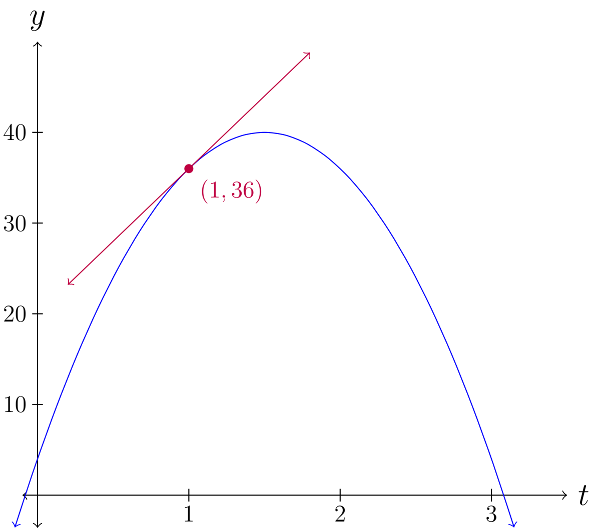
# Lecture 7: introduction to derivatives

Calculus I, section 10
September 26, 2023

In the worksheet for today's class, we looked at the example from the very beginning of the course, where we asked about the speed of a ball one second after being thrown upwards, i.e. the slope of the line tangent to this graph at $t = 1$:




We found that this slope, or speed, is given by

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

where $f(t) = -16t^2 + 48t + 4$, and then evaluated the limit to be 16.

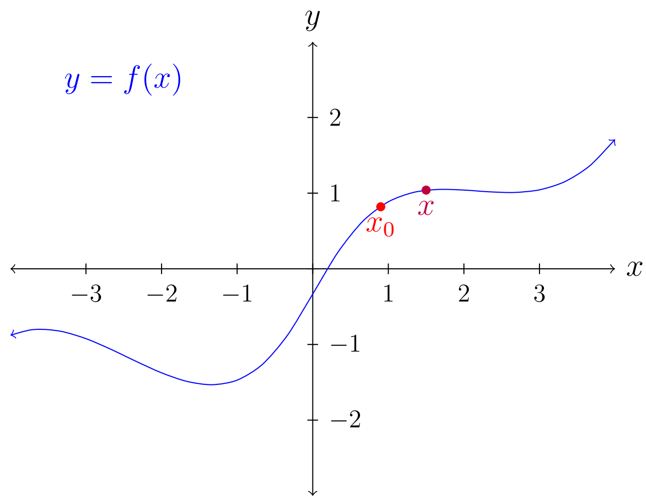
We might write this as $f'(1)$, the instantaneous velocity at $t = 1$. More generally, for any value $t_0$ we can find the instantaneous velocity at $t_0$ by a similar formula:

$$f'(t_0) = \lim_{t\to t_0} \frac{f(t) - f(t_0)}{t - t_0}.$$

This is the derivative!

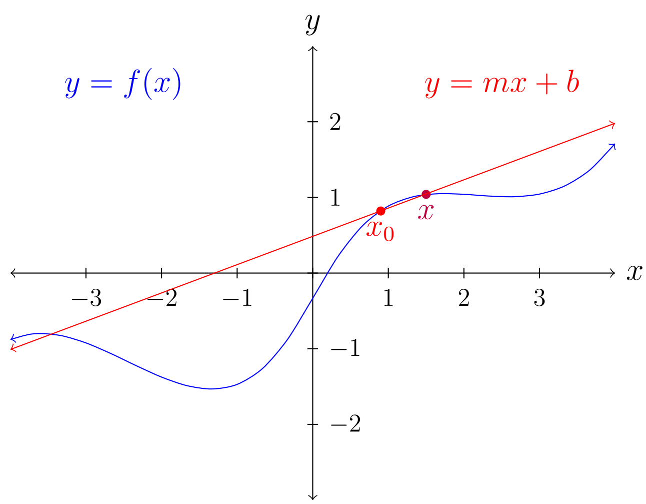
There are a few ways we can usefully think about derivatives. One, as we've seen, is the instantaneous rate of change: when the function $f(t)$ is measuring position with respect to time, then this rate of change is the speed. In general, it can be other things; for example, if $f$ is measuring speed, then the rate of change of the speed is the acceleration. Either way, we can find this formula by thinking about measuring the slope of a line connecting $(t_0, f(t_0))$ and $(t, f(t))$, which (as rise over run) is $\frac{f(t)-f(t_0)}{t-t_0}$, and then take the limit as $t \to t_0$ to get the slope of the tangent line at $t_0$.

Alternatively, we could think of it as giving a formula for "first-order approximation." To say what this means, let's pose another question: if we know the value of $f$ at some point $x_0$, how can we approximate $f(x)$ for $x$ very near $x_0$?




The simplest approximation—called the "zeroth-order approximation"—would be to simply say: if $x$ is very near $x_0$, then (so long as $f$ is continuous at $x_0$) we expect that $f(x)$ should be near $f(x_0)$, i.e. $f(x) \approx f(x_0)$.

A better approach would be to take into account the rate of change near $x_0$. In other words, if we said that near $x_0$, our function is approximated by some line $y = mx + b$, then since $f$ is approximately linear $f(x) - f(x_0)$ is approximately $m(x - x_0)$ and so we get the "first-order approximation" (also called linear approximation) $f(x) \approx f(x_0) + m(x - x_0)$.




(This is called a first-order approximation because the right-hand side is a degree 1 polynomial in $x$, or more precisely in $x - x_0$; taking the constant (zero-order) term gives $f(x) \approx f(x_0)$, the zeroth-order approximation.)

In order for this to be useful, we need some way of finding this number $m$. To find it, we just solve for $m$ in the approximation above:

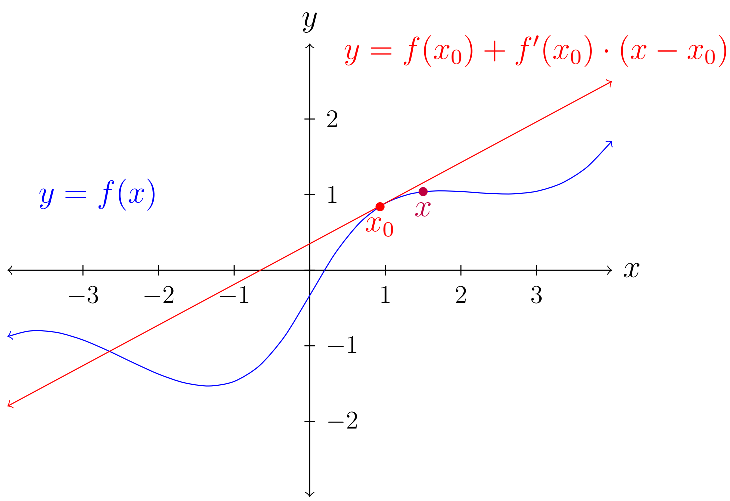
$$m \approx \frac{f(x) - f(x_0)}{x - x_0}.$$

This approximation gets better and better as $x$ gets closer to $x_0$, so to remove the dependence on $x$ (to make sure $m$ is a constant) we take the limit:

$$m = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0},$$

which is the same as the formula we found before for $f'(x_0)$, the derivative of $f$ at $x_0$. Therefore the full first-order approximation formula is

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$




The notation $f'(x_0)$ suggests that we can think of the derivative at a point $x_0$ as a value of a whole new function $f'$, which we form from $f$. This is true: the derivative is an operation that takes in a function $f(x)$ and outputs a new function $f'(x)$. To avoid confusion with $x$ and $x_0$, let's introduce a new formulation: suppose give the distance between $x$ and $x_0$ a name, say $h = x - x_0$, so that we expect $h$ to be very small. Then we could rewrite $x$ as $x_0 + h$, so that the derivative is

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{x_0 + h - x_0} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

Since there's now no $x$ in the picture, just $x_0$ and $h$, we can rename $x_0$ to $x$ to write

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

These two formulations for the derivative are equivalent; we'll use whichever is more convenient.

Let's try some examples. Suppose $f(x) = c$ is constant. Then

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x \to x_0} \frac{c - c}{x - x_0} = 0,$$

so all constant functions have derivative everywhere 0.

Next, suppose we have a linear function $f(x) = mx + b$. Then

$$\begin{aligned}
f'(x_0) &= \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \\
&= \lim_{x \to x_0} \frac{mx + b - mx_0 - b}{x - x_0} \\
&= \lim_{x \to 0} \frac{m(x - x_0)}{x - x_0} \\
&= m.
\end{aligned}$$

In particular if $f(x) = x$, then $f'(x) = 1$.

Next take $f(x) = x^2$. Let's use the formulation with $h$ this time:

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

We're now in a position to apply our linear approximation framework: say we want to approximate the value of $2.01^2$. We know that $2^2 = 4$, so we let $x_0 = 2$ and $x = 2.01$; then linear approximation tells us that

$$f(2.01) = 2.01^2 \approx f(2) + f'(2)(2.01 - 2),$$

where $f(x) = x^2$. We just computed that $f'(x) = 2x$, so $f(2) = 2^2 = 4$ and $f'(2) = 2 \cdot 2 = 4$, so this is

$$2.01^2 \approx 4 + 4 \cdot 0.01 = 4.04.$$

The true answer is 4.0401, so this is pretty good, and certainly significantly better than the zeroth-order approximation $2.01^2 \approx 2^2 = 4$ (already not bad in this case).

Before computing more examples, let's observe some properties of derivatives. We've already said this is an operator on functions that takes in $f(x)$ and produces $f'(x)$. For convenience, it's sometimes useful to have an "operator notation": given a function $f(x)$, we write $\frac{d}{dx} f(x)$, or just $\frac{df}{dx}$, for $f'(x)$, so $\frac{d}{dx}$ is the differentiation operator. Note that this is not literally a fraction! It's supposed to mean the ratio of an infinitesimal change $df$ in $f$ relative to the corresponding infinitesimal change $dx$ in $x$; but one cannot "cancel the $d$'s" or treat it like a fraction in most ways, it's purely notation.

Now we can note two properties of the derivative. One is that for any function $f(x)$ and constant $c$, the derivative of $c \cdot f(x)$ is $c\dot{f}'(x)$ (if it exists). So in operator notation:

$$\frac{d}{dx} cf(x) = c \cdot \frac{d}{dx} f(x).$$

The second is that for any two functions $f(x)$ and $g(x)$, the derivative of $f(x) + g(x)$ is $f'(x) + g'(x)$ (if both exist). So in operator notation,

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

These two properties together are called linearity: the operator $\frac{d}{dx}$ commutes with scalar multiplication and distributes over addition. One can check these properties from the definition of the derivative, together with limit laws.

Note also that these two properties together imply some others not explicitly stated. For example, we get a similar rule for subtraction, because

$$\begin{aligned}
\frac{d}{dx}(f(x) - g(x)) &= \frac{d}{dx}(f(x) + (-1) \cdot g(x)) \\
&= \frac{df}{dx} + \frac{d}{dx}(-1) \cdot g(x) \\
&= \frac{df}{dx} + (-1) \cdot \frac{dg}{dx} \\
&= \frac{df}{dx} - \frac{dg}{dx}.
\end{aligned}$$

For linear functions, this gives us another way to find the derivative:

$$\frac{d}{dx}(mx + b) = \frac{d}{dx}(mx) + \frac{d}{dx} b = m \cdot \frac{d}{dx} x + \frac{d}{dx} b,$$

so it's enough to know that $\frac{d}{dx} x = 1$ and $\frac{d}{dx} b = 0$ for any constant $b$; then we get $\frac{d}{dx}(mx+b) = m$ without any further computation.

Similarly, any quadratic function $ax^2 + bx + c$ is formed from $x^2$, linear functions, and scalar multiplication:

$$\frac{d}{dx}(ax^2 + bx + c) = a \cdot \frac{d}{dx} x^2 + b \cdot \frac{d}{dx} x + \frac{d}{dx} c = 2ax + b.$$

Recall our starting problem $y = f(t) = -16t^2 + 48t + 4$. We computed $f'(1)$ via a particular limit, but now we can compute it in general:

$$y' = \frac{dy}{dt} = f'(t) = -32t + 48.$$

Evaluating at $t = 1$ recovers $f'(1) = 16$.

Okay, so we know the derivatives of constants, of $x$, and of $x^2$, and we can use these (together with the linearity of the derivative) to compute derivatives of linear and quadratic functions. To compute the derivatives of all polynomials, we'd need to know the derivatives of $x^n$ for higher $n$. How can we do this?

Let's start with an example: $\frac{d}{dx}x^3$. To compute this, we first expand out

$$(x+h)^3 = x^3 + 3hx^2 + 3h^2x + h^3,$$

so

$$\begin{aligned}
\frac{d}{dx}x^3 &= \lim_{h\to 0}\frac{(x+h)^3 - x^3}{h} \\
&= \lim_{h\to 0}\frac{3hx^2 + 3h^2x + h^3}{h} \\
&= \lim_{h\to 0}(3x^2 + 3hx + h^2) \\
&= 3x^2.
\end{aligned}$$

This suggests a pattern:

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

To determine if this pattern is correct, we need a more general version of the expansion of $(x+h)^3$ above: the binomial theorem.

**Theorem** (Binomial theorem)**.** *For any real numbers $x$ and $y$ and any positive integer $n$, we have*

$$(x+y)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n-1}xy^{n-1} + \binom{n}{n}y^n,$$

*where $\binom{n}{k}$ are the binomial coefficients.*

To make use of this theorem, we need to know some things about the binomial coefficients $\binom{n}{k}$. There is a formula for these: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. It has a combinatorial interpretation: $\binom{n}{k}$ is the number of ways of choosing $k$ items out of $n$ options, e.g. there are $\binom{4}{2} = 6$ ways of choosing two students out of a group of four. They are the entries of Pascal's triangle, with $n$ corresponding to the row and $k$ the column:

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 \\
\end{array}$$

$$\begin{array}{cccccccc}
\iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

There are many fun properties of these numbers (e.g. in Pascal's triangle, each entry can be formed as the sum of the two numbers above it, and of course everything is symmetric: $\binom{n}{k} = \binom{n}{n-k}$), but for our purposes let's just make a couple of observations: $\binom{n}{0} = 1$ and $\binom{n}{1} = n$, for every $n$. These make sense combinatorially: there is only one way to pick no objects, and exactly $n$ ways to pick one out of $n$ objects. With that in hand, the binomial theorem tells us that

$$(x+y)^n = x^n + nx^{n-1}y + y^2 \cdot (\cdots),$$

where $(\cdots)$ is some polynomial in $x$ and $y$.

In our case of interest, we're looking at $(x+h)^n$ where $h$ is very small. Therefore $h^2$ is so small as to be negligible, so the idea is that we'll be able to approximate $(x+h)^n \approx x^n + nx^{n-1}h$. More precisely, we can compute:

$$\begin{aligned}\frac{d}{dx}x^n &= \lim_{h\to 0}\frac{(x+h)^n - x^n}{h}\\ &= \lim_{h\to 0}\frac{nx^{n-1}h + h^2(\cdots)}{h}\\ &= \lim_{h\to 0}\left(nx^{n-1} + h(\cdots)\right)\\ &= nx^{n-1}\end{aligned}$$

as $h(\cdots)$ will tend to 0 as $h \to 0$ no matter what $(\cdots)$ is (as it's a polynomial in $x$ and $h$, so will converge to some value as $h \to 0$). This is the power rule: if $f(x) = x^n$ for some positive integer $n$, then $f'(x) = nx^{n-1}$.

We've seen some special cases of this already. If $n = 0$, then $x^0$ is just the constant 1 (ignoring the $x = 0$ discontinuity), and so its derivative is $\frac{d}{dx}x^0 = 0 \cdot x^{0-1} = 0$, just as we found earlier. If $n = 1$, then $x^1 = x$ and we have $\frac{d}{dx}x^1 = 1 \cdot x^0 = 1$, just as above; and if $x = 2$ then $\frac{d}{dx}x^2 = 2x^1 = 2x$ again, and similarly $\frac{d}{dx}x^3 = 3x^2$.

As it turns out, the power rule turns out to be true for all real numbers $n$, not just positive integers. For example,

$$\frac{d}{dx}\sqrt{x} = \frac{d}{dx}x^{1/2} = \frac{1}{2}x^{-1/2} = \frac{1}{2\sqrt{x}},$$

or

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

The proof above doesn't really work more generally, but we'll get some tools next time which will help.

Just like for quadratics, knowing the derivatives of all the $x^n$ together with linearity lets us differentiate all polynomials! For example, say $f(x) = x^7 - 4x^3 + x + 2$. By linearity,

$$f'(x) = \frac{d}{dx}x^7 - 4 \cdot \frac{d}{dx}x^3 + \frac{d}{dx}x + \frac{d}{dx}2,$$

which by the power rule is

$$f'(x) = 7x^6 - 12x^2 + 1.$$

We can also iterate differentiation: the second derivative $f''(x)$ is the derivative of the derivative, i.e. $f''(x) = \frac{d}{dx}f'(x) = \frac{d}{dx}\left(\frac{d}{dx}f(x)\right)$, which is often written as $\frac{d^2}{dx^2}f(x)$ (again, purely notation). In our physics example before where $y = f(t)$ is position and $t$ is time, the first derivative $\frac{dy}{dt} = f'(t)$ is the velocity; the second derivative $\frac{d^2y}{dt^2} = f''(t)$ is the change in velocity, i.e. the acceleration.

We're now in a position where we can derive the equation of motion $y = f(t) = -16t^2 + 48t + 4$ from some basic laws of physics. Newton's second law is $F = ma$, where $F$ is the force exerted on our object (in this case the ball), $m$ is its mass, and $a$ is its acceleration. If $y = f(t)$ is the position of the ball, then we know that the acceleration $a = f''(t)$ is given by the second derivative, so using Newton's second law we get $f''(t) = \frac{F}{m}$. Our force is gravity, which is proportional to the mass; on the surface of the earth, the ratio $\frac{F}{m}$ is always a constant, about $-32$ feet per second squared (negative because gravity points down). Therefore $y = f(t) = -16t^2 + 48t + 4$ because $f(0) = 4$ gives the initial position of the ball, at four feet above the ground; $f'(t) = -32t + 48$, so $f'(0) = 48$, the initial velocity of the ball at 48 feet per second upwards; and $f''(t) = -32$, so the ball is undergoing constant acceleration of $-32$ feet per second squared downwards.