from these two examples — $\lim_{\Delta t \to 0}(s(t + \Delta t) - s(t))/\Delta t$ and $\lim_{h \to 0}(f(a+h) - f(a))/h$ — are of exactly the same form. Since the same expression has cropped up in two rather different-looking applications, it makes sense to look at the expression as an interesting object in its own right, and study its properties. That is exactly what we will do in this section.

8.2 The definition of the derivative

Let f be a function, and let f be defined at and near some number a (i.e., suppose there is some $\Delta > 0$ such that all of the interval $(a - \Delta, a + \Delta)$ is in the domain of f.

Definition of derivative Say that f is *differentiable* at a if

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

exists. If f is differentiable at a, then write $f'(a)$ for the value of this limit; $f'(a)$ is referred to as the *derivative* of f at a[105].

[105] There are many alternate notations for the derivative:

From the previous section, we obtain immediately two interpretations of the quantity $f'(a)$:

Velocity if $s(t)$ measures the position at time t of a particle that is moving along the number line, then $s'(a)$ measures the velocity of the particle at time a.

Slope $f'(a)$ is the slope of the tangent line of the graph of function f at the point $(a, f(a)$. Consequently the equation of the tangent line is

$$y = f'(a)(x - a) + f(a).$$

Once we have the notion of the derivate of a function at a point, it's a very short leap to considering the derivative as a *function*.

Definition of the derivative function If $f : D \to \mathbb{R}$ is a function defined on some domain D, then the *derivative* of f is a function, denoted f'[106], whose domain is $\{a \in D : f \text{ differentiable at } a\}$[107], and whose value at a is the derivative of f at a.

As we will see in a series of examples, the domain of f' may the same as the domain of f, or slightly smaller, or *much* smaller.

Before going on to the examples, we mention an alternate definition for of the definition of derivative:

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

Indeed, suppose $\lim_{b\to a}(f(b)-f(a))/(b-a)$ exists and equal L. Then for all $\varepsilon > 0$ there is $\delta > 0$ such that whenever b is within δ of a (but not equal to a), we have that $(f(b) - f(a))/(b - a)$ is within ε of L. Rewriting b as $a + h$ (so $b - a = h$), this says that whenever $a + h$ is within δ of a (but not equal to a), that is, whenever h is within δ of 0 (but not equal to 0), we have that $(f(a + h) - f(a))/h$ is within ε of L. This says $\lim_{h\to 0}(f(a + h) - f(a))/h$ exists and equal L. The converse direction goes along the same lines.

- $f'(a)$
- $\dot{f}(a)$
- $\frac{d}{dx} f(x)\,|_{x=a}$
- $\frac{df(x)}{dx}\,|_{x=a}$
- $\frac{df}{dx}\,|_{x=a}$
- $\frac{dy}{dx}\,|_{x=a}$ (if y is understood to be another name for f)
- $\dot{y}(a)$ (again, if y is another name for f).

We will almost exclusively use the first of these.

[106] or $\frac{df}{dx}$, or $\dot{f}$.

[107] We will shortly modify this definition slightly, to deal with functions which are defined on closed intervals such as $[0, 1]$; we will introduce a notion of "differentiable from the right" and "differentiable from the left" so as to be able to talk about what happens at the end points of the interval.

8.3 Some examples of derivatives

Given the work we have done on limits and continuity, calculating the derivatives of many simple function, even directly from the definition, is fairly straightforward. We give a bunch of examples here.

Constant function $f(x) = c$, where c is some fixed real number. Presumably, the derivative of this function is 0 at any real a, that is,

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

Notice that we can't verify this instantly by appealing to continuity of the expression $(f(a+h) - f(a))/h$, viewed as a function of h, at $h = 0$, and then just evaluating the expression at $h = 0$; the expression is not only not continuous at $h = 0$, it is not even defined at $h = 0$! This will be a common theme in computing derivatives: the expression $(f(a+h) - f(a))/h$ (viewed as a function of h), regardless of the f under consideration, will *always* not be defined at $h = 0$, since the numerator and the denominator both evaluate to 0 at $h = 0$. So here, and in all other examples that we do, we will have to engage in some algebraic manipulation of the expression $(f(a+h) - f(a))/h$. The goal of the manipulation is to try and find an alternate expression, that is equal to $(f(a+h) - f(a))/h$ for all h except (possibly) $h = 0$ (the one value of h we do not really care about); and then see if we can use some of our previous developed techniques to evaluate the limit as h goes to 0 of the new expression.

For any real a we have, for $h \neq 0$,

$$\frac{f(a+h) - f(a)}{h} = \frac{c - c}{h} = \frac{0}{h} = 0,$$

and so

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

from which we conclude that f is differentiable at all a, with derivative 0. (Of course: the line $y = c$ is clearly the tangent line to f at *any* point, and this line has slope 0; or, if a particle is located at the same position, c, on the line at all times, its velocity at all times is 0.)

In this example, f' is the constant 0 function, on the same domain ($\mathbb{R}$) as f.

This example is really simple, but it is worth doing in detail for two reasons. First, a philosophical reason: to act as a reality check for the definition, and our understanding of the definition. Second, a practical reason: to illustrate a subtlety of writing up proofs from first principles of derivatives of functions. It's very tempting to argue that $f'(a) = 0$ by writing

$$\text{“}f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = \lim_{h \to 0} \frac{c - c}{h} = \lim_{h \to 0} \frac{0}{h} = \lim_{h \to 0} 0 = 0.\text{”}$$

But this presentation, starting with the expression $f'(a)$, presupposes that the limit that defines the derivative actually exists. We'll come across *plenty* of examples where the limit doesn't exists. The more correct mathematical approach is to do the algebraic manipulation to $(f(a+h)-f(a))/h$ *first*, and then, when a nicer expression has been arrived at, whose limit near 0 can easily be computed, introduce the limit symbol. That was how we approached the write-up above, although frequently in the future we will be sloppy and write "$\lim - - -=$" before we've formally decided that the limit exists.[108]

Linear function Let $f(x) = mx + b$ for real constants m, b. Since the graph of f is a straight line with slope m, it should be its own tangent at all points, and so the derivative at all points should be m. We verify this. As discussed in the last example, we will do this slightly sloppily, beginning by assuming that the limit exists.

For each real a we have

$$\begin{aligned} f'(a) &= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \\ &= \lim_{h \to 0} \frac{m(a+h) + b - (ma+b))}{h} \\ &= \lim_{h \to 0} \frac{mh}{h} \\ &= \lim_{h \to 0} m \\ &= m, \end{aligned}$$

as we suspected. The key line was the second from last — dividing above and below by h was valid, because we never consider $h = 0$ when calculating the limit near 0. In the last line, we use that the constant function is continuous everywhere, so the limit can be evaluated by direct evaluation.

In this example, f' is the constant m function, on the same domain ($\mathbb{R}$) as f.

Quadratic function Let $f(x) = x^2$. There is every reason to expect that this function is differentiable everywhere — its graph, on any graphing utility, appears smooth. There is little reason to expect a particular value for the derivative, as we did in the last two examples[109]. We just go through the calculation, and see what comes out. This time, we'll do it in what might be called the "proper" way, not initially assuming the existence of the derivative.

[108]We actually already did this above, when we wrote "and so

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

[109]Not entirely true — when we motivate the product rule for differentiation, we see a good reason to expect that the derivative of $f(x) = x^2$ at a is $2a$.

For each real a, and for $h \neq 0$, we have

$$\begin{aligned} \frac{(a+h)^2 - a^2}{h} &= \frac{a^2 + 2ah + h^2 - a^2}{h} \\ &= \frac{2ah + h^2}{h} \\ &= 2a + h. \end{aligned}$$

Since $\lim_{h\to 0}(2a+h)$ evidently exists and equals $2a$, we conclude that $\lim_{h\to 0}((a+h)^2 - a^2)/h$ exists and equals $2a$, and so for all real a,

$$f'(a) = 2a.$$

In this example, f' is the linear function $x \mapsto 2x$, on the same domain ($\mathbb{R}$) as f.

Power function In general, calculating the derivative of $f(x) = x^n$ for $n \in \mathbb{N}$ at arbitrary real a is not much harder than in the special case of $n = 2$, just as long as we bring the right tool to the algebraic manipulation. Since we'll be faced with the expression $(a+h)^n - a^n$, it seems that the Binomial Theorem is probably the[110] right tool.

For each real a, and for $h \neq 0$, we have

$$\begin{aligned} \frac{(a+h)^n - a^n}{h} &= \frac{a^n + \binom{n}{1}a^{n-1}h + \binom{n}{2}a^{n-2}h^2 + \cdots + \binom{n}{n-1}ah^{n-1} + h^n - a^n}{h} \\ &= \frac{\binom{n}{1}a^{n-1}h + \binom{n}{2}a^{n-2}h^2 + \cdots + \binom{n}{n-1}ah^{n-1} + h^n}{h} \\ &= \binom{n}{1}a^{n-1} + \binom{n}{2}a^{n-2}h + \cdots + \binom{n}{n-1}ah^{n-2} + h^{n-1}. \end{aligned}$$

Now

$$\lim_{h\to 0} \binom{n}{1}a^{n-1} = \binom{n}{1}a^{n-1} = na^{n-1},$$

while

$$\lim_{h\to 0} \binom{n}{2}a^{n-2}h = \lim_{h\to 0} \binom{n}{3}a^{n-3}h^2 = \cdots = \lim_{h\to 0} \binom{n}{n-1}ah^{n-2} = \lim_{h\to 0} h^{n-1} = 0,$$

all these facts following from our previous work on continuity. So by the sum part of the sum/product/reciprocal theorem for limits, we conclude that

$$\lim_{h\to 0} \binom{n}{1}a^{n-1} + \binom{n}{2}a^{n-2}h + \cdots + \binom{n}{n-1}ah^{n-2} + h^{n-1} = na^{n-1}.$$

But then it follows that

$$\lim_{h\to 0} \frac{(a+h)^n - a^n}{h} = na^{n-1};$$

[110] or at least a

in other words, f is differentiable for all real a, with

$$f'(a) = na^{n-1}.$$

In this example, f' is the power function $x \mapsto nx^{n-1}$, on the same domain ($\mathbb{R}$) as f.

Quadratic reciprocal One final example in the vein of the previous ones: $f(x) = 1/x^2$. As long as $a \neq 0$, we have

$$\begin{aligned}
f'(a) &= \lim_{h\to 0} \frac{f(a+h) - f(a)}{h} \\
&= \lim_{h\to 0} \frac{\frac{1}{(a+h)^2} - \frac{1}{a^2}}{h} \\
&= \lim_{h\to 0} \frac{a^2 - (a+h)^2}{(a+h)^2a^2h} \\
&= \lim_{h\to 0} \frac{-2ah - h^2}{(a+h)^2a^2h} \\
&= \lim_{h\to 0} \frac{-2a - h}{(a+h)^2a^2} \\
&= \frac{-2a}{a^2a^2} \\
&= \frac{-2}{a^3}.
\end{aligned}$$

In this example, f' is the function $x \mapsto -2x^3$, on the same domain ($\mathbb{R} \setminus \{0\}$) as f.

Absolute value function Here we consider $f(x) = |x|$. We would strongly expect that for $a > 0$, we have f differentiable at a, with derivative 1, because a little neighborhood around such a, we have that $f(x) = x$; indeed, for $a > 0$ we have that for all sufficiently small h (say, for all $h < a/2$)

$$\frac{|a+h| - |a|}{h} = \frac{a+h-a}{h} = \frac{h}{h} = 1,$$

and so $\lim_{h\to 0}(|a+h| - |a|)/h = \lim_{h\to 0} 1 = 1$. We can similarly verify that for all $a < 0$, $f'(a) = -1$. But at $a = 0$, something different happens:

$$\frac{|0+h| - |0|}{h} = \frac{|h|}{h},$$

and we know that $\lim_{h\to 0} |h|/h$ *does not exist*. So, this is our first example of a function that is *not* always differentiable; the domain of f' here is $\mathbb{R} \setminus \{0\}$ while the domain of f is $\mathbb{R}$.

We should not have expected $f(x) = |x|$ to be differentiable at 0, as there is no coherent "direction" that the graph of the function is going near 0 — if we look to the right of zero, it is increasing consistently at rate 1, while if we look to the left of zero, it is *decreasing* consistently at rate 1. Nor is there obviously an unambiguous tangent line.

The comments in the previous paragraph suggest that it might be useful to define notions of right and left derivatives, as we did with continuity. Say that f is *right differentiable* at a, or *differentiable from the right*, or *differentiable from above*, if

$$\lim_{h\to 0^+} \frac{f(a+h)-f(a)}{h}$$

exists, and if it does, denote by $f'_+(a)$ the value of the limit. Say that f is *left differentiable* at a, or *differentiable from the left*, or *differentiable from below*, if

$$\lim_{h\to 0^-} \frac{f(a+h)-f(a)}{h}$$

exists, and if it does, denote by $f'_-(a)$ the value of the limit. It's a (hopefully routine, by now) exercise to check that

> f is differentiable at a if and only if f is both left and right differentiable at a, and the two one-sided derivatives have the same value; in this case that common value is the value of the derivative at a.

So, for example, with $f(x) = |x|$ we have

$$f'_+(0) = \lim_{h\to 0^+} \frac{|h+0|-|0|}{h} = \lim_{h\to 0^+} \frac{h}{h} = 1$$

while

$$f'_-(0) = \lim_{h\to 0^-} \frac{|h+0|-|0|}{h} = \lim_{h\to 0^+} \frac{-h}{h} = -1,$$

so that f is not differentiable at 0.

Piecewise defined functions Consider

$$f(x) = \begin{cases} x^2 & \text{if } x < 1 \\ ax+b & \text{if } x \geq 1 \end{cases}$$

where a, b are some constants. What choices of a, b make f both continuous and differentiable on all reals?

Well, clearly f is both continuous and differentiable on all of $(-\infty, 1)$ and on all of $(1, \infty)$. What about at 1? We have

$$\lim_{x\to 1^+} f(x) = \lim_{x\to 1^+} ax+b = a+b$$

and

$$\lim_{x\to 1^-} f(x) = \lim_{x\to 1^-} x^2 = 1,$$

so in order for f to be continuous at 1, we require $a+b=1$. For differentiability, at 1, we have

$$\lim_{h\to 0^+}\frac{f(1+h)-f(1)}{h}=\lim_{h\to 0^+}\frac{a+ah+b-(a+b)}{h}=\lim_{h\to 0^+}\frac{ah}{h}=\lim_{h\to 0^+}a=a,$$

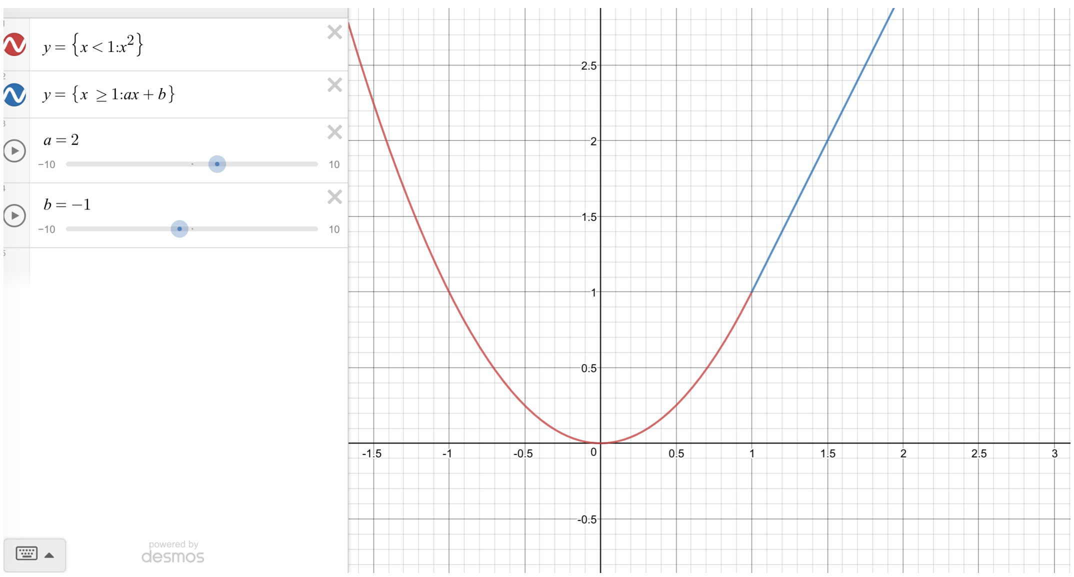
and (recalling that $a+b=1$, since we require f to be continuous at 1)

$$\lim_{h\to 0^-}\frac{f(1+h)-f(1)}{h}=\lim_{h\to 0^-}\frac{(1+h)^2-(a+b)}{h}=\lim_{h\to 0^-}\frac{2h+h^2}{h}=\lim_{h\to 0^-}(2+h)=2.$$

So, for f to be differentiable at 1 we require $a=2$; and since $a+b=1$ this says $b=-1$. The function we are considering is thus

$$f(x)=\left\{\begin{array}{cl} x^2 & \text{if } x<1 \\ 2x-1 & \text{if } x\geq 1. \end{array}\right.$$

Here is the graph. It shows the two pieces not justing fitting together at 1, but fitting together *smoothly*.



The square root function Consider $f(x)=\sqrt{x}$, defined on $[0,\infty)$. Tho compute its derivative at any $a\in(0,\infty)$ we proceed in the usual way:

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

So f is differentiable on $(0, \infty)$, with derivative $f'(a) = 1/2\sqrt{a}$.

What about at 0? Because f is not defined for negative inputs, we must consider a one sided derivative, in particular the right derivative, and we have

$$\lim_{h \to 0^+} \frac{\sqrt{0+h} - \sqrt{0}}{h} = \lim_{h \to 0^+} \frac{1}{\sqrt{h}}.$$

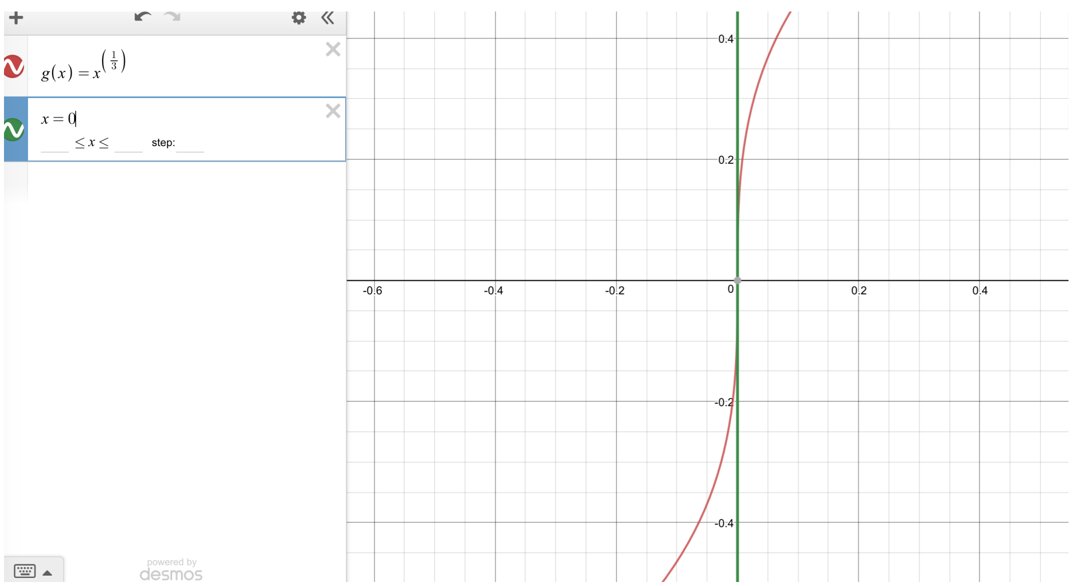
This limit does not exist, so f is not left differentiable at 0.

A more dramatic example in a similar vein comes from considering $g(x) = x^{1/3}$, which has all of $\mathbb{R}$ as its domain. By a similar calculation to above, we get that f is differentiable at all $a \neq 0$, with derivative $f'(a) = 1/(3a^{2/3})$. At $a = 0$ we have

$$\lim_{h \to 0} \frac{(0+h)^{1/3} - 0^{1/3}}{h} = \lim_{h \to 0} \frac{1}{h^{2/3}},$$

which again does not exist, so g is not differentiable at 0.

What's odd about this is that from a drawing of the graph of g, it seems that g has an unambiguous slope/tangent line at the point $(0, 0)$:



It is the *vertical* line, $x = 0$. We are failing to see this in the math, because the vertical line has infinite slope, and we have no real number that captures that.[111]

$\sin(1/x)$ **and variants** Consider the three functions

$$f_1(x) = \sin(1/x), \;\; x \neq 0$$
$$f_2(x) = x\sin(1/x), \;\; x \neq 0$$
$$f_3(x) = x^2\sin(1/x), \;\; x \neq 0,$$

[111]Shortly we will talk about "infinite limits" and rectify this deficiency.

with $f_1(0) = f_2(0) = f_3(0) = 0$.

All three of these functions have domain $\mathbb{R}$. What about the domains of their derivatives? Presumably they are all differentiable at all non-zero points.[112]

What about at 0? For f_1, using $f_1(0) = 0$ we have (if the limit exists)

$$f'(0) = \lim_{h \to 0} \frac{\sin(1/h)}{h}.$$

It's a slightly tedious, but fairly instructive, exercise to verify that this limit does not exist; so f_1 is not differentiable at 0 (and maybe we shouldn't have expected it to be: it's not even continuous at 0).

For f_2, which is continuous at 0, we have a better chance. But

$$\lim_{h \to 0} \frac{f_2(0+h) - f_2(0)}{h} = \lim_{h \to 0} \frac{h \sin(1/h)}{h} = \lim_{h \to 0} \sin(1/h),$$

which does not exist; so f_2 is not differentiable at 0, either.

For f_3, however, we have

$$f_3'(0) = \lim_{h \to 0} \frac{h^2 \sin(1/h)}{h} = \lim_{h \to 0} h \sin(1/h) = 0,$$

so f_3 *is* not differentiable at 0, with derivative 0.

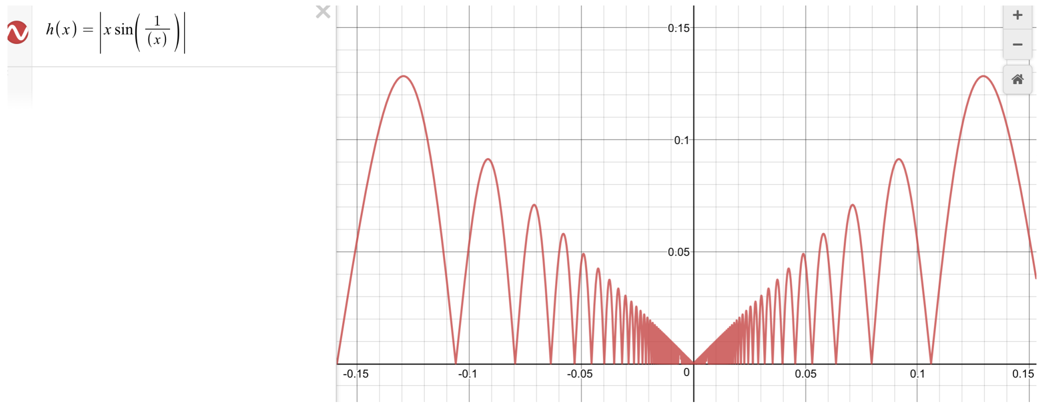
We will return to this example when considering a function which is k times differentiable, but not $(k+1)$ times differentiable.

Weierstrass function It's easy to come up with an example of a function that is defined *everywhere* differentiable *nowhere* — the Dirichlet function works well. But this seems a cheat. The Dirichlet function is nowhere continuous, and if we imagine differentiability to be a more stringent notion of smoothness than continuity, then we might expect that non-continuous functions are for some fairly trivial reason non-differentiable.[113]

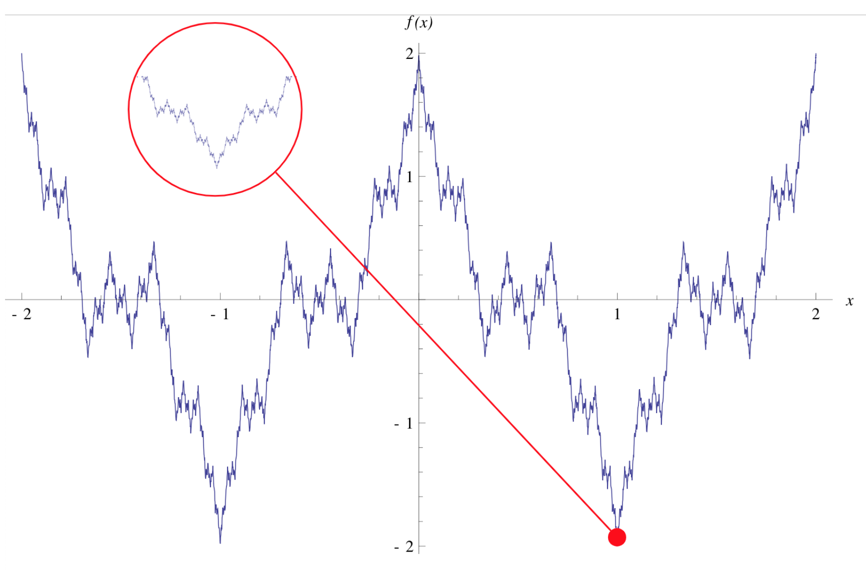
So what about a *continuous* function that is nowhere differentiable? It's fairly easy to produce an example of a function that is continuous everywhere, but that has infinitely many points of non-differentiability; even an infinite collection of points that get arbitrarily close to each other. For example, the function $f(x) = x \sin(1/x)$ is continuous everywhere on its domain $\mathbb{R} \setminus \{0\}$, and presumably differentiable everywhere on its domain; but if we take its absolute value we get a function that is still continuous, but has a sharp point (so a point of non-differentiability) at each place where it touches the axis (see figure below). In other words, if $h(x) = |x \sin(1/x)|$ then while $\text{Domain}(h) = \mathbb{R} \setminus \{0\}$ we have $\text{Domain}(h') = \mathbb{R} \setminus \{0, \pm 1/\pi, \pm 2/\pi, \pm 3/\pi, \ldots\}$.

[112] We will verify this informally soon, using or informal/geometric definition of the trigonometric functions.

[113] This is true, as we'll see in a moment.



It is far less easy to come with an example of a function which is continuous *everywhere*, but differentiable *nowhere*; nor is it easy to imagine what such a function could look like. There *are* examples[114], but they are not as easy to explain as the Dirichlet function (our example of a function that is defined everywhere but continuous nowhere). The first such example was found by Karl Weiestrass in 1872, and so is traditionally called the *Weierstrass function*. It is infinitely jagged, and displays a self-similarity or fractal behavior: zoom in on any portion of the graph, and you see something very similar to the global picture (see figure below).



Higher derivatives Let f be a function on some domain D. As we have been discussing in these examples, there may be some points in the domain of f at which f is differentiable, leading to a function f', the derivative function, which might have a smaller domain than D. But the function f' may itself be differentiable at some points, leading to a function $(f')'$ (which might have a smaller domain than that of f'). Rather than

[114]In fact, in a quite precise sense *most* continuous function are nowhere differentiable.

working with this ungainly notation, we denote the second derivative by f''. Formally, the second derivative of a function f at a point a is defined to be

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

assuming that limit exists — which presupposes that f is both defined at and near a, and is differentiable at and near a.

We may further define the third derivative function, denoted f''', as the derivative of the second derivative function f''. And we can go on; but even without the parentheses, this "prime" notation gets a little ungainly, quickly. We use the notation $f^{(k)}$ to denote the kth derivative of f, for and natural number k (so $f^{(3)} = f'''$ and $f^{(1)} = f'$). By convention, $f^{(0)} = f$.

Physically, if $f(t)$ is the position of a particle at time t, then

- $f'(t)$ is velocity at time t (rate of change of position with respect to time);
- $f''(t)$ is the acceleration at time t (rate of change of velocity with respect to time);
- $f'''(t)$ is the *jerk* at time t (rate of change of acceleration with respect to time), and so on.

Consider, for example, $f(x) = 1/x$, with domain all reals except 0. We have

- $f'(x) = -1/x^2$, domain $\mathbb{R} \setminus \{0\}$;
- $f''(x) = 2/x^3$, domain $\mathbb{R} \setminus \{0\}$;
- $f'''(x) = -6/x^3$, domain $\mathbb{R} \setminus \{0\}$, and so on.

As another example, consider the function that is obtained by splicing the cube function and the square function, i.e.

$$f(x) = \begin{cases} x^3 & \text{if } x \le 0 \\ x^2 & \text{if } x \ge 0. \end{cases}$$

By looking at one sides limits, it is easy to check that f is continuous at 0, differentiable at 0, and even twice differentiable at 0, but *not* thrice differentiable. A homework problem asks for an example of a function that, at least at some points, is differentiable k times, but not $k+1$ times.

Before moving on to some more theoretical properties of the derivative, we mention one more motivation. The tangent line to a curve at a point, as we have defined it, seems to represent a good approximation to the curve, at least close to the point of tangency. Now the tangent line is a *straight* line, and it is relatively easy to calculate exact values of points along a straight line, while the graph of a typical function near a point may well be *curved*, and the graph may be that of a function whose values are hard to compute (we may be dealing with $f(x) = x \sin x / \cos^2(\pi(x + 1/2))$, for example).

This suggests that it might be fruitful to use the point (x, y) on the tangent line at $(a, f(a))$ to the graph of function f, as an approximation for the point $(x, f(x))$ (that's actually on the graph); and to use y as an approximation for $f(x)$. It seems like thais might be particularly useful if x is relatively close to a.

Recalling that the equation of the tangent line at $(a, f(a))$ to the graph of function f is $y = f(a) + f'(a)(x - a)$, we make the following definition:

Linearization of a function at a point The *linearization* $L_{f,a}$ of a function f at a at which the function is differentiable is the function $L_{f,a} : \mathbb{R} \to \mathbb{R}$ given by

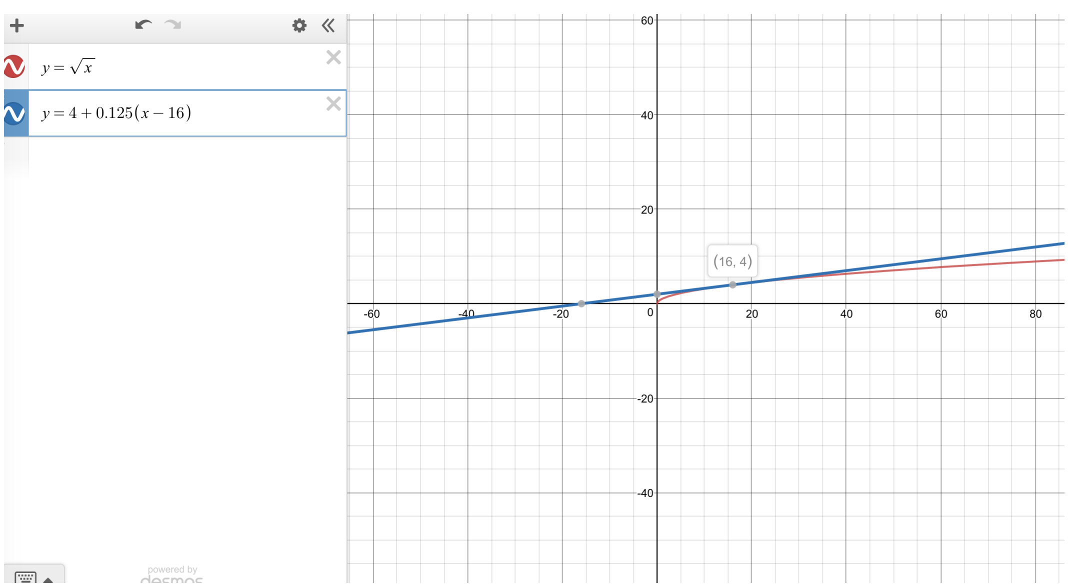
$$L_{f,a}(x) = f(a) + f'(a)(x - a).$$

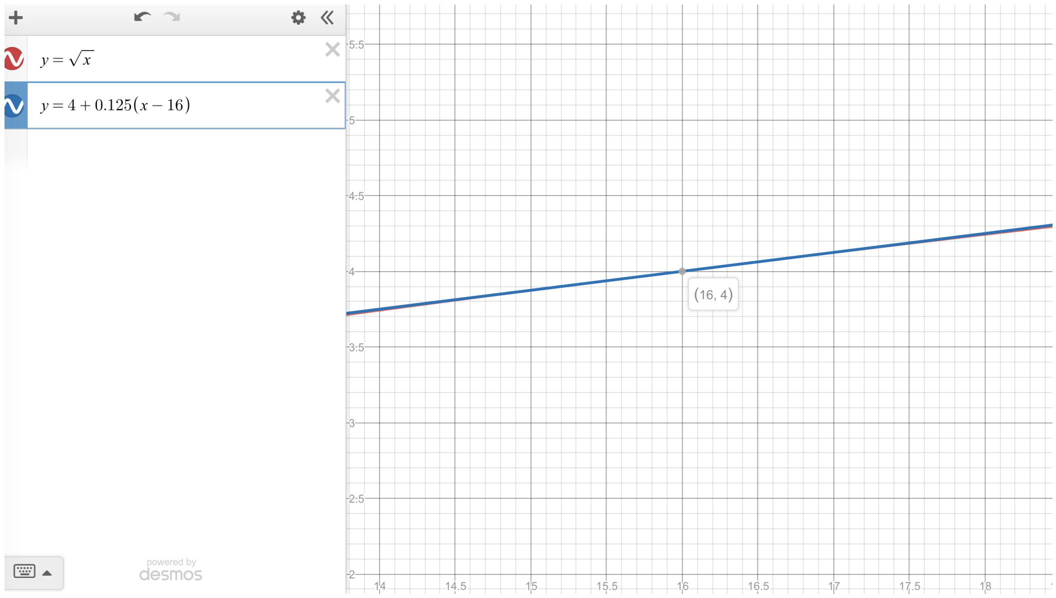
Notice that the linearization of f at a agrees with f at a: $L_{f,a}(x) = f(a)$. The intuition of the definition is that *near* a, $L_{f,a}(x)$ is a good approximation for $f(x)$, and is (often) much easier to calculate than $f(x)$.

The linearization is particularly useful if the point a is such that both $f(a)$ and $f'(a)$ are particularly "nice". Here's an example: consider $f(x) = \sqrt{x}$. It's not in general easy to calculate values of f, but there are some places where it *is* easy, namely at those x which are perfect squares of integers $(1, 4, 9, \ldots)$. So, take $a = 4$. We have $f(a) = f(16) = 4$, and, since $f'(x) = 1/(2\sqrt{x})$, we have $f'(a) = f'(16) = 1/8 = 0.125$. That means that the linearization of f at 16 is the function

$$L_{f,16}(x) = 4 + 0.125(x - 16).$$

Here are two pictures showing the graphs of both f and $L_{f,16}$, one from a wide view, and the other focussing in on what happens near the point $(16, 4)$. Notice that near to 16 on the x-axis, the two graphs are very close to each other; this is especially apparent from the second, close-up, picture, where is is almost impossible to tell the two graphs apart.





If we use $L_{f,16}$ to approximate $\sqrt{14}$, we get

$$\sqrt{14} = f(14) \approx L_{f,16}(14) = 4 + 0.125(14 - 16) = 3.75.$$

This is not too bad! A calculator suggests that $\sqrt{14} = 3.7416\cdots$, so the linearization gives an answer with an absolute error of around 0.0083, and a relative error of around 2.2%.

Of course, the situation won't always be so good: if we use $L_{f,16}$ to approximate $\sqrt{100}$, we get an estimate of $4+0.125(100-16) = 14.5$, which differs from the true value (10) by a large amount[115]; and if we use it to estimate $\sqrt{-8}$ we get an estimate of $4 + 0.125(-8 - 16) = 1$ for a quantity that doesn't exist!

This leads to the first of two natural questions to ask about the linearization (the second, you are probably already asking yourself):

- *How good is the linearization as an approximation tool, precisely?*: It's easy to approximate *any* function, at *any* possible input: just say "7". An approximation is only useful if it comes with some guarantee of its accuracy, such as "$\sqrt{14}$ is approximately 3.75; and this estimate is accurate to error ± 0.2", meaning that "$\sqrt{14}$ is certain to lie in the interval $(3.55, 3.95)$". The linearization does come with a guarantee of accuracy, but we will not explore it until next semester, when we consider the much more general (and powerful) Taylor polynomial.

- *Why use a scheme like this, to estimate the values of complicated functions, when we could just use a calculator?*: To answer this, ask another question: how does a *calculator* figure out the values of complicated functions?!?

Here's a theoretical justification for the linearization as a tool for approximating the values of a function, near the point around which we are linearizing: it's certainly the case

[115]Not too surprising, since by most measures 100 is *not* close to 16.

that
$$\lim_{x\to a}(f(x)-L_{f,a}(x)) = \lim_{x\to a} f(x) - \lim_{x\to a} L_{f,a}(x) = f(a)-f(a) = 0,$$
which says that as x approaches a, the linearization gets closer and closer to f (makes smaller and smaller error). But this is true of lots and lots of candidates for a simple approximating function; in particular it's true about the constant function $f(a)$, but something as naive as that can hardly be considered as a good tool for approximating the function f away from a (it takes into account nothing except the value of the function at a). The linearization takes a little more into account about the function; it consider the direction in which the graph of the function is moving, at the point $(a, f(a))$. As a consequence of this extra data being built into the linearization, we have the following fact:
$$\begin{aligned}
\lim_{x\to a}\frac{f(x)-L_{f,a}(x)}{x-a} &= \lim_{x\to a}\frac{f(x)-f(a)-f'(a)(x-a)}{x-a} \\
&= \lim_{x\to a}\left(\frac{f(x)-f(a)}{x-a} - f'(a)\right) \\
&= \lim_{x\to a}\frac{f(x)-f(a)}{x-a} - \lim_{x\to a} f'(a) \\
&= f'(a)-f'(a) \\
&= 0.
\end{aligned}$$
In other words, not only does the value of the linearization get closer and closer to the value of f as x approaches a, but also

> the linearization get closer and closer to f as x approaches a, even when the error is measured relative to $x-a$

(a stronger statement, since $x-a$ is getting smaller as x approaches a).[116]

8.4 The derivative of sin

Here we go through an informal calculation of the derivative of the sin function. It is informal, because we have only informally defined sin. Next semester, we will give a proper definition of sin (via an integral), from which all of its basic properties will emerge quite easily.

Along the way, we will derive the important and non-obvious trigonometric limit
$$\lim_{h\to 0}\frac{\sin h}{h} =?.$$
Because we haven't yet rigorously defined sin, the treatment here will be quite casual and intuitive. But at least it will give a sense of the behavior of the trigonometric functions vis a vis the derivative, and allow us to add sin and cos to the army of functions that we can differentiate.

Recall how we (informally, geometrically) defined the trigonometric functions sin and cos:

[116]The linearization is actually the *unique* linear function with this property. We'll have much more to say about this next semester, when we look at Taylor series.

If P is a point on the unit circle $x^2 + y^2 = 1$, that is a distance θ from $(1, 0)$, measured along the circle in a counterclockwise direction (starting from P), then the x-coordinate of P is $\cos\theta$, and the y-coordinate is $\sin\theta$.

It's typical to refer to the angle made at $(0, 0)$, in going from P to $(0, 0)$ to $(1, 0)$, as θ; see the picture below.

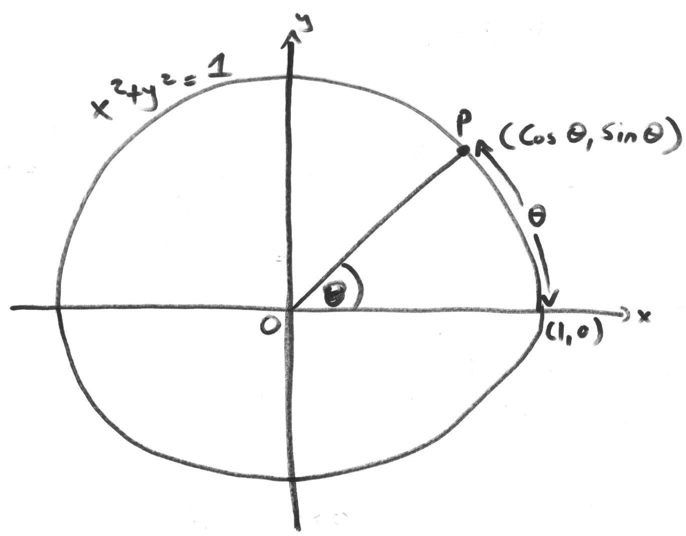


FIGURE 1 → DEF. OF SIN, COS.

And once we have said what "angle" means, it is easy to see (by looking at ratios of side-lengths of similar triangles) that this definition of sin and cos coincides with the other geometric definition you've seen: if triangle ABC has a right angle at B, and an angle θ at A, then

$$\sin\theta = \frac{\mathrm{BC}}{\mathrm{AC}} = \frac{\text{opposite}}{\text{hypothenuse}}, \qquad \cos\theta = \frac{\mathrm{BA}}{\mathrm{AC}} = \frac{\text{adjacent}}{\text{hypothenuse}}.$$

What is the derivative of sin? By definition, $\sin'\theta$ is

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

There's no obvious algebraic manipulation we can do here to make this limit easy to calculate. We need the *sine sum formula*:

for any angles α, β, $\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$.

Here is a picture, that leads to a proof of this formula (square brackets indicate right angles):

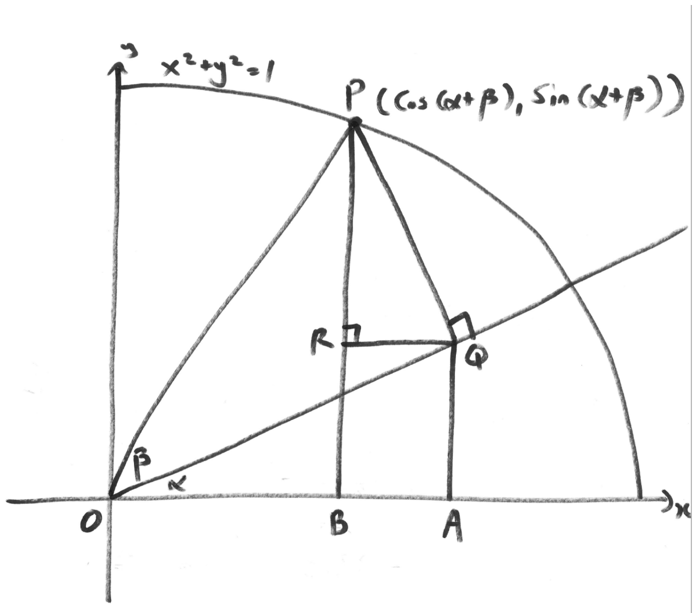


FIGURE 2 → SIN SUM FORMULA

Question 1: Why does this prove

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta?$$

Answer:

- First argue that angle RPQ is α (look first at OQA, then RQO, then RQP, then RPQ).
- Argue (from the definition of sin, and similar triangles) that PQ is $\sin\beta$, and so PR is $\cos\alpha\sin\beta$.
- Argue similarly that OQ is $\cos\beta$, and so AQ is $\sin\alpha\cos\beta$.
- Since AQ is the same as RB, and since PB is known, conclude that

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

Of course, this geometric proof only works for $\alpha, \beta \geq 0$, $\alpha+\beta \leq \pi/2$; but similar pictures can be drawn for all other cases.[117]

Now using the sin sum formula, we have (throughout assuming that all the various limits

[117] Here is another picture that justifies the trigonometric sum formulae, due to Tamás Görbe:

in fact exist):

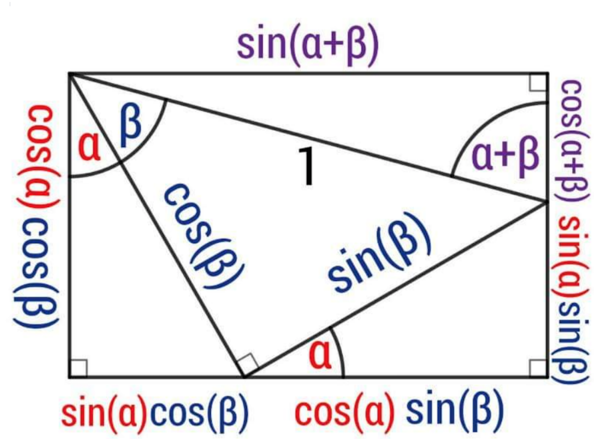
$$\begin{aligned}
\sin'\theta &= \lim_{h\to 0}\frac{\sin(\theta+h)-\sin\theta}{h}\\
&= \lim_{h\to 0}\frac{\sin\theta\cos h+\cos\theta\sin h-\sin\theta}{h}\\
&= \lim_{h\to 0}\frac{\sin\theta(\cos h-1)+\cos\theta\sin h}{h}\\
&= \lim_{h\to 0}\frac{\sin\theta(\cos h-1)}{h}+\lim_{h\to 0}\frac{\cos\theta\sin h}{h}\\
&= \sin\theta\lim_{h\to 0}\frac{\cos h-1}{h}+\cos\theta\lim_{h\to 0}\frac{\sin h}{h}.
\end{aligned}$$

We have reduced to two limits, neither of which look any easier than the one we started with! But, it turns out they they are essentially the same limit:

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

In the second from last line we used continuity of sin and cos, and in the last line, we used the (as yet unjustified) fact that $(\sin h)/h$ actually tends to a limit, as h nears 0. On this assumption, we get

$$\sin'\theta = \cos\theta\lim_{h\to 0}\frac{\sin h}{h}.$$



So now we have one limit left to consider, and it is a *little* bit simpler than the limit we started with.

We now claim that

$$\lim_{h\to 0} \frac{\sin h}{h} = 1.$$

Here is a picture, that leads to a proof of this claim:

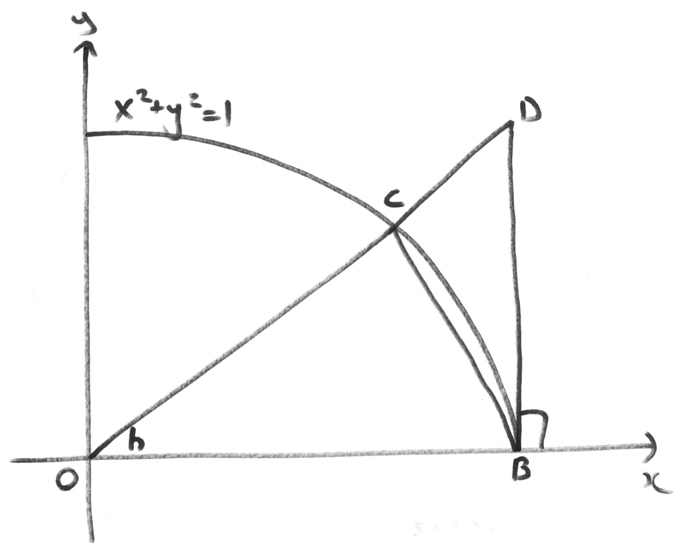


FIGURE 3 → LIMIT OF $\frac{\sin h}{h}$

Question 2: Why does this prove

$$\lim_{h\to 0} \frac{\sin h}{h} = 1?$$

Answer:

- What is the area of the triangle OBC?
- What is the area of the wedge of the circle between OC and OB?
- What is the area of the triangle OBD?
- What is the inequality relation between these three areas?
- Conclude that
$$\frac{\sin h}{2} \le \frac{h}{2} \le \frac{\sin h}{2\cos h}.$$
- Conclude that
$$\cos h \le \frac{\sin h}{h} \le 1.$$

- Use continuity of cos, and the squeeze theorem, to get the result.

Of course, the picture only shows that $\lim_{h\to 0^+}(\sin h)/h = 1$; but a similar picture gives the other one-sided limit.

We now get to conclude that
$$\sin'\theta = \cos\theta.$$

What about the derivative of cos? We could play the same geometric game to derive

$$\cos'\theta = -\sin\theta;$$

after we've seen the chain rule, we'll give an alternate derivation.

8.5 Some more theoretical properties of the derivative

In this section, rather than looking at specific examples to bring out properties of the derivative, we derive some more general properties (that, incidentally, will allow us to discover many more specific examples of derivatives).

We have observed intuitively that differentiability at a point is a more stringent "smoothness" condition that simple continuity; in other words, it's possible for a function to be continuous at a point but not differentiable there ($f(x) = |x|$ at $x = 0$ is an example), but it should not be possible for a function to be differentiable at a point without first being continuous. We'll now turn this intuition into a proven fact.

The hard way to do this is to start with a function which is defined at some point a, but not continuous there, and then argue that it cannot be differentiable at that point (the non-existence of the continuity limit somehow implying the non-existence of the differentiability limit). This is the hard way, because there are many different ways that a function can fail to be continuous, and we would have to deal with all of them in this approach.

The soft way is to go via the contrapositive, and prove that if f is differentiable at a point, then it most also be continuous there (the existence of the differentiability limit somehow implying the existence of the continuity limit). This is easier, because there's only one way for a limit to exist; and it immediately implies that failure of continuity implies failure of differentiability.

Claim 8.1. *Suppose that f is defined at and near a, and is differentiable at a. Then f is continuous at a.*

Proof: Since f is differentiable at a, we have that for some real number a,

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

But also, $\lim_{h\to 0} h = 0$. By the product part of the sum/product/reciprocal theorem for limits, we can conclude that

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

so (by the sum part of the sum/product/reciprocal theorem for limits)

$$\lim_{h\to 0} f(a+h) = f(a),$$

which says that f is continuous at a.[118] □

The take-away from this is:

continuity is necessarily for differentiability, but not sufficient.

We now derive some identities that will allow us to easily compute some new derivatives from old ones.

Claim 8.2. *Suppose f and g are functions that are both differentiable at some number a, and that c is some real constant. Then both $f+g$ and cf are differentiable at a, with*

$$(f+g)'(a) = f'(a) + g'(a)$$

and

$$(cf)'(a) = cf'(a).$$

Proof: Both statements follow quickly from previously established facts about limits. We have that

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

and

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

and so

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

[118] This is not exactly the definition of continuity at a; but you can prove that this is an equivalent definition, just as we proved earlier that $\lim_{b\to a}(f(b)-f(a))/(b-a)$ is the same as $\lim_{h\to 0}(f(a+h)-f(a))/h$.

which exactly says that $f+g$ is differentiable at a, with derivative $f'(a)+g'(a)$.

Similarly,

$$\begin{aligned} cf'(a) &= c\lim_{h\to 0}\frac{f(a+h)-f(a)}{h} \\ &= \lim_{h\to 0}\frac{c(f(a+h)-f(a))}{h} \\ &= \lim_{h\to 0}\frac{(cf)(a+h)-(cf)(a)}{h} \end{aligned}$$

which exactly says that cf is differentiable at a, with derivative $cf'(a)$. □

Note that the second part of the above claim should alert us to an important insight: we should *not* expect (as we might have, by analogy with similar properties for limits) that the derivative of the product of a pair of functions is the product of the derivatives. If that *were* the case, then, since the derivate of the constant function c is 0, we would have that the derivative of cf at any point is also 0.

Another reason we should not expect that the derivative of the product of a pair of functions is the product of the derivatives is through a dimension analysis. Suppose that $f(x)$ is measuring the height (measured in meters) of a square at time x (measured in seconds), and that $g(x)$ is measuring the width of the square. Then $(fg)(a)$ is measuring the area of the square (measured in meter squared) at time a. Now, the derivative of a function at a can be thought of as measuring the *instantaneous* rate at which the function is changing at a, as the input variable changes — indeed, for any $h\neq 0$ the quantity $(f(a+h)-f(a))/h$ is measuring the average change of f over the time interval $[a,a+h]$ (or $[a+h,a]$, if $h<0$), so if the limit of this ratio exists at h approaches 0, it makes sense to declare that limit to be the instantaneous rate of change at a.

So $(fg)'(a)$ is measuring the rate at which the area of the square is changing, at time a. This is measured in meters squared per second . But $f'(a)g'(a)$, being the product of two rates of changes of lengths, is measured in meters squared per second *squared*. The conclusion is that $(fg)'(a)$ and $f'(x)g'(a)$ have different dimensions, so we should not expect them to be equal in general.

What *should* we expect $(fg)'(a)$ to be? The linearizations of f and g provide a hint. For x near a, we have

$$f(x)\approx L_{f,a}(x)=f(a)+f'(a)(x-a)$$

and

$$g(x)\approx L_{f,a}(x)=g(a)+g'(a)(x-a).$$

Using the linearizations to approximate f and g at $a+h$ we get

$$\begin{aligned} (fg)(a+h) &= f(a+h)g(a+h) \\ &\approx (f(a)+f'(a)h)(g(a)+g'(a)h) \\ &= f(a)g(a)+f'(a)g(a)h+f(a)g'(a)h+f'(a)g'(a)h^2 \end{aligned}$$

and so

$$\frac{(fg)(a+h)-(fg)(a)}{h} \approx f'(a)g(a)+f(a)g'(a)+f'(a)g'(a)h.$$

Considering what happens as $h \to 0$, this strongly suggests that $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$, and this is indeed the case.

Claim 8.3. *(Product rule for differentiation) Suppose f and g are functions that are both differentiable at some number a. Then both fg is differentiable at a, with*

$$(fg)'(a) = f'(a)g(a) + f(a)g'(a).$$

Proof: The proof formalizes the intuition presented above. We begin by assuming that fg is indeed differentiable at a, and try to calculate its derivative.

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

Going backwards through the chain of equalities we see that all manipulations with limits are justified, by

- repeated applications of the sum/product/reciprocal theorem for limits,
- the differentiability of f and g at a, and
- the continuity of f at a (needed for $\lim_{h\to 0} f(a+h) = f(a)$), which itself follows from the differentiability of f at a.

$\square$

As a first application of the product rule, we re-derive the fact that if $f_n : \mathbb{R} \to \mathbb{R}$ ($n \in \mathbb{N}$) is given by $f_n(x) = x^n$, then f_n' (the derivative, viewed as a function) has domain $\mathbb{R}$ and is given by $f_n'(x) = nx^{n-1}$. We prove this by induction on n; we've already done the base case $n = 1$. For $n \geq 2$ we have that $f_n = f_1 f_{n-1}$, so, for each real a, by the product rule (applicable by the induction hypothesis: f_{n-1} and f_1 are both differentiable at a) we have

$$\begin{aligned} f_n'(a) &= (f_1 f_{n-1})'(a) \\ &= f_1'(a)f_{n-1}(a) + f_1(a)f_{n-1}'(a) \quad \text{(product rule)} \\ &= 1 \cdot a^{n-1} + a \cdot (n-1)a^{n-2} \quad \text{(induction)} \\ &= a^{n-1} + (n-1)a^{n-1} \\ &= na^{n-1}, \end{aligned}$$

which completes the induction step.

We extend this now to negative exponents, giving the presentation in a slightly more streamlined way. For $n \in \mathbb{N}$, define g_n by $g_n(x) = 1/x^n$ (so the domain of g_n is $\mathbb{R} \setminus \{0\}$). We claim that

$$g_n'(x) = \frac{-n}{x^{n+1}}$$

(again with domain $\mathbb{R} \setminus \{0\}$). We prove this by induction on n. The base case $n = 1$ claims that if g_1 is defined by $g_1(x) = 1/x$ then for all non-zero x g_1 is differentiable with derivative $-1/x^2$. This is left as an exercise — it's very similar to a previous example.

For the induction step, assume that $g_{n-1}' = -(n-1)/x^n$. We have

$$\begin{aligned} g_n'(x) &= (g_1 g_{n-1})'(x) \\ &= g_1'(x)g_{n-1}(x) + g_1(x)g_{n-1}'(x) \quad \text{(product rule)} \\ &= \frac{(-1)(1)}{(x^2)(x^{n-1})} + \frac{(1)(-(n-1))}{(x)(x^n)} \quad \text{(induction)} \\ &= \frac{-n}{x^{n+1}}, \end{aligned}$$

which completes the induction step.

Both the sum rule for derivatives and the product rule extend to sums and products of multiple functions. For sums, the conclusion is obvious:

if $f_1, f_2, \ldots, f_n$ are all differentiable at a, then so is $\sum_{k=1}^n f_k$, and

$$\left(\sum_{k=1}^n f_k\right)'(a) = \sum_{k=1}^n f_k'(a).$$

The proof is an easy induction on n, as it is left as an exercise.

For products, the conclusion is less obvious. But once we apply the product rule multiple times to compute the derivative of the product of *three* functions, a fairly clear candidate conclusion emerges. We have (dropping reference to the particular point a, to keep the notation readable)

$$(fgh)' = ((fg)h)' = (fg)'h + (fg)h' = ((f'g) + (fg'))h + (fg)h' = f'gh + fg'h + fgh'.$$

This suggests:

Claim 8.4. *(Product rule for product of n functions)*[119] *If $f_1, f_2, \ldots, f_n$ are all differentiable at a, then so is $\prod_{k=1}^n f_k$, and*

$$\left(\prod_{k=1}^n f_k\right)'(a) = \sum_{k=1}^n f_1(a)f_2(a)\cdots f_{k-1}(a)f_k'(a)f_{k+1}(a)\cdots f_{n-1}(a)f_n(a).$$

Proof: Strong induction on n, with $n = 1$ trivial and $n = 2$ being the product rule. For $n > 2$, write $f_1f_2\cdots f_n$ as $(f_1\cdots f_{n-1})(f_n)$, and apply the product rule to get

$$(f_1f_2\cdots f_n)'(a) = (f_1\cdots f_{n-1})'(a)f_n(a) + (f_1\cdots f_{n-1})(a)f_n'(a).$$

The second term here, $(f_1\cdots f_{n-1})(a)f_n'(a)$, gives the term corresponding to $k = n$ in $\sum_{k=1}^n f_1(a)\cdots f_k'(a)\cdots f_n(a)$. By induction the first term is

$$\begin{aligned}(f_1\cdots f_{n-1})'(a)f_n(a) &= \left(\sum_{k=1}^{n-1} f_1(a)\cdots f_k'(a)\cdots f_{n-1}(a)\right)f_n(a)\\ &= \sum_{k=1}^{n-1} f_1(a)\cdots f_k'(a)\cdots f_{n-1}(a)f_n(a),\end{aligned}$$

and this gives the remaining terms ($k = 1, \ldots, n-1$) in $\sum_{k=1}^n f_1(a)\cdots f_k'(a)\cdots f_n(a)$. This completes the induction. □

As application of this generalized product rule, we give a cute derivation of

$$h_n'(x) = \frac{1}{n}x^{\frac{1}{n}-1}$$

where $h_n(x) = x^{1/n}$, $n \in \mathbb{N}$. Recalling the previously used noatation $f_1(x) = x$, we have

$$h_n(x)h_n(x)\cdots h_n(x) = f_1(x),$$

where there are n terms in the product. Differentiating both sides, using the product rule for the left-hand side — and noting that, since all the terms in the product are the same, it follows that all n terms in the sum that determines the derivative of the product are the same — we get

$$nh_n'(x)h_n(x)^{n-1} = 1,$$

which, after a little algebra, translates to

$$h_n'(x) = \frac{1}{n}x^{\frac{1}{n}-1}.$$

[119] In words: the derivative of a product of n functions is the derivative of the first times the product of the rest, plus the derivative of the second times the product of the rest, and so on, up to plus derivative of the last times the product of the rest.

Although this is cute, it's a little flawed — we assumed that $h_n'(x)$ exists. So essentially what we have done here is argued that *if* the nth root function is differentiable *then* its derivative must be what we expect it to be. To actually verify that the root function is differentiable, we need to go back to the definition, as we did with the square root function.

Another way to generalize the product rule is to consider higher derivatives of the product of two functions. We have

$$(fg)^{(0)} = fg = f^{(0)}g^{(0)},$$
$$(fg)^{(1)} = (fg)' = fg' + f'g = f^{(0)}g^{(1)} + f^{(1)}g^{(0)},$$

and

$$(fg)^{(2)} = (fg)'' = (fg' + f'g)' = fg'' + 2f'g' + f''g'' = f^{(0)}g^{(2)} + 2f^{(1)}g^{(1)} + f^{(2)}g^{(0)}.$$

There seems to be a pattern here:

$$(fg)^{(n)} = \sum_{k=0}^{n} (\text{SOME COEFFICIENT DEPENDING ON } n \text{ and } k)\ f^{(k)}g^{(n-k)}.$$

A homework problem asks you to find the specific pattern, and prove that is correct for all $n \geq 0$.

After the product rule, comes the quotient rule. We work up to that by doing the reciprocal rule first.

Claim 8.5. *(Reciprocal rule for differentiation) Suppose g is differentiable at some number a. If $g(a) \neq 0$ Then $(1/g)$ is differentiable at a, with*

$$\left(\frac{1}{g}\right)'(a) = -\frac{g'(a)}{g^2(a)}.$$

Proof: Since g is differentiable at a, it is continuous at a, and since $g(a) \neq 0$, $g(x) \neq 0$ for all x in some interval around a. So $1/g$ is defined in an interval around a.

We have

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

where, as usual, going backwards through the chain of equalities we see that all manipulations with limits are justified (and all claimed limits exist), by

- repeated applications of the sum/product/reciprocal theorem for limits,
- the differentiability of g at a, and
- the continuity of g at a.

$\square$

The reciprocal rule allows an alternate derivation of the derivative of $g_n(x) = x^{-n}$ ($n \in \mathbb{N}$). Since $g_n(x) = 1/f_n(x)$ (where $f_n(x) = x^n$) and $f_n'(x) = nx^{n-1}$, we have

$$g_n'(x) = -\frac{nx^{n-1}}{x^{2n}} = -\frac{n}{x^{n+1}}.$$

The rule for differentiating the quotient of functions follows quickly from the reciprocal, by combining it with the product rule:

Claim 8.6. *(Quotient rule for differentiation) Suppose f and g are differentiable at some number a. If $g(a) \neq 0$ then (f/g) is differentiable at a, with*

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g^2(a)}.$$

Proof: We view f/g as $(f)(1/g)$, and apply product and quotient rules to get

$$\begin{aligned}\left(\frac{f}{g}\right)'(a) &= f'(a)\left(\frac{1}{g(a)}\right) + f(a)\left(\frac{1}{g}\right)'(a)\\ &= \frac{f'(a)}{g(a)} - \frac{f(a)g'(a)}{g^2(a)}\\ &= \frac{f'(a)g(a) - f(a)g'(a)}{g^2(a)}.\end{aligned}$$

$\square$

With all of these rules, we can easily differentiate any rational function, and all root functions, but we cannot yet differentiate a function like $f(x) = \sqrt{x^2+1}$, or like $f(x) = \sin^2 x$ (unless we go back to the definition, which is nasty). What we need next is rule saying what happens when we try to differentiate the *composition* of two known functions. This rule is probably the most important one in differential calculus, so we give it its own section.

8.6 The chain rule

Suppose that g is differentiable at a, and that f is differentiable at $g(a)$. We would expect that $f \circ g$, the composition of f and g, should be differentiable at a, but what should the derivative be? To get an intuition, we do what we did before deriving the product rule, and consider the linearizations of g and f near a and $g(a)$, respectively. We have, for any number a at which g is differentiable,

$$g(a+h) - g(a) \approx g'(a)h, \ (\star)$$

and for any number A at which f is differentiable,

$$f(A+k) - f(A) \approx f'(A)k, \ (\star\star)$$

both approximations presumably reasonable when h and k are small (and in particular, getting better and better as h and k get smaller). So

$$\begin{aligned}
\frac{(f \circ g)(a+h) - (f \circ g)(a)}{h} &= \frac{f(g(a+h)) - f(g(a))}{h} \\
&\approx \frac{f(g(a) + g'(a)h) - f(g(a))}{h} \text{ (applying } (\star)) \\
&\approx \frac{f'(g(a))g'(a)h}{h} \text{ (applying } (\star\star) \text{ with } A = g(a) \text{ and } k = g'(a)h) \\
&= f'(g(a))g'(a).
\end{aligned}$$

This suggest an answer to the question "what is the derivative of a composition?", and it turns out to be the correct answer.

Claim 8.7. *(Chain rule for differentiation) Suppose that g is differentiable at a, and that f is differentiable at $g(a)$. Then $f \circ g$ is differentiable at a, and*

$$(f \circ g)'(a) = f'(g(a))g'(a).$$

A word of warning: $f'(a)$ means the derivative of f at input a; so $f'(g(a))$ is the derivative of f evaluated at $g(a)$, **NOT** the derivative of (f composed with g) evaluated at a (that's $(f \circ g)'(a)$). These two things — $f'(g(a))$ and $(f \circ g)'(a)$ — are usually different. Indeed,

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

while

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

Usually $g(a) + h \neq g(a+h)$ (consider, for example, $g(x) = x^3$: we have

$$g(a) + h = a^3 + h \neq (a+h)^3 = g(a+h)).$$

There is one exception: when $g(x) = x + c$, for some constant c, we have

$$g(a) + h = (a+c) + h = (a+h) + c = g(a+h).$$

One upshot of this is that if $h(x) = f(x+c)$ then $h'(x) = f'(x+c)$. But in general if a function h is defined as the composition $f \circ g$ (here g was $g(x) = x + c$), you need to use the chain rule to evaluate the derivative of h.

Before giving the proof of the chain rule, we present some examples.

- $h_1(x) = \sin^2 x$. This is the composition $h = \text{square} \circ \sin$, where $\text{square}(x) = x^2$. By the chain rule
$$h_1'(x) = \text{square}'(\sin(x))\sin'(x) = 2\sin x \cos x.$$

- $h_2(x) = \sin(x^2)$. This is the composition $h = \sin \circ \text{square}$. By the chain rule
$$h_2'(x) = \sin'(\text{square}(x))\square'(x) = (\cos(x^2))2x = 2x\cos x^2.$$

  Notice that $h_1'(x) \neq h_2'(x)$ (in general); but since composition is not commutative, there is no particular reason to expect that these two functions h_1, h_2 would end up having the same derivative.

- $f(x) = 1/x^n$. We can view this as the composition "reciprocal after nth power, and find, via chain rule (and the fact that we have already computed the derivatives of both the reciprocal function and the nth power function), that
$$f'(x) = \frac{-1}{(x^n)^2} n x^{n-1} = \frac{-n}{x^{n+1}}.$$
  Or, we can view f as the composition "nth power after reciprocal" to get
$$f'(x) = n(1/x)^{n-1} \cdot \frac{-1}{x^2} = \frac{-n}{x^{n+1}}.$$
  Either way we get the same answer.

- The derivative of $\cos x$. We have observed that the derivative of cos can be obtained by geometric arguments in a manner similar to the way we derived the derivative of sin. Another approach is to consider the equation $\sin^2 x + \cos^2 x = 1$. The right- and left-hand sides here are both functions of f, so both can be differentiated as functions of x. Using the chain rule for the right-hand side, we get
$$2\sin x \cos x + 2\cos x \cos' x = 0$$
  or, dividing across by $\cos x$[120],
$$\cos' x = -\sin x$$
  (as expected).

- Composition of three functions. Consider $f(x) = (\sin x^3)^2$. This is the composition of squaring (on the outside), sin (in the middle), cubing (inside), so $f_1 \circ f_2 \circ f_3$, where f_1 is the square function, f_2 the sin function, and f_3 the cube function. The chain rule says
$$\begin{aligned}(f_1 \circ f_2 \circ f_3)'(a) &= (f_1 \circ (f_2 \circ f_3))'(a) \\ &= f_1'((f_2 \circ f_3)(a))((f_2 \circ f_3)'(a)) \\ &= f_1'((f_2 \circ f_3)(a)) f_2'(f_3(a)) f_3'(a) \\ &= f_1'(f_2(f_3(a))) f_2'(f_3(a)) f_3'(a).\end{aligned}$$

[120]This is a little informal; it really only works in some interval around 0 where $\cos x \neq 0$. But that's ok; we only have informal definitions of sin and cos to start with. We will shore all this up next semester.

So we get

$$f'(x) = 2(\sin x^3)(\cos x^3)3x^2 = 6x^6(\sin x^3)(\cos x^3).$$

The chain rule pattern (what happens for the derivative of a composition of four, or five, or six, or more, functions, should be fairly clear from this example. In applying the chain rule on a complex composition, you should get used to "working from the outside in".

We now present a formalization of the heuristic argument given above for the chain rule. We will give *two* approaches; the second is perhaps the easier. These are somewhat different justification to the one presented by Spivak.

First proof of chain rule: Note that given our definition of differentiation, saying that g is differentiable at a automatically says that it is defined in some interval (b, c) with $b < a < c$, and saying that f is differentiable at $g(a)$ automatically says that it is defined in some interval (b', c') with $b' < g(a) < c'$.

We start the proof by observing that since g is differentiable at a, we have (for some number $g'(a)$)

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

which says that as h approaches 0, the expression

$$\frac{g(a+h) - g(a)}{h} - g'(a)$$

approaches 0. Denoting this expression by $\varepsilon(h)$ (a function of h, named ε), we get that

$$g(a+h) - g(a) = g'(a)h + \varepsilon(h)h \tag{3}$$

where $\varepsilon(h) \to 0$ as $h \to 0$. The function $\varepsilon(h)$ is defined *near* 0, but not *at* 0; however, the fact that $\varepsilon(h) \to 0$ as $h \to 0$ means that if we extend ε by declaring $\varepsilon(0) = 0$ then not only is ε defined at 0, but it is continuous at 0.

Similarly

$$f(g(a) + k) - f(g(a)) = f'(g(a))k + \eta(k)k \tag{4}$$

where $\eta(k) \to 0$ as $k \to 0$; as before, we extend η to a function that is continuous at 0 by declaring $\eta(0) = 0$. Notice that (4) remains true at $k = 0$.

We now study the expression $f(g(a+h)) - f(g(a))$. Applying (3) we get

$$f(g(a+h)) - f(g(a)) = f(g(a) + g'(a)h + \varepsilon(h)h) - f(g(a)). \tag{5}$$

Now, for notational convenience, set $k = g'(a)h + \varepsilon(h)h$ (notice that this depends on h, so we really should think of k as $k(h)$; but to keep the notation manageable, we will mostly just write k). Applying (4) to (5) we get

$$f(g(a+h)) - f(g(a)) = f(g(a)) + f'(g(a))k + \eta(k)k - f(g(a)) = (f'(g(a)) + \eta(k))k.$$

Now notice that k consists of two terms both of which are multiples of h, so we may divide through by h to obtain

$$\begin{aligned}\frac{f(g(a+h))-f(g(a))}{h} &= (f'(g(a))+\eta(k))(g'(a)+\varepsilon(h)) \\ &= f'(g(a))g'(a)+f'(g(a))\varepsilon(h)+\eta(k)(g'(a)+\varepsilon(h)).\end{aligned}$$

As h approaches 0, it is certainly that case that $f'(g(a))\varepsilon(h)$ approaches 0, since $\varepsilon(h)$ does. If we could show that $\eta(k) \to 0$ as $h \to 0$, then we would also have $\eta(k)(g'(a)+\varepsilon(h)) \to 0$ as $h \to 0$, and so we could conclude that

$$\frac{f(g(a+h))-f(g(a))}{h} \to f'(g(a))g'(a) \quad \text{as} \quad h \to 0,$$

which is exactly what the chain rule asserts.

It seems clear that $\eta(k) \to 0$ as $h \to 0$, since we know that $\eta(k)$ approaches 0 as the argument k approaches 0, and we can see from the equation $k = g'(a)h + \varepsilon(h)h$ that k approaches 0 as h approaches 0. Making this precise requires an argument very similar to the one that we used to show that the composition of continuous functions is continuous.

Let $\varepsilon > 0$ be given (this has nothing to do with the function $\varepsilon(h)$ introduced earlier). Since $\eta(x) \to 0$ as $x \to 0$, there is a $\delta > 0$ such that $0 < |x| < \delta$ implies $|\eta(x)| < \varepsilon$. But in fact, since η is continuous at 0 (and takes the value 0) we can say more: we can say that $|x| < \delta$ implies $|\eta(x)| < \varepsilon$. (In a moment we'll see why this minor detail is important).

Now consider $k = k(h) = g'(a)h + \varepsilon(h)h$. As observed earlier, $k(h)$ approaches 0 as h approaches 0, so, using the definition of limits, there is a $\delta' > 0$ such that $0 < |h| < \delta'$ implies that $|k(h)| < \delta$ (the same δ from the last paragraph). From the last paragraph we conclude that $0 < |h| < \delta'$ in turn implies $|\eta(k(h))| < \varepsilon$, and since $\varepsilon > 0$ was arbitrary, this shows that $\eta(k) \to 0$ as $h \to 0$, finishing the proof of the chain rule.

Notice that if we only knew that $0 < |x| < \delta$ implies $|\eta(x)| < \varepsilon$ (i.e., if we didn't have continuity of η at 0), then knowing that $0 < |h| < \delta'$ implies that $|k(h)| < \delta$ would allow us to conclude *nothing* — for those h for which $k(h) = 0$ (for which $g'(a) = -\varepsilon(h)$), we would be unable to run this argument.

Second proof of chain rule: We begin with a preliminary observation about the linearization of a function. Suppose that a function f is differentiable at a. Then we can write, for all h,

$$f(a+h) = L_{f,a}(a+h) + \mathrm{err}_{f,a}(h),$$

where $\mathrm{err}_{f,a}(h)$, the error in using $L_{f,a}(a+h)$ to estimate $f(a+h)$, is defined by

$$\mathrm{err}_{f,a}(h) := f(a+h) - f(a) - f'(a)h.$$

Notice that (by continuity of f at a) we have that $\mathrm{err}_{f,a}$ approaches limit 0 near 0, i.e.,

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

This is no surprise — it's just saying that the linearization of f agrees with f at a.

But more is true. We have that

$$\frac{\mathrm{err}_{f,a}(h)}{h} = \frac{f(a+h)-f(a)}{h} - f'(a) \to f'(a) - f'(a) = 0 \ \text{ as } h \to 0 \quad (\star)$$

(the last step valid because f is not just continuous but *differentiable* at a). In other words,

> the error we make in using $L_{f,a}(a+h)$ to approximate $f(a+h)$ goes to zero as h approaches zero, even when scaled relative to h.

We will return to this idea next semester, when discussing Taylor series.

Armed with $(\star)$, we can formally prove the chain rule, without using a ε-δ proof. Recall that we are assuming that g is differentiable at a and that f is differentiable at $g(a)$. Using first the linearization of g at a (with error $\mathrm{err}_{g,a}$) and then the linearization of f at $g(a)$ (with error $\mathrm{err}_{f,g(a)}$), we have that $\frac{f(g(a+h))-f(g(a))}{h}$

$$\begin{aligned}
&= \frac{f(g(a)+hg'(a)+\mathrm{err}_{g,a}(h))-f(g(a))}{h} \\
&= \frac{f(g(a))+(hg'(a)+\mathrm{err}_{g,a}(h))f'(g(a))+\mathrm{err}_{f,g(a)}(hg'(a)+\mathrm{err}_{g,a}(h))-f(g(a))}{h} \\
&= f'(g(a))g'(a) + \frac{\mathrm{err}_{g,a}(h)}{h}f'(g(a)) + \frac{\mathrm{err}_{f,g(a)}(hg'(a)+\mathrm{err}_{g,a}(h))}{h}. \quad (\star\star)
\end{aligned}$$

We claim that $(\star\star) \to f'(g(a))g'(a)$ as $h \to 0$; if we can show this then we have proved the chain rule. Indeed, by $(\star)$ we have that

$$\frac{\mathrm{err}_{g,a}(h)}{h} \to 0 \quad \text{as } h \to 0.$$

To deal with the last term in $(\star\star)$ (which we also want to tend to 0) we have to work a little harder. We have

$$\begin{aligned}
\frac{\mathrm{err}_{f,g(a)}(hg'(a)+\mathrm{err}_{g,a}(h))}{h} &= \left(\frac{\mathrm{err}_{f,g(a)}(hg'(a)+\mathrm{err}_{g,a}(h))}{h\left(g'(a)+\frac{\mathrm{err}_{g,a}(h)}{h}\right)}\right)\left(g'(a)+\frac{\mathrm{err}_{g,a}(h)}{h}\right) \\
&= \left(\frac{\mathrm{err}_{f,g(a)}(hg'(a)+\mathrm{err}_{g,a}(h))}{hg'(a)+\mathrm{err}_{g,a}(h)}\right)\left(g'(a)+\frac{\mathrm{err}_{g,a}(h)}{h}\right)
\end{aligned}$$

The second term on the right-hand side goes to $g'(a)$ as $h \to 0$, by $(\star)$. The first term can be viewed as a composition $(a \circ b)(h)$ where

- $b(h) = hg'(a) + \mathrm{err}_{g,a}(h)$ — a function which is defined and continuous everywhere
- $a(x) = \frac{\mathrm{err}_{f,g(a)}(x)}{x}$ — a function which is defined everywhere *except* $x = 0$, and is continuous everywhere it is defined.

However, from $(\star)$ we know that $\lim_{x\to 0} a(x) = 0$, so we can extend a to a function that is defined *and continuous* everywhere, by setting $a(0) = 0$.

We have previously proven that the composition of continuous functions is continuous, so $(a \circ b)$ is a function that is defined and continuous everywhere. In particular that means that

$$\lim_{h\to 0} \frac{\mathrm{err}_{f,g(a)}(hg'(a) + \mathrm{err}_{g,a}(h))}{hg'(a) + \mathrm{err}_{g,a}(h)} = \lim_{h\to 0}(a \circ b)(h) = (a \circ b)(0) = a(b(0)) = a(0) = 0,$$

and the proof of the chain rule is complete.