3 Derivatives

3.1 Review of the Slope Problem

Remember from Limits (I) the problem that started us out in inventing limits. We wanted to compute the slope of the tangent line to a curve at a given point. In particular, we asked about the slope of the curve $y = x^2$ at the point $(1, 1)$. Our approach to this was to approximate the tangent line to the curve by using a secant line, a line cutting the curve not just at one point but at 2 close together points, as pictured in Limits, Figure 1. The second point at which the secant line meets the curve is often written as $(x+\Delta x, (x+\Delta)^2)$, since this description highlights the fact that we are trying to pick a point a small distance Δx away from 1. The algebra looks a little neater, though, if we use the simpler h instead of Δx, and write the second point as $(x + h, f(x + h))$. The slope of the secant line is then

$$\frac{(1 + h)^2 - 1}{(1 + h) - 1}.$$

The secant line will get closer and closer to the tangent line as $h \to 0$, which leaves us saying that the slope of the tangent line should be

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

This agreed with what numerical experimentation and graphing had already led us to conjecture.

3.2 Definition of the derivative.

Now let's do the same thing for an arbitrary function f at an arbitrary point x. To compute the slope of the tangent line to $y = f(x)$ at the particular point $(x, f(x))$, one takes a second point $(x + h,\ f(x + h))$ on the curve, and one computes the slope of the secant line joining these 2 points. The result is

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

The slope of the tangent line at $(x, f(x))$ should be the limit of the slope of the secant lines as $h \to 0$. This slope is called the *derivative* of f at the point x, and it is written

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

This whole process is shown pictorially in Figure 1.

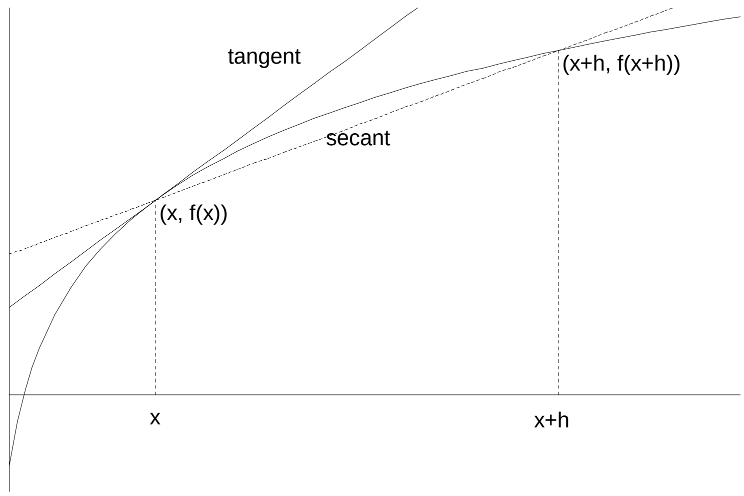

Figure 1: Tangent and secant lines.

The slope of the solid tangent line in Figure 1 is $f'(x)$. The slope of the dashed secant line is

$$\frac{\Delta y}{\Delta x} = \frac{f(x+h) - f(x)}{h}.$$

If we imagine the h shrinking toward 0, so that $x + h \to x$, the secant line will swing up closer and closer to the tangent line, until their slopes nearly coincide. The limit of the slopes of the secant lines will be the slope of the tangent line:

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

3.3 A Notational Interlude.

There are a many alternative ways to write the definition of the derivative, some simpler to look at than others, but all saying the same thing. As mentioned above, some people don't like using h as the distance between the two points. They prefer Δx, which looks obviously like a distance between two values of x. For these people, the two points on the curve would be written $(x, f(x))$ and $(x + \Delta x,\ f(x + \Delta x))$. The derivative would then be

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

Other people don't like the sum in $f(x+h)$ or $f(x+\Delta x)$. These people all the two points in the figure $(a, f(a))$ and $(x, f(x))$, and they set out to compute the slope $f'(a)$ of $y = f(x)$ at the point $(a, f(a))$. They therefore label the picture as shown in Figure 2.

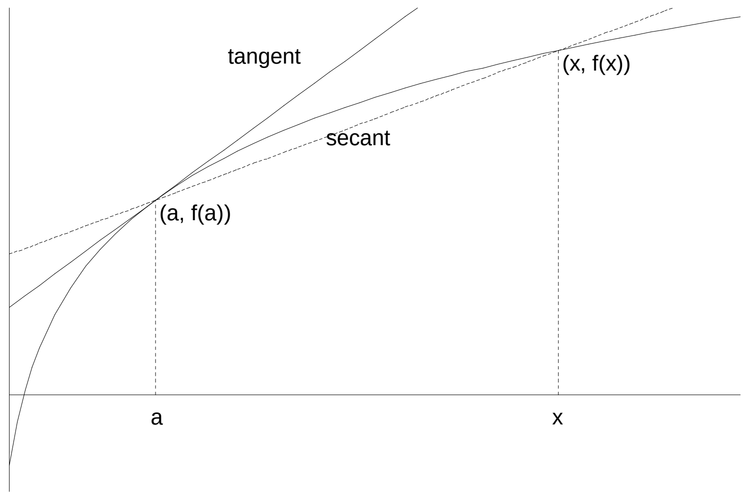

Figure 2: Alternative notation for tangents, secants.

Now the slope of the solid tangent line in is $f'(a)$. The slope of the dashed secant line is

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

The secant line will swing up closer and closer to the tangent line as x moves closer and closer to a; so now we have to take a limit as $x \to a$. The limit of the slopes of the secant lines will be the slope of the tangent line:

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

All these formulas are saying exactly the same thing; the only difference is the names of the points. Use whichever one makes you happiest. It's worth remembering they all exist, though, both so that you can talk to other people with different notational preferences and because there are times when one of these choices results in simpler arithmetic than another.

The derivative is also not always written as $f'(x)$. If $y = f(x)$, then the derivative of f at x can be written

$$f'(x) = y' = \frac{df}{dx} = \frac{dy}{dx} = \frac{d}{dx} f(x) = D_x f = D_x(y),$$

plus variants on all these. The derivative at the particular point $y = 2$ would then be

$$f'(2) = y'(2) = \left(\frac{dy}{dx}\right)(2) = \left.\frac{df}{dx}\right|_{x=2}$$

etc. Again, all these mean exactly the same thing: the slope of the tangent line to $y = f(x)$ at a generic point $(x, f(x))$ and the slope of the tangent line to the curve $y = f(x)$ at the particular point $(2, f(2))$.

The notations that look like df/dx are particularly interesting historically. The two main inventors of calculus were Sir Isaac Newton in England and Gottfried Wilhelm, Freiherr von Leibniz in Germany. Newton used notation sort of like y' for the derivative (actually, he used $\dot{y}$), while Leibniz used $\frac{dy}{dx}$. Leibniz had no notion of limits (nor had Newton), so his thought was this: the slope of the secant line between the point $(x, f(x))$ and the point $(x + \Delta x,\ f(x + \Delta x))$ is

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

Leibniz imagined computing the slope of the tangent line by taking 2 points that were only an infinitesimal distance apart—a distance too small to be written in decimal form, but still not zero. He denoted this tiny—and imaginary—distance as dx. The corresponding change in the height of the curves then $dy = df$; and the slope of the secant line, which was really equal to the tangent line, was $\frac{dy}{dx} = \frac{df}{dx}$. This idea is really rather confused, and confusing, but it gives rise to notation that captures an important insight, that the derivative is almost equal to the slope $\frac{\Delta y}{\Delta x}$ of a secant line, for very, very small values of Δy and Δx. It's also a very nice shorthand—to take a limit, just change alphabets, replacing the Greek Δ with a Latin d. In condensed, schematic form,

$$\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \frac{df}{dx}.$$

3.4 Derivatives as Rates of Change.

Here's another way to think about derivatives. Suppose $f(t)$ were the position of an object at time t. Then $f(t + \Delta t) - f(t)$ would be the distance travelled between time t and time $t + \Delta t$. The difference quotient

$$\frac{f(t + \Delta t) - f(t)}{\Delta t}$$

would be the average velocity between these two times, i.e., the average rate of change of the position f between times t and $t + \Delta t$. The derivative

$$f'(t) = \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}$$

is the limit of that average velocity over very small intervals including time t. In other words, it is the instantaneous velocity of the object exactly at time , the instantaneous rate of change of f at time t. The limit concept lets us make sense of the velocity or rate of change of a quantity not over an interval—the average velocity or rate of change—but at an instant in time—the instantaneous velocity or rate of change.

The reason we care so much about the slope problem, and the reason calculus is so important a part of the mathematical scientist's armamentarium is that in computing slopes, we are really computing instantaneous rates of change. Our ability to do this is critical to our ability to model any quantity in the universe that is able to change.

3.5 Some first computations.

On the second day of class, we made a conjecture, that the slope of the tangent line to the curve $y = x^2$ at the point (x, x^2) would be $2x$. We gave qualitative reasons to think this was true. We know it is exactly true at $x = 0$. The calculation at the top of this handout shows that it is exactly true at $x = 1$. By symmetry, it must be true at $x = -1$ as well. Now let's do it in general. By the definition of the derivative, the slope of the tangent line to $y = x^2$ at the point $(x, f(x))$ is

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

The conjecture we made at the beginning of the term that at least for the function $y = x^2$, the area and slope problems are inverses of one another, is therefore exactly correct; and thecalculation is not even very hard. We've learned a lot!

It's worth plotting the function $f(x) = x^2$ and its derivative $f'(x) = 2x$ side by side in order to think about the geometric relation between them.

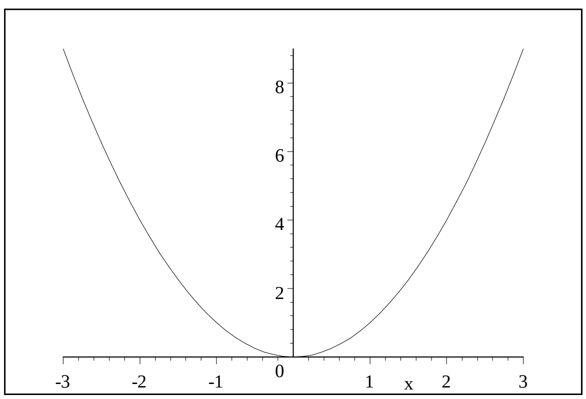

Figure 3: $y = x^2$

The derivative of f at the point x is the slope of the tangent line to f at the point $(x, f(x))$. That is, at any given point x on the x-axis, the height of the curve in Figure 4 should be the slope of the curve in Figure 3. Try to estimate the slope in Figure 3 at a few points and to convince yourself that the curve in Figure 4 is really giving you these slopes.

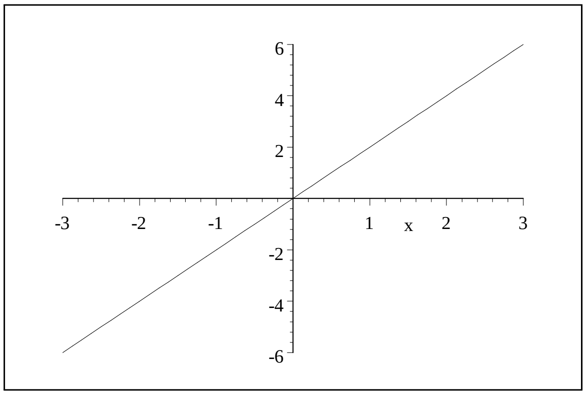

Figure 4: $y = 2x$

Let's try 2 more examples before we pause and look around. Let $g(x) = x^3$, and let us compute the slope of the tangent line to $y = g(x)$ at the point $(x, g(x))$. By the definition of the derivative, this should be

$$\begin{aligned} g'(x) &= \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} \\ &= \lim_{h\to 0} \frac{(x+h)^3 - x^3}{h} \\ &= \lim_{h\to 0} \frac{(x^2 + 3x^2h + 3xh^2 + h^3) - x^3}{h} \\ &= \lim_{h\to 0} \frac{3x^2h + 3xh^2 + h^3}{h} \\ &= \lim_{h\to 0} (3x^2 + 3xh + h^2) = 3x^2. \end{aligned}$$

The graphs of $y = g(x)$ and $y = g'(x)$ are shown in Figures 5 and 6, resp. Again, the height of the curve in Figure 6 should be the slope of the curve in Figure 5.

Finally, what if $k(x) = \frac{1}{x}$? Then the slope of the tangent line to $y = k(x)$ at

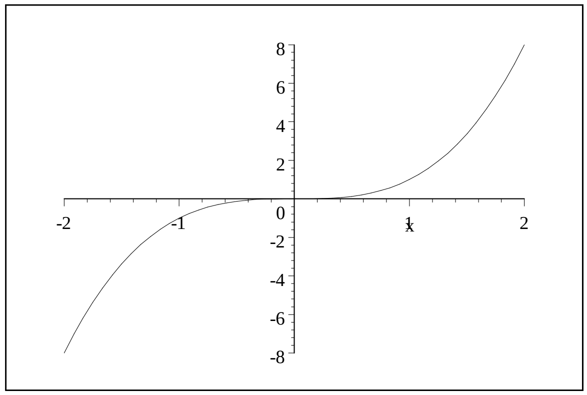

Figure 5: $y = x^3$

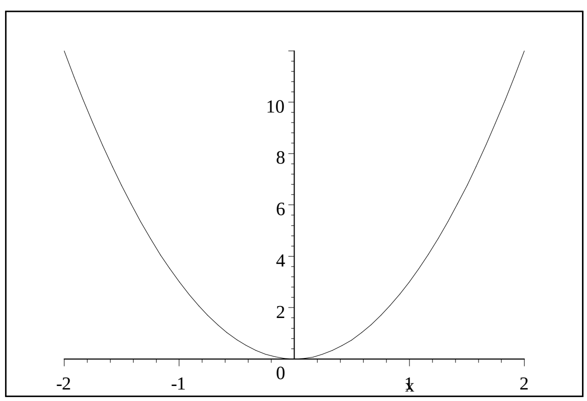

Figure 6: $y = 3x^2$

the point $(x, \frac{1}{x})$ should be

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

The graphs of $y = k(x)$ and $y = k'(x)$ are shown in Figures 7 and 8, resp. The height of the curve in Figure 8 should be the slope of the curve in Figure 7.

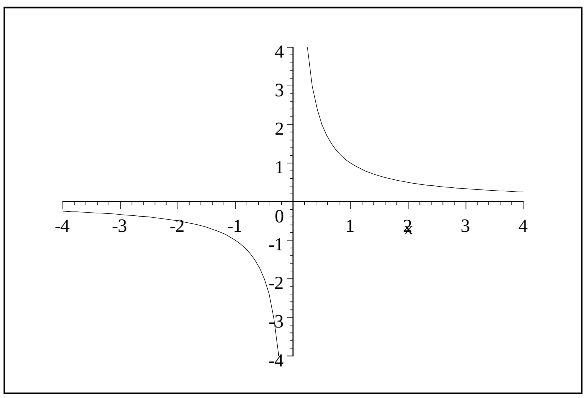

Figure 7: $y = 1/x$

3.6 Reality Check.

If $y = mx + b$ is a straight line, we now have 2 methods for computing the slope of y at the point $(x, mx + b)$. We could remember from school that the slope is always m, or we could compute the derivative. Would we get the same answer? Let's find. To make the calculation more interesting, I'll make some different notational choices than we have previously.

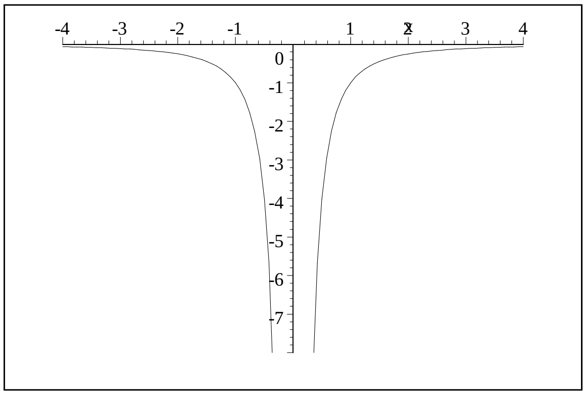

Figure 8: $y = -1/x^2$

The slope of the curve $y = mx + b$ at the point $(x, mx + b)$ is supposed to be

$$\begin{aligned} \frac{\Delta y}{\Delta x} &= \lim_{\Delta x \to 0} \frac{[m(x + \Delta x) + b] - [mx + b]}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{mx + m\Delta x + b - mx - b}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{m\Delta x}{\Delta x} = \lim_{\Delta x \to 0} m = m. \end{aligned}$$

This is exactly what they taught us in school—and what we used as our starting point for this whole theory. So our new method of computing slopes doesn't force us to change the meaning of slopes for the only curve for which we could compute the slope without calculus, the straight line.

3.7 Typical Problems.

3.7.1 Problem 1.

Compute the slope of the tangent line to the curve $y = x^2 - 3x + 2$ at the point $(2, 0)$.

One way to do this problem would be in the concrete style we used for our first slope problem. The secant line joining the point at $(2, 0)$ and another point at $(2 + \Delta x,\ (2 + \Delta x)^2 - 3(2 + \Delta x) + 2)$ has slope

$$\frac{\big[(2 + \Delta x)^2 - 3(2 + \Delta x) + 2\big] - \big[2^2 - 3(2) + 2\big]}{\Delta x} = \frac{\big[(2 + \Delta x)^2 - 3(2 + \Delta x) + 2\big] - [0]}{\Delta x}.$$

The tangent line at $(2,0)$ would therefore have slope

$$\begin{aligned} y' &= \lim_{\Delta x \to 0} \frac{\left[(2+\Delta x)^2 - 3(2+\Delta x) + 2\right] - [0]}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{4 + 4\Delta x + (\Delta x)^2 - 6 - 3\Delta x + 2}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\Delta x + (\Delta x)^2}{\Delta x} \\ &= \lim_{\Delta x \to 0} (1 + \Delta x) = 1. \end{aligned}$$

An alternative approach would be to compute the derivative at an arbitrary point x, and then to plug in $x = 2$. In this method, we would first compute

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{\left[(x+h)^2 - 3(x+h) + 2\right] - \left[x^2 - 3x + 2\right]}{h} \\ &= \lim_{h \to 0} \frac{x^2 + 2xh + h^2 - 3x - 3h + 2 - x^2 + 3x - 2}{h} \\ &= \lim_{h \to 0} \frac{2xh + h^2 - 3h}{h} \\ &= \lim_{h \to 0} (2x + h - 3) = 2x - 3. \end{aligned}$$

This is the slope of the tangent line at the point $(x, x^2 - 3x + 2)$. The slope of the tangent line at $(2,0)$ is therefore $f'(2) = 2(2) - 3 = 1$, the same answer we got before.

3.7.2 Problem 2.

Compute the equation of the tangent line to the curve $y = x^2 - 3x + 2$ at the point $(2,0)$.

The first part of this problem would be to do Problem 1, computing the slope of the tangent line. Once we've done that, the problem is easy. The tangent line has slope 1, and it goes through the point $(2,0)$; so its equation must be $(y-0) = 1(x-2)$, i.e., $y = x - 2$. As a check, the curve and this line are plotted together in Figure 9. They look pretty tangent.

3.8 Functions without derivatives.

We've worked out now how to compute the derivatives of quite a lot of functions, and we're about to start looking for general rules that will let us differentiate practically every function. But before we do that, we ought to pause and clarify slightly our idea of derivatives by observing that not every function has a derivative at every point. First of all, only a function that is continuous at $x = a$ can have a derivative at $x = a$. This seems reasonable enough: surely any function that can't even be drawn at $x = a$ without picking up ones pencil isn't be smooth enough to have a tangent line at a. Here's a clever argument

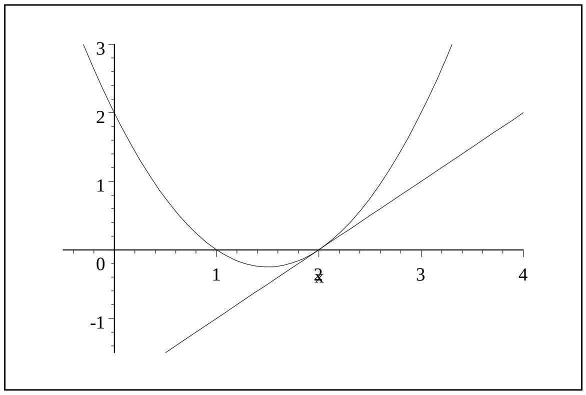

Figure 9: $y = x^2 - 3x + 2$.

that actually proves this fact analytically using our definitions of derivative and of continuity: Suppose

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

exists. Then

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

which means that

$$\begin{aligned} \lim_{x \to a} f(x) &= \lim_{x \to a} ((f(x) - f(a)) + f(a)) \\ &= \lim_{x \to a} (f(x) - f(a)) + \lim_{x \to a} f(a) \\ &= 0 + f(a) = f(a), \end{aligned}$$

which means that f is continuous at a.

Does every function that is continuous at a and have a derivative at a? No. Consider, for instance, $f(x) = |x|$ at $a = 0$. The derivative $f'(0)$, if it existed, would be

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

If $h > 0$, then $|h| = h$, while if $h < 0$, then $|h| = -h$. (Think about $|2|$ and $|-2|$ if you don't see this.) Thus,

$$\lim_{h \to 0^+} \frac{|h|}{h} = \lim_{h \to 0^+} 1 = 1,$$

but

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

Since the left and right hand limits differ, $\lim_{h \to 0} \frac{|h|}{h}$ does not exist, and $y = |x|$ does not have a derivative at $x = 0$. This seems reasonable if you think of the graph of $|x|$ (Figure 10). What could one mean by the slope of the tangent line to this function at $x = 0$, where the graph comes to a sharp point?

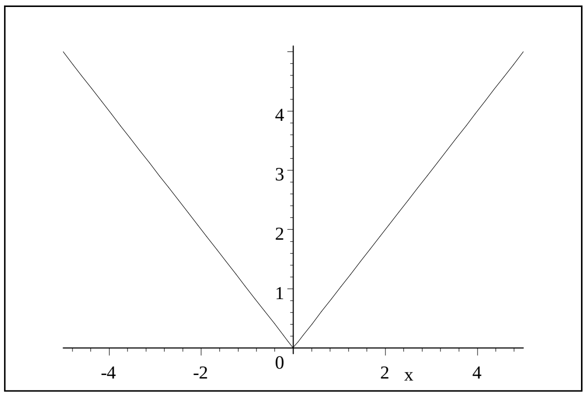

Figure 10: $y = |x|$.

3.9 Basic rules for computing derivatives.

3.9.1 Powers of x.

So far, we know that

$$\begin{aligned} \frac{d}{dx}(x) &= 1 \\ \frac{d}{dx}\left(x^2\right) &= 2x \\ \frac{d}{dx}\left(x^3\right) &= 3x^2. \end{aligned}$$

Is there any pattern here? How about

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

To prove that this conjectured formula works for any positive integer value of n, we would have to let $f(x) = x^n$ and to evaluate

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

The second form above seems to be the easier one to work with, since we can imagine just doing long division to divide $x - a$ into $x^n - a^n$ (try it!) to get

$$x^n - a^n = (x-a)\left(x^{n-1} + ax^{n-2} + a^2x^{n-3} + \cdots + a^{n-2}x + a^{n-1}\right).$$

This means that

$$\begin{aligned} \lim_{x\to a} \frac{x^n - a^n}{x - a} &= \lim_{x\to a}\left(x^{n-1} + ax^{n-2} + a^2x^{n-3} + \cdots + a^{n-2}x + a^{n-1}\right) \\ &= \left(a^{n-1} + aa^{n-2} + a^2a^{n-3} + \cdots + a^{n-2}a + a^{n-1}\right) \\ &= na^{n-1}, \end{aligned}$$

just the formula we were hoping for. If you want an interesting algebraic challenge, you might enjoy seeing that you get the same result when you compute

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

Remembering at least roughly what the Binomial Theorem says will be a help here.

The formula that $(x^n)' = nx^{n-1}$ works more generally than just for positive integer values of n, though. For instance, we have already seen that

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

which is the same formula with $n = -1$.

We could also try the case $n = \frac{1}{2}$, which means we have to compute the derivative of $f(x) = x^{1/2} = \sqrt{x}$. We get

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

the result of the formula nx^{n-1} with $n = \frac{1}{2}$.

We'll come back later to the formula

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

later, and see if we can prove its validity for all real n.

3.9.2 Constant multiples.

Now that we know how to differentiate at least every positive integer power of x, let's see about combining these powers to make polynomials.

The first thing we need to be able to do is to multiply powers of x by constants to get functions like $3x^5$. Here the theorem is simple: If c is a constant, then

$$(cf(x))' = c \cdot f'(x).$$

To prove this, we compute

$$\begin{aligned}
\frac{d}{dx}\left(cf(x)\right) &= \lim_{h\to 0}\frac{cf(x+h)-cf(x)}{h} \\
&= \lim_{h\to 0}\left(c\cdot\frac{f(x+h)-f(x)}{h}\right) \\
&= \left(\lim_{h\to 0} c\right)\left(\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}\right) \\
&= c\cdot f'(x).
\end{aligned}$$

We therefore know things like

$$\frac{d}{dx}\left(3x^5\right) = 3\frac{d}{dx}\left(x^5\right) = 3\cdot 5x^4 = 15x^4.$$

3.9.3 Sums and Differences.

To get polynomials, we now have to add up terms that look like $3x^5$. To take the derivative of a sum of functions is also simple:

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

Again, the proof of this is a simple computation:

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

By exactly the same argument, the derivative of the difference of two functions is

$$(f(x)-g(x))' = f'(x)-g'(x).$$

By combining the results we have so far, we can compute the derivative of any polynomial. For instance,

$$\begin{aligned}\frac{d}{dx}\left(x^{17}-11x^{7}+6x^{2}+13\right) &= \left(x^{17}\right)'-\left(11x^{7}\right)'+\left(6x^{2}\right)'+(13)' \\ &= \left(x^{17}\right)'-11\left(x^{7}\right)'+6\left(x^{2}\right)'+(13)' \\ &= 17x^{16}-77x^{6}+12x+0.\end{aligned}$$

(Remember that a constant function is a horizontal straight line with slope 0, so that its derivative should be 0. You can also get this right out of the definition of derivative applied to a constant function.)

3.10 Preview.

We can now compute the derivatives of an enormous collection of functions. Are we done? No. There are still lots of functions we can't differentiate. These include

Trig functions, like $\sin x$.

Products of functions, like $x^2 \cdot \sin x$.

Quotients of functions, like $\frac{x^2+x-1}{5x+2}$.

Powers of functions, like $\left(4x^2+x+11\right)^{100}$, or in general, functions of functions, like $\sin\left(x^2\right)$.

There's also the matter of non-positive-integer powers of x, like x^{-5} or $x^{22/7}$.

So we still need some more rules, but we have made enormous headway in very short order.

3.11 Derivatives of $\sin x$ and $\cos x$.

If we were to guess what the derivatives of the sine and cosine functions were like, what would we say? The graph of $y = \sin x$ is shown in Figure 11.

The derivative of $y = \sin x$ has to be something like the function graphed in Figure 12.

The function in Figure 12 looks at least superficially like the function $y = \cos x$, so a reasonable conjecture would be that

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

Similarly, the graph of $y = \cos x$ and and approximation to its derivative are shown in Figures 13 and 14, resp.

It would appear that perhaps

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

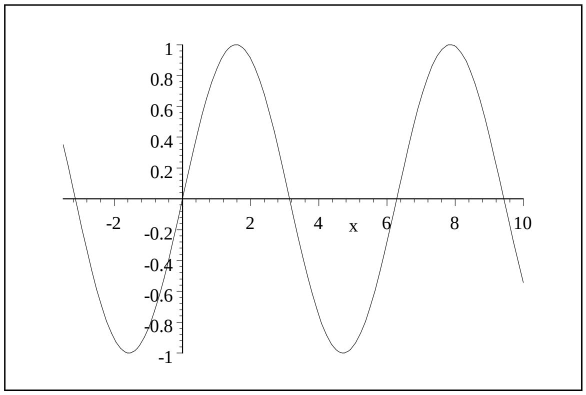

Figure 11: $y = \sin x$.

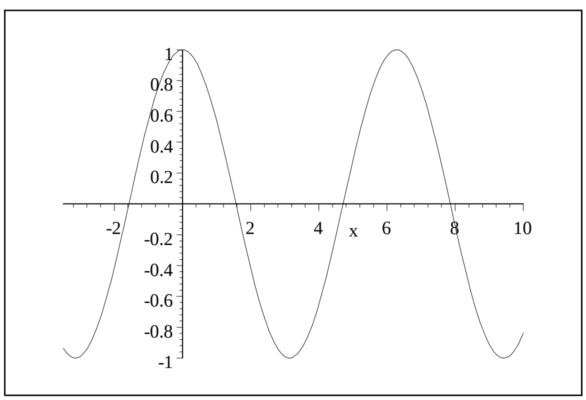

Figure 12: The derivative of $\sin x$.

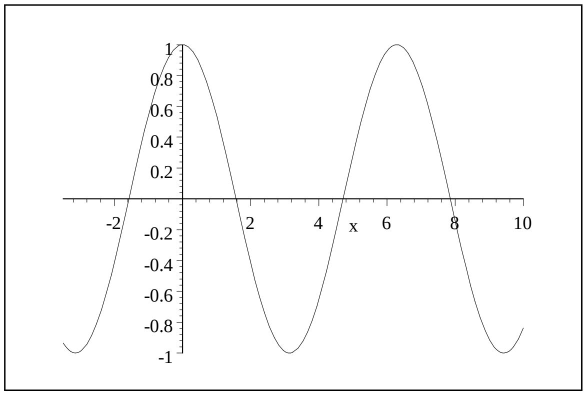

Figure 13: $y = \cos x$.

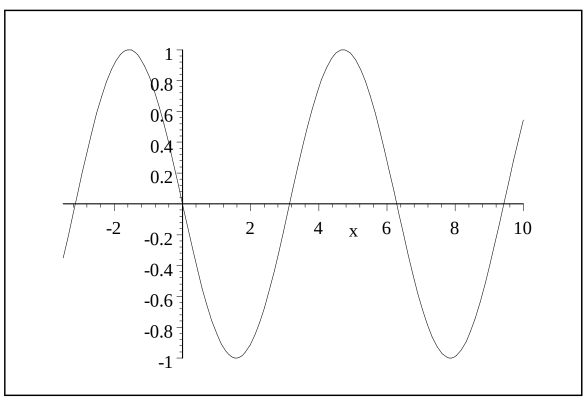

Figure 14: The derivative of $\cos x$.

Can we prove these claims? It turns out (maybe surprisingly) that the answer is yes. To get the derivative of $\sin x$, we say, if $f(x) = \sin x$, then

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

In line 2 of this calculation, we have used the sum formula for the sine function. In line 6, we use 2 of the limits we conjectured in the lab and later proved in class.

To compute the derivative of $\cos x$, we could try a similar calculation using the sum formula for the cosine. We could also observe that the graph of the cosine function is the graph of the sine function moved left by a distance $\pi/2$.

$$\cos x = \sin\left(x + \frac{\pi}{2}\right).$$

The derivative of the cosine function should therefore be the derivative of the sine function moved left by a distance $\pi/2$. That is,

$$\frac{d}{dx}(\cos x) == \frac{d}{dx}\left(\sin\left(x + \frac{\pi}{2}\right)\right) = \cos\left(x + \frac{\pi}{2}\right).$$

Looking at the unit circle or using some of our earlier trig identities shows that $\cos(x + \pi/2) = -\sin x$, so that

$$\frac{d}{dx}(\cos x) = \cos\left(x + \frac{\pi}{2}\right) = -\sin x.$$

3.12 The Product Rule.

The Product Rule says

$$\frac{d}{dx}(f(x)g(x)) = f'(x)g(x) + f(x)g'(x).$$

or, in shorthand,

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

The proof is a by now standard trick of adding and subtracting the same thing. Much of mathematics consists of exactly this: adding 0 and multiplying by 1 in clever ways.

$$\begin{aligned}
&\frac{d}{dx}(f(x)g(x)) \\
= \quad & \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\
= \quad & \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x+h) + f(x)g(x+h) - f(x)g(x)}{h} \\
= \quad & \lim_{h\to 0} \left[\frac{f(x+h)g(x+h) - f(x)g(x+h)}{h} + \frac{f(x)g(x+h) - f(x)g(x)}{h}\right] \\
= \quad & \lim_{h\to 0} \left[\frac{f(x+h)g(x+h) - f(x)g(x+h)}{h}\right] + \lim_{h\to 0} \left[\frac{f(x)g(x+h) - f(x)g(x)}{h}\right] \\
= \quad & \left(\lim_{h\to 0} \frac{f(x+h) - f(x)}{h}\right)\left(\lim_{x\to 0} g(x+h)\right) + \left(\lim_{h\to 0} f(x)\right)\left(\lim_{h\to 0} \frac{g(x+h) - g(x)}{h}\right) \\
= \quad & f'(x)g(x) + f(x)g'(x).
\end{aligned}$$

It may initially seem surprising that $(fg)' \neq f'g'$. Here's a picture that may clarify what's happeneing in the Product Rule. Suppose $f(t)$ represents the length of a rectangle and $g(t)$ represents its width. Both these quantities are changing with time—the rectangle is growing. The function $f(t)g(t)$ then represents the area of the rectangle; and $\frac{d}{dt}(f(t)g(t))$ is the rate of change of this area.

The picture in Figure 15 shows the rectangle at time t (solid) along with the growth of the rectangle between time t and time $t + \Delta t$ (dashed).

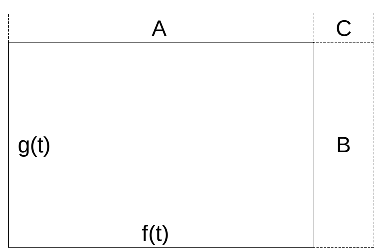

Figure 15: Rectangles and the Product Rule.

Between time t and time $t + \Delta t$, the area of the solid rectangle increases because of the addition of three smaller rectangles, A, B, and C. The rectangle A has length $f(t)$ and width about $g'(t)\Delta t$. Rectangle B has length $g(t)$ and width about $f'(t)\Delta t$. The heigt and width of rectangle C are about $f'(t)\Delta t$ and $g'(t)\Delta t$. Thus, the increase in area of the reactangle between time t and time

$t + \Delta t$ is roughly

$$f(t)g'(t)\Delta t + g(t)f'(t)\Delta t + f'(t)g'(t)(\Delta t)^2.$$

The rate of increase of the area is about

$$\frac{f(t)g'(t)\Delta t + g(t)f'(t)\Delta t + f'(t)g'(t)(\Delta t)^2}{\Delta t} = f(t)g'(t)+g(t)f'(t)+f'(t)g'(t)\Delta t.$$

As $\Delta t \to 0$, this rate of change approaches

$$\frac{d}{dt}(f(t)g(t)) = f(t)g'(t) + g(t)f'(t),$$

which is just what the Product Rule claimed. The two terms in the Product Rule therefore correspond to the rates of increase of area caused by expansion of the two sides of the rectangle, resp.

3.13 The Quotient Rule

The rule for differentiating reciprocals is

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

One can convince onself of this by just doing the only calculation one could do:

$$\begin{aligned}
\frac{d}{dx}\left(\frac{1}{g(x)}\right) &= \lim_{h\to 0} \frac{\frac{1}{g(x+h)} - \frac{1}{g(x)}}{h} \\
&= \lim_{h\to 0} \frac{1}{h}\left[\frac{1}{g(x+h)} - \frac{1}{g(x)}\right] \\
&= \lim_{h\to 0} \frac{1}{h}\left[\frac{g(x) - g(x+h)}{g(x)g(x+h)}\right] \\
&= \lim_{h\to 0} \left[\frac{1}{g(x)g(x+h)} \cdot \frac{g(x) - g(x+h)}{h}\right] \\
&= \left(\lim_{h\to 0} \frac{1}{g(x)g(x+h)}\right)\left(\lim_{h\to 0} \frac{g(x) - g(x+h)}{h}\right) \\
&= \frac{1}{g(x)^2}(-g'(x)) = -\frac{g'(x)}{g(x)^2}.
\end{aligned}$$

The formula for the derivative of an arbitrary quotient can now be derived by combining the Product and Reciprocal Rules:

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

The Quotient Rule is harder to remember than the Product Rule because it has that minus sign. You have to get the terms in the right order. One solution to this is a bit of mnemonic doggerel: to differentiate $\frac{Hi}{Ho}$, we say, “Ho-de-Hi minus Hi-de-Ho, Square the bottom, and away we go!” “De” here means derivative. The rhyme forces you to get the terms in the right order; since the alternative is “Hi-de-Ho minus Ho-de-Hi, Square the bottom, and we all fry!”

3.14 Other Trig Functions.

An immediate use of the quotient rule is to compute the derivatives of the other trig functions.

$$\begin{aligned}
\frac{d}{dx}\sec x &= \frac{d}{dx}\left(\frac{1}{\cos x}\right)\\
&= -\frac{(d/dx)(\cos x)}{\cos^2 x}\\
&= \frac{\sin x}{\cos^2 x}\\
&= \frac{1}{\cos x}\cdot\frac{\sin x}{\cos x} = \sec x\tan x.
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx}\tan x &= \frac{d}{dx}\left(\frac{\sin x}{\cos x}\right)\\
&= \frac{(\cos x)((d/dx)\sin x) - (\sin x)((d/dx)\cos x)}{\cos^2 x}\\
&= \frac{(\cos x)(\cos x) - (\sin x)(-\sin x)}{\cos^2 x}\\
&= \frac{1}{\cos^2 x} = \sec^2 x.
\end{aligned}$$

Calculations just like these give the derivatives of the cosecant and cotangent functions. If we list all these results together in a table, one can see some of the symmetries involved.

$f(x)$	$f'(x)$
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\tan x$	$\sec^2 x$
$\cot x$	$-\csc^2 x$
$\sec x$	$\sec x \tan x$
$\csc x$	$-\csc x \cot x$

3.15 Examples.

$$\begin{aligned}
\frac{d}{dx}(x \sin x) &= 1 \sin x + x \cos x \\
\frac{d}{dx}(\sin^2 x) &= \frac{d}{dx}(\sin x \cdot \sin x) = \cos x \sin x + \sin x \cos x = 2 \sin x \cos x \\
\frac{d}{dx}\left(\frac{x^2+3x+5}{7x^3+11}\right) &= \frac{(7x^3+11)(2x+3) - (x^2+3x+5)(21x^2)}{(7x^3+11)^2} \\
\frac{d}{dx}\left(\frac{\sin x \cos x}{x^2+1}\right) &= \frac{(x^2+1)\frac{d}{dx}(\sin x \cos x) - (\sin x \cos x)2x}{(x^2+1)^2} \\
&= \frac{(x^2+1)(\cos^2 x - \sin^2 x) - (\sin x \cos x)2x}{(x^2+1)^2} \\
\frac{d}{dx}\left(\frac{x}{2x+3} \cdot \frac{\tan x}{\sqrt{x}}\right) &= \frac{(2x+3)1 - x(2)}{(2x+3)^2}\frac{\tan x}{\sqrt{x}} + \frac{x}{2x+3}\frac{\sqrt{x}\sec^2 x - (\tan x)\left(\frac{1}{2\sqrt{x}}\right)}{x}.
\end{aligned}$$

3.16 Higher Order Derivatives.

We've been starting with a function $f(x)$ and computing its derivative $f'(x)$. But $f'(x)$ is a function in its own right. We can therefore perfectly well imagine taking its derivative to get yet another function, $f''(x)$. We could, if we wanted, keep going and take the derivative of $f''(x)$, which is called $f'''(x)$ and so on. These functions are called the second derivative of f, the third derivative of f, and so on. Derivatives higher than the third are normally written using a superscript in parentheses to denote the order of the derivative. Thus, we write

$$\begin{aligned}
f'''(x) &= f^{(3)}(x), \\
f''''(x) &= f^{(4)}(x), \\
f'''''(x) &= f^{(5)}(x),
\end{aligned}$$

and so on. Occasionally, it is also convenient to use this notation for lower order derivatives, usually in formulas in which we are summing over many derivatives. We could then write

$$\begin{aligned}
f''(x) &= f^{(2)}(x) \\
f'(x) &= f^{(1)}(x) \\
f(x) &= f^{(0)}(x).
\end{aligned}$$

If we are using Leibniz' notation for derivatives, and if $y = f(x)$, then

$$\begin{aligned} f''(x) &= \frac{d}{dx}\left(\frac{df}{dx}\right) = \frac{d^2 f}{dx^2} = \frac{d^2 y}{dx^2} = \frac{d^2}{dx^2} f(x), \\ f'''(x) &= \frac{d}{dx}\left(\frac{d^2 f}{dx^2}\right) = \frac{d^3 f}{dx^3} = \frac{d^3 y}{dx^3} = \frac{d^3}{dx^3} f(x), \end{aligned}$$

and so on.

For polynomials, higher order derivatives eventually get simpler and simpler. Here's an example:

$$\begin{aligned} f(x) &= x^3 - 27x^2 + 11x + 6 \\ f'(x) &= 3x^2 - 54x + 11 \\ f''(x) &= 6x - 54 \\ f'''(x) &= 6 \\ f^{(4)}(x) &= f^{(5)}(x) = f^{(6)}(x) = ... = 0. \end{aligned}$$

For other functions, the higher order derivatives may get more complicated, not less.

$$\begin{aligned} f(x) &= \frac{x+1}{3x^2-1} \\ f'(x) &= \frac{(3x^2-1)1 - (x+1)(6x)}{(3x^2-1)^2} \\ &= -\frac{3x^2+6x+1}{9x^4-6x^2+1} \\ f''(x) &= -\frac{(9x^4-6x^2+1)(6x+6) - (3x^2+6x+1)(36x^3-12x)}{(9x^4-6x^2+1)^2} \\ &= \frac{18x^3+54x^2+18x+6}{27x^3-27x^2+9x-1}. \end{aligned}$$

For the sine and cosine functions, the derivatives cycle:

$$\begin{aligned} f(x) &= \sin x \\ f'(x) &= \cos x \\ f''(x) &= -\sin x \\ f'''(x) &= -\cos x \\ f^{(4)}(x) &= \sin x \\ f^{(5)}(x) &= \cos x. \end{aligned}$$

3.17 A Physical Interpretation of Higher Derivatives.

I once rode an elevator down from the 39th floor apartment of my friend Charlie Fefferman, who at the time was 22, a full professor at Chicago and at Princeton, and one of the best mathematical analysts in the world. I commented that the elevator was very fast but very smooth. Charlie nodded. "Small third derivative," he said.

Could he possibly have been right? Can one feel third derivatives?

To investigate this question, suppose you wake up one morning and find yourself in an elevator. This would raise all manner of questions that are outside the scope of this class, but it would also raise questions we can actually address.

Let $s(t)$ be the position of the elevator as a function of time. In American units, $s(t)$ might be measured in feet above the ground. When you wake up in the elevator, you have no way to directly experience the value of $s(t)$. To be on the first floor or the 50th floor feels the same. The only way to determine $s(t)$ is to read the lights or listen to the bells showing the floor. If you can't see the lights or hear the bells, then you don't know where you are.

The derivative $s'(t)$ is the rate of change of the height of the elevator. We know that

$$s'(t) = \lim_{\Delta t \to 0} \frac{s(t + \Delta t) - s(t)}{\Delta t}.$$

The numerator, Δs, in the difference quotient is measured in feet, and the denominator, Δt, is measured in seconds. The rate of change, $s'(t)$, will therefore have units of feet per second. $s'(t)$ is the velocity at which the elevator is moving.

Now, one can imagine being able to determine whether or not the elevator was moving by hearing the sound of the motor or by feeling its vibration. But if the motor is very smooth or if one cannot hear it, then there really is no way to tell directly when you wake up whether the elevator is moving or not. Indirectly, one can measure the value of $s'(t)$ by observing how fast the floor numbers are changing. If the spacing between floors is about 10 feet, and if in one second you drop from floor 17 to floor 14, then roughly, $s'(t) = -30\,\mathrm{ft/s}$.

So far, we haven't yet hit a function we can directly experience; but what about the second derivative, $s''(t)$? The second derivative is the rate of change of the velocity.

$$s''(t) = \lim_{\Delta t \to 0} \frac{s'(t + \Delta t) - s'(t)}{\Delta t},$$

so $s''(t)$ will be measured in units of feet per second per second, or $\mathrm{ft/s^2}$. This rate of change is called the *acceleration* of the elevator. A large acceleration, say $30\,\mathrm{ft/s^2}$, would mean that each second, the elevator was going upward at a speed of $30\,\mathrm{ft/s}$ faster than the previous second. An acceleration like $-0.03\,\mathrm{ft/s^2}$ would mean that each second the upward velocity was $0.03\,\mathrm{ft/s}$ less than it had been a second before.

How do you measure acceleration in the elevator? Well, you could estimate velocity at 2 points and look at its rate of change, but this would be complicated and error-prone. But there's a simpler way. When you accelerate upwards, you weigh more; when you accelerate downwards, you weigh less. You can feel acceleration in the pit of your stomach, whether you can read the floor numbers or not.

In fact, one can even be quantitative about this observation. According to Newton, $F = ma$, where F is the force on an object, m is its mass, and

$a = s''(t)$ is its acceleration. So if you know your mass, and if you plan ahead and take a bathroom scale with you when you go into the elevator so that you can measure quantitatively the force on your body, then you can immediately infer the acceleration.

Now on to the third derivative, $s'''(t)$. The third derivative is the rate of change of the acceleration. It will therefore be large in absolute value if the acceleration changes rapidly. Rapid change in acceleration would result, for instance, in you weighing a lot one moment, and very little the next; or in you weighing very little one moment and a lot the next. That is, the third derivative will be large during a jerk. Just as $s'(t)$ is called the velocity and $s''(t)$ is called the acceleration, $s'''(t)$ also has a name—it's called the *jerk*, which is intended to capture its physical meaning. So my friend Charlie Fefferman was right: one can feel the third derivative directly, and a smooth ride means a small third derivative.

3.18 The Chain Rule

We've seen in lab the last situation in which we need to compute derivatives, the case of functions built by composing one function with another to produce $f(g(x))$. Our conjecture is that

$$\frac{d}{dx} f(g(x)) = f'(g(x))g'(x).$$

This conjecture is in fact correct; it is called in the trade the Chain Rule. Before we try to prove it, though, let's pause for a moment and make sure we can parse the notation. On the left hand side, $f(g(x))$ says to take $f(x)$ and to replace x by $g(x)$ everywhere. $\frac{d}{dx} f(g(x))$ says to take the derivative of the resulting expression. So the left hand side means to substitute $g(x)$ for x and then to differentiate.

On the right hand side, $f'(g(x))$ says to first take the derivative of f, and then to substitute $g(x)$ for x. The Chain Rule tells us that the order in which we do the substitution and differentiation matters. It is not the case that $\frac{d}{dx} f(g(x)) = f'(g(x))$. Rather, one has to multiply the right hand side by $g'(x)$ in order to get equality.

The point of the Chain Rule, though, is not really that the order in which we differentiate and substitute matters. Rather, it tells us that if we can differentiate f and g, then we can also differentiate $f(g(x))$. Since every function one can write explicitly seems to be built up from powers of x or from sines and cosines by addition, subtraction, multiplication, division, or composition (the process of forming $f(g(x))$), it seems that we have now completed the process of learning how to differentiate everything.

3.18.1 A Modern "Proof" of the Chain Rule.

To prove the Chain Rule is actually a bit delicate. Let me give what is essentially the modern argument, with a bit of technical detail left out. If we just plunge boldly into a calculation, we would use the definition of the derivative to write

$$\frac{d}{dx}f(g(x)) = \lim_{\Delta x \to 0} \frac{f(g(x+\Delta x)) - f(g(x))}{\Delta x}.$$

It's not initially clear what to do with this limit. How could we simplify it at all? An answer turns out to be to try to bridge the gap between the numerator and denominator by multiplying and dividing by something algebraically intermediate between the two. (We multiply by 1 in a complicated way.) That is, we write

$$\begin{aligned}
\frac{d}{dx}f(g(x)) &= \lim_{\Delta x \to 0} \frac{f(g(x+\Delta x)) - f(g(x))}{\Delta x} \\
&= \lim_{\Delta x \to 0} \frac{f(g(x+\Delta x)) - f(g(x))}{g(x+\Delta x) - g(x)} \cdot \frac{g(x+\Delta x) - g(x)}{\Delta x} \\
&= \left[\lim_{\Delta x \to 0} \frac{f(g(x+\Delta x)) - f(g(x))}{g(x+\Delta x) - g(x)}\right] \cdot \left[\lim_{\Delta x \to 0} \frac{g(x+\Delta x) - g(x)}{\Delta x}\right] \\
&= \left[\lim_{\Delta x \to 0} \frac{f(g(x+\Delta x)) - f(g(x))}{g(x+\Delta x) - g(x)}\right] g'(x).
\end{aligned}$$

We've now made real progress. We've pulled the derivative $\frac{d}{dx}f(g(x))$ apart into 2 factors, one of which is $g'(x)$, just as the Chain Rule say it should be. What's left to do is to argue that the remaining ugly limit is in fact $f'(g(x))$.

To do this, remember that if g is differentiable at x, then g is continuous at x. This means that as $\Delta x \to 0$, $g(x+\Delta x) \to g(x)$. To clean up our notation a bit, define

$$\begin{aligned}
\Delta g &= g(x+\Delta x) - g(x) \\
z &= g(x).
\end{aligned}$$

We know that when $\Delta x \to 0$, $\Delta g \to 0$ as well. So the remaining limit in our expression for $\frac{d}{dx}f(g(x))$ can be rewritten as

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

Plugging this into the formula above completes the proof.

Honesty compels me to point out that there is a slight problem with this argument. There is no reason there should not be points $x+\Delta x$ near x at which $g(x+\Delta x) - g(x) = 0$. At points like this, the expression $g(x+\Delta x) - g(x)$ we have multiplied and divided by is 0, which is a problem, since we can't divide by 0. A completely rigorous proof would need to address this concern. Come back in Analysis A, and we'll do that.

3.18.2 A Classical Proof of the Chain Rule.

The modern proof above seems to require one to be awfully clever. Is this really what the initial architects of calculus did? No, and it couldn't have been. Newton and Leibniz didn't even know about limits, after all. A lot of insight can be gained by looking at how Leibniz would have thought about the Chain Rule. First, let's rewrite the Chain Rule by just changing the name of g, which looks like it could only be a function, to u, which looks like it could be either a function or a variable. The Chain Rule would then say in something like Newton's notation that

$$(f(u(x)))' = f'(u(x))u'(x).$$

How would this look in Leibniz' notation? We know Leibniz would have written $u'(x)$ as $\frac{du}{dx}$. How would he have written the other two derivatives in the formula? He would have been rather less explicit than modern notation causes one to be. He would have said, $f = f(u) = f(u(x))$ can be though of either as a function of u where we know u really depends on x, or as a function of x. Regarding f as a function of x, Leibniz would write the derivative $(f(u(x)))'$ as $\frac{df}{dx}$. Regarding f as a function of u, Leibniz would write the derivative $f'(u(x))$ as $\frac{df}{du}$. So the Chain Rule written by Leibniz would say

$$\frac{df}{dx} = \frac{df}{du}\frac{du}{dx}.$$

How would Leibniz prove this? He'd cancel the du and be done!

In Leibniz' understanding of the calculus, this would make perfect sense. Leibniz regarded the derivative dy/dx as really being a difference quotient $\Delta y/\Delta x$ in which the numerator and denominator were both infinitesimally small; so for him, the derivatives above were really fractions, and the cancellation was completely legitimate. For us, the situation is more complicated. $\Delta y/\Delta x$ is not a fraction, but a limit of difference quotients. It doesn't really have a numerator and a denominator that can be cancelled. We therefore need the more complex argument. But if you look again at our proof, you still see the traces of Leibniz' understanding. Leibniz multiplies and divides by du. Inside the limit, we multiply and divide by $g(x + \Delta x) - g(x) = \Delta g = \Delta u$ (remember that g and u were different names for the same thing). So Leibniz' proof is still alive in the modern argument.

Even though Leibniz' interpretation of calculus in terms of infinitesimals isn't the post-1850 understanding of the subject, arguments treating dy and dx as if they were numbers and dy/dx as if it were a fraction can nearly always be rephrased in terms of limits to yield valid modern proofs. Indeed, if this were not so, then it would probably have been necessary to base calculus on some concept other than that of limit in order to preserve the validity of these Leibnizean arguments. Physicists and chemists, economists and engineers, and other practitioners of calculus other than pure mathematicians often work and think in terms of these infinitesimals, knowing or trusting that the professional mathematicians could, if needed, render rigorous the simple arguments they produce with Leibniz' tools.

Leibniz' argument for the Chain Rule also holds a simple intuitive understanding of what's going on inside the daunting looking expression $f'(u(x))u'(x)$. Suppose, for concreteness, that x represents time, that u represents the market value of some investment, and that f represents the blood pressure of the owner of that investment. The value of the investment has begun to drop in eager trading, and the investor's blood pressure, which depends on the value of his investment, has begun to rise.

We know that the value u of the investment is a function of the time x. We should therefore write f as $f(u(x))$, since f depends on u which depends on x. The derivative $\frac{d}{dx}(f(u(x)))$ therefore represents the rate of change of the investor's blood pressure as a function of time. The derivative $u'(x)$ represents the rate of change of the value of the investment as a function of time, and the derivative $f'(u(x)) = \frac{df}{du}$ represents the rate of change of the blood pressure as a function of the value. What the Chain Rule says is that

$$\frac{df}{dx} = \frac{df}{du} \cdot \frac{du}{dx},$$

i.e., that to find how fast the investor's blood pressure is changing with time, we should find the rate of change of the value of his investment as a function of time, and multiply by the rate of change of his blood pressure as a function of the value of his investment. This seems to me to be eminently reasonable.

3.18.3 Examples of the Chain Rule.

$$\frac{d}{dx}(\sin(x^2)) = \cos(x^2) \cdot 2x.$$

(Here $f(x) = \sin x$ and $g(x) = x^2$.)

$$\frac{d}{dx}(\sin^2 x) = 2 \sin x \cos x.$$

(Here $f(x) = x^2$ and $g(x) = \sin x$.)

$$\frac{d}{dx}\left(\left(\frac{3x+5}{x^2+1}\right)^{100}\right) = 100\left(\frac{3x+5}{x^2+1}\right)^{99} \cdot \frac{(x^2+1)3 - (3x+5)2x}{(x^2+1)^2}.$$

(Here $f(x) = x^{100}$ and $g(x) = \frac{3x+5}{x^2+1}$.)

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

(Here we needed to use first the Product Rule and the the Chain Rule.)

$$\begin{aligned} \frac{d}{dx}(f(g(h(x)))) &= f'(g(h(x))) \cdot \frac{d}{dx}(g(h(x))) \\ &= f'(g(h(x)))g'(h(x))h'(x). \end{aligned}$$

For instance,

$$\frac{d}{dx}\left(\sin\sqrt{x^3+11}\right) = \cos\sqrt{x^3+11}\cdot\frac{1}{2\sqrt{x^3+11}}\cdot 3x^2.$$

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

If we can compute the derivative of this, then we really can compute the derivative of just about anything.

3.19 Implicit Differentiation and Related Rates

These are two interesting ideas that represent side lights in our current quest for understanding, but that are worth casting a glance at as we go past them. There are situations in which each is critically important.

3.19.1 Implicit Differentiation

A first example The simplest way to write the equation of the unit circle in Cartesian coordinates is $x^2 + y^2 = 1$. Suppose we wanted to know the slope of the tangent line to this curve at the point $\left(\frac{3}{5}, \frac{4}{5}\right)$. (This is actually easy to get from elementary geometry, but bear with me for the example.)

The obvious naive approach would be to solve for y and the differentiate. We'd get

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

so that

$$y' = \frac{1}{2\sqrt{1-x^2}}(-2x) = -\frac{x}{\sqrt{1-x^2}}.$$

At $x = \frac{3}{5}$, $y' = -\frac{3/5}{\sqrt{1-(3/5)^2}} = -\frac{3}{4}$.

There are two problems with this naive approach. First, the initial equation wouldn't have to get much more complicated than $x^2 + y^2 = 1$ before it became difficult or impossible to solve. Second, the fact that we get a complicated expression when we solve for y means that our expression for y' is guaranteed to be a mess. Is there any way to resolve these concerns?

Yes. There's a clever second way to compute the slope in situations like this, called Implicit Differentiation. The idea is to start with the original equation $x^2 + y^2 = 1$ and to differentiate everything in sight without bothering to solve for y first. When we do differentiate, though, we have to remember that x is just the variable, but that $y = y(x)$ is really a function of x. This means that the derivative of x is 1, but that the derivative of y is y'. When we differentiate, we therefore get

$$\begin{aligned} x^2 + y^2 &= 1 \\ (x^2 + y^2)' &= 1' \\ (x^2)' + (y^2)' &= 1' \\ 2x + 2yy' &= 0. \end{aligned}$$

If that looks confusing, then you could write explicitly that $y = f(x)$, so that the last 2 line of the calculation would look like

$$\begin{aligned} (x^2)' + (f(x)^2)' &= 1' \\ 2x + 2f(x)f'(x) &= 0, \end{aligned}$$

which is a straightforward application of the Chain Rule.

Now that we've done the differentiating to get $2x + 2yy' = 0$, it's easy to solve to get

$$y' = -\frac{x}{y}.$$

We wanted the derivative at the point $\left(\frac{3}{5}, \frac{4}{5}\right)$, which must therefore be

$$y' = -\frac{x}{y} = -\frac{3/5}{4/5} = -\frac{3}{4},$$

the same result we got by first solving for y and then differentiating.

A second example To see just how powerful the technique of Implicit Differentiation is, let's try a second example. Suppose we had a curve given by the equation

$$x^2y^5 - 3xy^2 + x - y + 2 = 0.$$

We are asked to find the slope of the tangent line to this curve at the point $(1, -1)$.

Now, the starting equation is a 5th degree equation for y. If it had been a quadratic for y, we could have used the Quadratic Formula to solve it. If it had been even an equation of degree 3 or 4, we could have used formulas like the Quadratic Formula due to the 16th Century Italian mathematicians Tartaglia and Cardano to solve it in terms of third and fourth roots; though the solutions would take several pages to write. We'd then have to differentiate them. But fifth degree equations are even worse. Before his death at 27, the Norwegian mathematician Abel proved that a general 5th degree equation cannot be solved in radicals—there is'nt a formula like the Quadratic Formula for solving equations of degree 5 or higher. Even stronger and much deeper results about solvability were sketched by Evariste Galois before his death in a duel at age 20, and have grown into a rich and beautiful discipline called Galois Theory; but I'm now getting distracted. The critical point for us right now is that I can't solve the equation above for y, and neither can you. Our initial approach to differentiating it therefore can be proven not to work.

With Implicit Differentiation, though, it's still easy. First, take the derivative of the equation, remembering that x is the variable and that y is a function of x. You get

$$2xy^5 + x^2(5y^4y') - 3y^2 - 3x(2yy') + 1 - y' = 0.$$

Now group the terms containing y', and solve.

$$\begin{aligned} \left(5x^2y^4 - 6xy - 1\right) y' &= -2xy^5 + 3y^2 - 1 \\ y' &= \frac{-2xy^5 + 3y^2 - 1}{5x^2y^4 - 6xy - 1}. \end{aligned}$$

We want to know the value of y' at the point $(1, -1)$, so we replace x with 1 and y with -1 and we get

$$y' = \frac{2+3-1}{5+6-1} = \frac{2}{5}.$$

Is this amazing, or what?

Did I just cheat? Implicit Differentiation is an amazing technique, but I have perhaps made it seem even more amazing than it is. I said that the equation $x^2y^5 - 3xy^2 + x - y + 2 = 0$ could not be solved for y. If this is so, then how was it possible to find the y coordinate of the point $(1, -1)$? Didn't I just cheat?

Not really. In saying that we can't solve $x^2y^5 - 3xy^2 + x - y + 2 = 0$, what I mean to say is that we can't find a formula for y as a function of x, which is what we need in order to be able to differentiate y. But if we know the value of x, say, $x = 1$, then the equation simplifies to

$$y^5 - 3y^2 - y + 3 = 0.$$

To find numerical values of y satisfying this equation, we could do someting as simple as graphing the function and zooming in on the points where it crosses the axis. This is way simpler than solving the original equation analytically. In fact, a graph of $f(y) = y^5 - 3y^2 - y + 3$ (with the y-axis horizontal) is shown in Figure 16. It's easy to see from the Figure that there is a root very close to $y = -1$, as well as two positive roots.

3.19.2 Related Rates

Implicit Differentiation is a technique for finding the derivative y' in the situation where the variable x and the function y are related by some equation. The idea of Related Rates is that if instead of a function and a variable, there are two functions related by some equation, then the derivatives of these functions are also related.

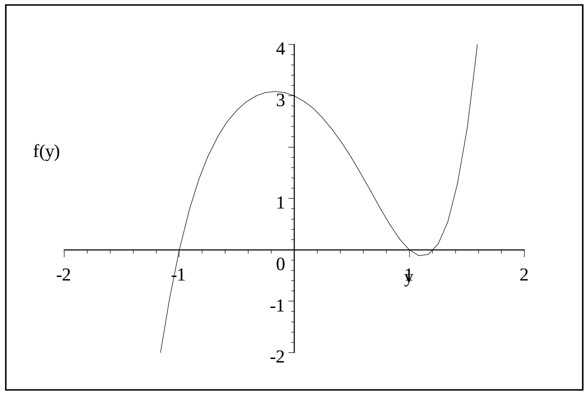

Figure 16: $f(y) = y^5 - 3y^2 - y + 3$.

Some questions Consider the following situation. A person 5 feet tall is in a lighted parking lot at night. The lot contains a single street light on a 15 foot tall pole. As the person walks away from the light, her shadow gets longer and longer. The person is not necessarily walking at a constant speed, but when she is 30 feet from the light, she is moving away from the light at $4\,\mathrm{ft/\,s}$. How fast is the length of the person's shadow growing at this moment, which is sketched in Figure 17?

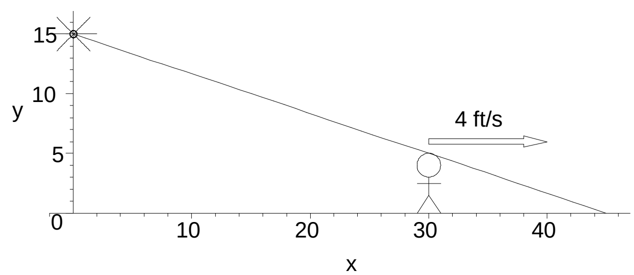

Figure 17: Street lights and shadows.

One can ask another, related question that I always find intriguing. Suppose the person's velocity does happen to be constant at $4\,\mathrm{ft/\,s}$ away from the light. Does the length of the shadow grow quickly when the person is close to the light and then more slowly later, or does it grow faster and faster as the person moves away from the light, or does it grow at a constant rate? Although I've walked away from an awful lot of street lights in my life, I find it rather hard to answer this question by thinking back on my experience. Indeed, I can imagine a variety of convincing sounding answers:

When the person is close to the street light, the angle from the person'r head to the lamp to the pole is increasing rapidly, which means the shadow is increasing rapidly. When the person is far from the light, this angle is hardly changing at all, so the length of the shadow probably isn't changing much. So the length of the shadow grows fast at first, then slowly.

Or try this: When the person is far from the light, you know the shadow is really long. It must be growing pretty fast to have gotten that long. Further, even a tiny change in the angle will make a huge difference in the length of a really long shadow. So the length of the shadow grows slowly at first, then fast.

Or maybe somehow everything cancels, and the rate of growth is a constant.

Which analysis, if any, is right?

Some answers Let's start with the first question, about the rate of growth of the shadow at the moment the person is 30 feet from the light. There are really two functions involved here: $p(t)$, the distance from the light post to the person at time t, and $s(t)$, the length of the shadow at time t. These functions are related by the fact that the trangles in the picture are similar. The small triangle, whose vertical side is the 5 foot person, has horizontal side $s(t)$. The large triangle, whose vertical side is the 15 foot light post, has horizontal side $p(t) + s(t)$. We therefore have

$$\frac{s(t)}{5} = \frac{p(t) + s(t)}{15}.$$

The problem tells us $p(t)$ and $p'(t)$ at a particular time t, and asks us to find $s'(t)$ at that time. To do this, we differentiate the equation above to get

$$\frac{s'(t)}{5} = \frac{p'(t) + s'(t)}{15}.$$

Since we know that at the moment in question, $p'(t) = 4\,\mathrm{ft/\,s}$, we have

$$\frac{s'(t)}{5} = \frac{4 + s'(t)}{15},$$

which solves to give $s'(t) = 2$. The shadow is growing at $2\,\mathrm{ft/\,s}$.

In fact, this calculation also answers the second question above, about how the rate of growth of the shadow changes with time. The expression we just got for $s'(t)$, namely $s'(t) = 2$, doesn't contain $p(t)$ at all. As long as the person keeps walking at $4\,\mathrm{ft/\,s}$, the analysis above is valid, and $s'(t) = 2\,\mathrm{ft/\,s}$. The rate of growth of the shadow is therefore a constant, regardless of how far the person is from the light.

Another example Related Rates seem to be a popular topic with authors of calculus texts, in part because they give rise to lots of entertaining word problems. Let me just do one more here.

Part of a class project not assigned by the professor, and not yet involving the professor, requires filling a really big water balloon. This is a word problem, not real life—and I'd like you to keep it that way—so the balloon is assumed to be a perfect sphere at all times. The tap from which the balloon is being filled produces 1 cubic foot of water per minute. How fast is the radius of the balloon increasing when the radius is 3 inches? 6 inches?

To solve this problem, you need to remember that the radius and volume of a sphere are related by the equation

$$V = \frac{4}{3}\pi r^3.$$

The problem seems to be considering both the radius and the volume as functions of time, so maybe we should really write

$$V(t) = \frac{4}{3}\pi r(t)^3.$$

If we differentiate everything in sight, we get

$$V' = \frac{4}{3}\pi(3r^2 r') = 4\pi r^2 r'.$$

We are told that $V' = 1\,\mathrm{ft}^3/\,\mathrm{min} = 1728\,\mathrm{in}^3/\,\mathrm{min}$. To find the rate of change of the radius when $r = 3\,\mathrm{in}$, we just plug in those values and solve:

$$\begin{aligned} 1728 &= 4\pi(9r') \\ r' &= \frac{1728}{36\pi} = \frac{48}{\pi} \doteq 15.279\,\mathrm{in}/\,\mathrm{min}. \end{aligned}$$

When $r = 6\,\mathrm{in}$, we have

$$\begin{aligned} 1728 &= 4\pi(36r') \\ r' &= \frac{1728}{144\pi} = \frac{12}{\pi} \doteq 3.8197\,\mathrm{in}/\,\mathrm{min}. \end{aligned}$$

It's a good exercise in the development of your intuition to see if you can convince yourself intuitively that it makes sense that the rate of change of the radius should be less when $r = 6$ than when $r = 3$. Why should it be exactly one quarter as great?