2 First Order Approximation: Derivatives and Linear Approximation

In section 1 we talked about approximating functions in the simplest possible way: if x is close to a then we approximate $f(x) \approx f(a)$. In this section we want to refine this estimate. How far apart are $f(x)$ and $f(a)$? How does the estimate get worse as x gets further from a?

2.1 Linear Approximation

In order to get a better estimate that the zero-order estimate of last section, we need to use a more complicated formula. But we want to keep the amount of complexity under control. So we want to use a simple function to approximate $f(x)$. The simplest possible function is a constant function; and that's exactly what we used last section. If a is a fixed number then $f(a)$ is a constant, and thus $f(x) \approx f(a)$ approximates f with a constant function.

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

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

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

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

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

The equation for our linear approximation is

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

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

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

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

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

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

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

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

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

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

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

2.2 The Derivative

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

So let's look at our formula from equation (1) again. We want to understand $f'(a)$, so we'll solve the equation for that:

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

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

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

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

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

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

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

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

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

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

Example 2.5. 1. Let $f(x) = x^2 + 1$. Then

$$f'(2) = \lim_{h\to 0} \frac{f(2+h) - f(2)}{h} = \lim_{h\to 0} \frac{(2+h)^2 + 1 - 2^2 - 1}{h} = \lim_{h\to 0} \frac{4h + h^2}{h} = 4,$$

and more generally, for any number a we have

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

2. Let $f(x) = x^3$, and let's find the derivative at a point a. Then

$$\begin{aligned} f'(a) &= \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lim_{x\to a} \frac{x^3 - a^3}{x - a} \\ &= \lim_{x\to a} \frac{(x-a)(x^2 + ax + a^2)}{x - a} = \lim_{x\to a} x^2 + ax + a^2 = 3a^2. \end{aligned}$$

Notice that it wasn't obvious that we could factor $x^3 - a^3$ this way. We could notice this by noticing that plugging in a gives us zero; in general, if plugging a into a polynomial gives zero, we can always factor out a $(x - a)$ term. In this case, though, it might have been easier to just start with the limit as $h \to 0$, in which case the problem would have essentially solved itself.

3. Let $f(x) = \sqrt{x}$. Then given a number a, we have

$$f'(a) = \lim_{h\to 0} \frac{\sqrt{a+h} - \sqrt{a}}{h} = \lim_{h\to 0} \frac{(a+h) - a}{h(\sqrt{a+h} + \sqrt{a})} = \lim_{h\to 0} \frac{1}{\sqrt{a+h} + \sqrt{a}} = \frac{1}{2\sqrt{a}}$$

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

4. Let $g(x) = \sqrt[3]{x}$. Then we can compute $g'(0)$ and we get

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

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

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

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

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

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

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

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

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

which is precisely waht it means to be continuous.

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

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

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

but

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

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

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

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

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

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

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

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

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

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

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

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

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

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

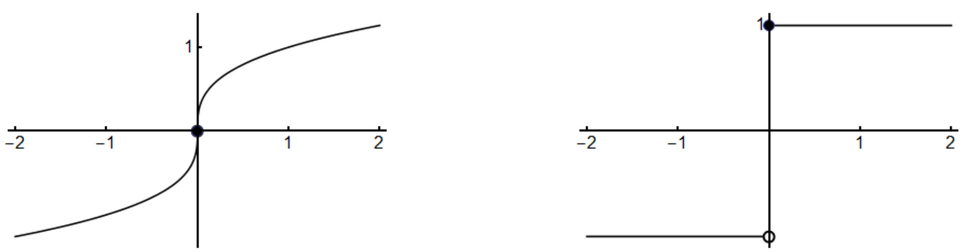

Figure 2.1: A vertical tangent line and a discontinuous function

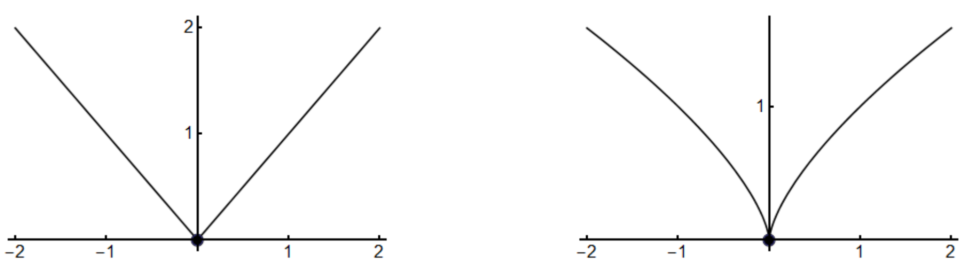

Figure 2.2: A corner and a cusp

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

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

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

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

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

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

Example 2.9.

1. If $f(x) = x^2 + 1$, we computed that $f'(x) = 2x$. The domain of f is all reals, and so is the domain of $f'(x)$.
2. If $g(x) = \sqrt{x}$ then $g'(x) = \frac{1}{2\sqrt{x}}$. The domain of g is all reals ≥ 0, and the domain of g' is all reals > 0.
3. We saw above that if $a(x) = |x|$, then

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

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

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

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

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

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

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

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

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

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

We won't say much more about the second derivative now, but we'll discuss it extensively in section 4.

2.3 Computing Derivatives

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

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

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

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

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

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

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

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

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

   Geometrically, this is a line with slope 1.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Example 2.19.

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

  Alternatively,

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

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

2.4 Trigonometric derivatives

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

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

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

Using this we can compute

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

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

Further using the product and quotient rules, we observe that

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

- $$(\cot(x))' = \left(\frac{\cos x}{\sin x}\right)' = \frac{-\sin^2(x)-\cos^2(x)}{\sin^2(x)} = \frac{-1}{\sin^2(x)} = -\csc^2(x)$$

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

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

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

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

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

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

3. IF $g(\theta) = \theta \sin\theta + \frac{\cos\theta}{\theta}$, then
$$g'(\theta) = (\sin\theta + \theta\cos\theta) + \frac{-\theta\sin\theta - \cos\theta}{\theta^2}.$$
4. If $h(x) = \frac{x}{2-\tan x}$, then
$$h'(x) = \frac{(2-\tan x) + x\sec^2 x}{(2-\tan x)^2}.$$
5. We can also compute second derivatives. $\sin'' x = -\sin x$. $\cos'' x = -\cos x$.
$$\tan'' x = (\sec x \sec x)' = \sec x \tan x \sec x + \sec x \tan x \sec x = 2\sec^2 x \tan x.$$

2.5 The Chain Rule

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

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

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

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

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

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

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

Proof.

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

□

Remark 2.22.

1. When we write $f'(g(x))$, we mean the function f' evaluated at the point $g(x)$, or in other words, the derivative of f at the point $g(x)$.
2. It can be helpful as a way of remembering the chain rule that
$$\frac{d(f \circ g)}{dx} = \frac{d(f \circ g)}{dg} \cdot \frac{dg}{x}.$$
Don't take this too seriously as actively meaning anything, since it only sort of does, but it's quite helpful for the memory.

Example 2.23.

1. $(x+1)^2 = (f \circ g)(x)$ where $g(x) = x + 1$ and $f(x) = x^2$. Then $f'(x) = 2x$ and $g'(x) = 1$, so
$$(f \circ g)'(x) = f'(g(x)) \cdot g'(x) = 2(g(x)) \cdot 1 = 2(x+1) \cdot 1 = 2x + 2.$$
Sanity check:
$$(f \circ g)'(x) = (x^2 + 2x + 1)' = 2x + 2.$$
2. $(x^2+1)^2 = (f \circ g)(x)$ where $g(x) = x^2 + 1$ and $f(x) = x^2$. Then $f' = 2x$, $g' = 2x$, and
$$(f \circ g)'(x) = f'(g(x)) \cdot g'(x) = 2(g(x)) \cdot 2x = 2(x^2+1) \cdot 2x = 4x^3 + 4x.$$
Sanity check:
$$(f \circ g)'(x) = (x^4 + 2x^2 + 1)' = 4x^3 + 4x.$$
3. $\sin^2(x) = (f \circ g)(x)$ where $g(x) = \sin x$ and $f(x) = x^2$. Then $f'(x) = 2x$, $g'(x) = \cos x$, and we have
$$(f \circ g)'(x) = 2(g(x)) \cdot \cos x = 2(\sin x) \cos x.$$

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

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

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

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

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

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

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

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

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

Example 2.24. Find
$$\frac{d}{dx}\sec(x^2+\sqrt{x^3+1}).$$

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

Example 2.25. Find
$$\frac{d}{dx}\frac{\sin(x^2)+\sin^2(x)}{x^2+1}$$

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

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

Example 2.26. Find
$$\frac{d}{dx}\sqrt{\frac{\sqrt{x}+1}{(\cos x+1)^2}}$$

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

Example 2.27. Calculate
$$\frac{d}{dx}\left(\frac{\sin^2\left(\frac{x^2+1}{\sqrt{x-1}}\right)+\sqrt{x^3-2}}{\cos(\sqrt{x^2+1}+1)-\tan(x^4+3)}\right)^{5/3}$$