Last time, we talked about limits for two purposes: making sense of the values of power series, and for understanding functions at points where they might not be defined. Today, we will make use of limits to define the derivative. This mainly involves limits in the second sense.

Lots of functions don't have nice formulas that we can write down and evaluate (most of the time, only polynomials have this good property). To work around this, we can find polynomial approximations to our functions. Adding more and more terms to these polynomials gives us more flexibility to produce better approximations, but as we saw with power series, such approximations tend to only behave well for sufficiently small input values.

If we're given a function $f(x)$, we might hope that it can be approximated well by polynomials, or even a power series, for small values of x; we might say that we are *approximating* $f(x)$ *near zero.*

The "zeroth" approximation one might consider is just a constant. If we want to approximate $f(x)$ near zero, the only sensible choice is the constant $f(0)$, which is exact at $x = 0$, and hopefully not too bad of an estimate when x is very small.

We could refine this to a to consider is by a degree 1 polynomial, or in other words a linear approximation of the form $f(0) + mx$. These can be quite accurate for very small values of x - if you think about "zooming in" on a curve like $x^2 + 2x + 1$ near $x = 0$, it looks less and less curved as one increases the magnification, and eventually looks like a line with slope 2, and so that's the value of m that gives the best approximation. This value is exactly the derivative!

The *derivative* of $f(x)$ at zero is the slope of the best linear approximation to $f(x)$ near zero. This quantity is written $f'(0)$. To be precise, we want to be able to write

$$f(x) = f(0) + f'(0)x + \text{error}(x)$$

The error should be small near zero, and in particular it shouldn't overwhelm the linear term - otherwise we picked the wrong slope for $f'(0)$ and it should be changed. It will almost always suffice to think of the error on a purely intuitive level – everything you would expect to be true is (adding small errors is just a small error, the product of small errors with anything is a small error, etc). We can make this quantitative with limits, although we won't often worry about this:

$$\lim_{x\to 0} \frac{\text{error}(x)}{x} = 0$$

which means that for x very small, the error at x should be much much smaller than x. So if we know $f(0)$ and $f'(0)$, we should be able to extrapolate nearby values of $f(x)$ from the line $f(0) + f'(0)x$.

It is easy to see that for a polynomial, the above condition on the error is satisfied by every term with degree bigger than 1.

Warning: to ask for number $f'(0)$ to exist is very very restrictive. There are all kinds of functions for which it's not possible to pick such a constant and have a well-behaved error term. However, in this course we won't tend to work with such functions!

Naturally, we might want to define linear approximations and derivatives at points besides zero. The template is the same: if we want the find the derivative at a, denoted $f'(a)$, then we want to be able to write

$$f(x) = f(a) + f'(a)(x-a) + \text{error}(x)$$

where the error should be very small *near* a. Quantitatively, this means

$$\lim_{x \to a} \frac{\text{error}(x)}{x-a} = 0$$

The reason we now have $x - a$ on the right is that this is the distance from our point of interest, x, to the point that we have information about, a. We want our approximation to be accurate near a, or in other words when $x - a$ is small.

Another way to view this is to observe that linear approximations should be preserved by translation. So if we move a to zero, then the derivative at 0 in these new coordinates should be the derivative at a in the old coordinates. In other words, $f'(a) = g'(0)$ where $g(x) = f(x+a)$. This is (for now) the easiest way to calculate derivatives.

Let's work out some simple examples with a polynomial. Let $f(x) = x^3 + 2x^2 - 4x + 1$.
To find the derivative at zero, notice we can group the terms of the polynomial as follows:

$$f(x) = 1 - 4x + (2x^2 + x^3)$$

Comparing with the definition of the derivative,

$$f(x) = f(0) + f'(0)x + \text{error}(x),$$

we see that this is already in the correct form if we take the error to be $2x^2 + x^3$, and this error satisfies our "smallness" requirement:

$$\lim_{x \to 0} \frac{2x^2 + x^3}{x} = \lim_{x \to 0} 2x + x^2 = 0$$

So the derivative $f'(0)$ is just the coefficient of x, which is -4.

What if we wanted to calculate $f'(1)$? Well, let's use the translation trick mentioned above: if we let $g(x) = f(x+1)$, then $g'(0)$ should equal $f'(1)$. Since $g(x)$ is a polynomial, it should fairly straightforward. Let's expand out $g(x)$:

$$g(x) = f(x+1) = (x+1)^3 + 2(x+1)^2 - 4(x+1) + 1 = x^3 + 5x^2 + 3x$$

Looking at the definition,

$$g(x) = \underbrace{0}_{constant} + \underbrace{3x}_{g'(0)x} + \underbrace{x^3 + 5x^2}_{error}$$

and so $f'(1) = g'(0) = 3$.