Why limits?
Math 120 Calculus I
Fall 2015

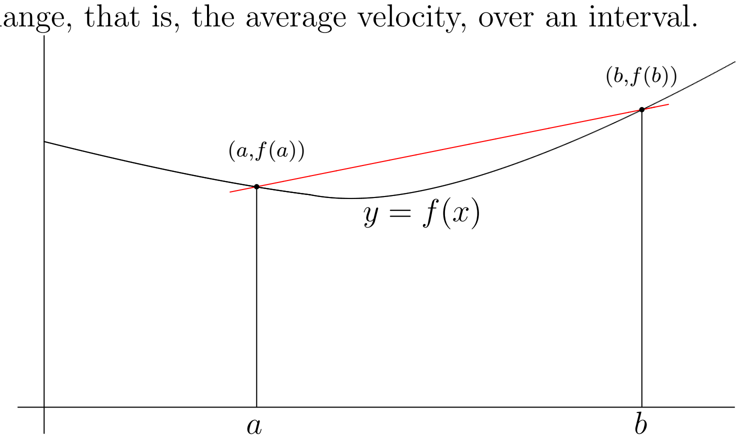
Why limits? We'll spend the next few weeks studying "limits." Naturally, the question is "why limits?" Why not just go on to derivatives?

The answer involves the character of the course. This is not just a course about how to use calculus, but a mathematics course about what calculus is.

How do derivatives depend on limits? Just how are derivatives supposed to depend on limits? A derivative is supposed to be the rate of change of a function at an instant, what we'll call "instantaneous rate of change." On the face of it, that makes no sense at all, since an instant is a point in time, and nothing changes at a point in time. You need a time interval for anything to change.

Start with a function $y = f(x)$. To help us understand, let's take x to be time, measured in some convenient time unit, and let's take $y = f(x)$ to be the distance travelled at time x, measured in some convenient distance unit. Then the derivative is what we know as velocity. The velocity doesn't have to be constant, but may change over time. It might slowly at first with a small velocity, and later quickly with a large velocity, or vice versa. We're trying to determine how to find the velocity $f'(x)$ when we know the distance $f(x)$. Finding the derivative $f'(x)$ when you know $f(x)$ is called *differentiation*.

Average rates of change and slopes of secant lines. We can fairly easily compute the average rate of change, that is, the average velocity, over an interval.



Suppose we take the time interval $[a, b]$ which starts at time $x = a$ and ends at time $x = b$. We can compute the distance travelled over that interval as the difference $f(b) - f(a)$. But the length of the time interval is $b-a$. That says the object travels a distance of $f(b)-f(a)$ units over a time interval of length $b-a$. Therefore, the average rate of change is

$$\frac{f(b)-f(a)}{b-a}.$$

If a quantity changed uniformly over the interval $[a,b]$, this is the rate of change necessary for the dependent variable $y=f(x)$ to change from $f(a)$ to $f(b)$.

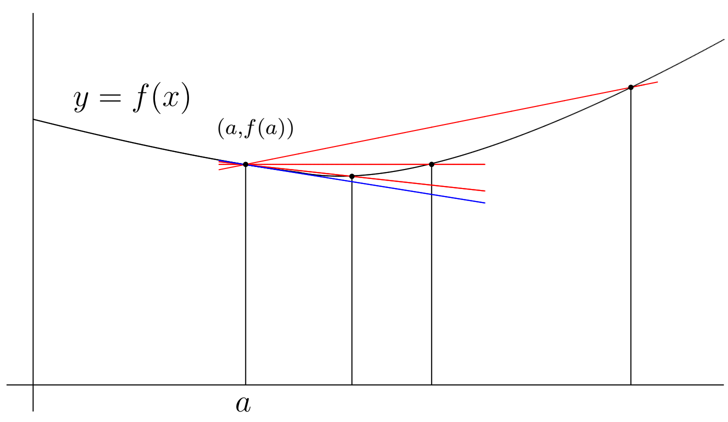
This average rate of change is the slope of the secant line connecting the point $(a,f(a))$ to the point $(b,f(b))$ on the graph $y=f(x)$ of the function.

It's called a *secant line* because the line cuts across the graph at these two points, *secantem* being Latin for "a cutting." The term has been used in mathematics since 1583. The slope of this line is just the change in y, namely $f(b)-f(a)$, divided by the change in x, namely, $b-a$.

Instantaneous rates of change and slopes of tangent lines, Fermat's method. The rate of change at a point $x=a$, which we'll call the *derivative* and denote $f'(a)$, cannot be determined just by the value of function at that point. That is, in order to know how f is changing at $x=a$, it's not enough just to know $f(a)$. You need to know the values of f near $x=a$ as well.

In the 1620s Fermat developed algorithms to find rates of change at points, which he identified with slopes of tangent lines. The idea was that if the two points where a secant line crossed the curve were to get closer together and coincide, then the secant line would become a tangent line.

Imagine in the figure that the point a is fixed, but the point b moves closer to a and eventually coincides with a.



As b approaches a, the red secant lines get closer to the curve $y=f(x)$ near $x=a$, that is, they approach the blue tangent line at $x=a$. Unlike the secant lines, the tangent line doesn't cross the curve but only touches it. The word "tangent" comes from the Latin *tangentem* for "a touching." It's been used in mathematics since 1583.

To illustrate his method, take the function $f(x)=x^2$, the simplest nonlinear function. Its graph is a parabola. First compute the average rate of change over an interval $[a,b]$. You'll get

$$\frac{f(b)-f(a)}{b-a}=\frac{b^2-a^2}{b-a}.$$

You can factor the numerator as $(b+a)(b-a)$ then cancel the factor $b-a$ with the denominator to conclude that the average rate of change over an interval $[a, b]$ is $b+a$. Following Fermat's method, to get the instantaneous rate of change at a point, just set b to equal a. Then the instantaneous rate of change at the point a is $2a$.

What's the problem with this method? We divided by $b - a$, but since $b = a$, that means we divided by 0.

Nonetheless, the method works. To justify it, you have to recognize that you actually take a limit as b approaches a.

Limits. Newton also recognized that the instantaneous rate of change at an instant $x = a$ can be defined as the limit of the average rates of change over an interval $[a, b]$ as the length of the interval goes to 0, that is, as b approaches a. Long after Newton, a special notation has evolved to abbreviate this expression. Namely,

$$\lim_{b \to a}$$

stands for "the limit as b approaches a." With that notation we can symbolically express the instantaneous rate of change at $x = a$ as

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

Note for the limit $\lim\limits_{b \to a}$ that b is never taken to equal a. That would lead to division by 0.

But what, exactly is a limit? Newton founded his calculus on intuitive concepts of limits. He recognized that an instantaneous rate of change could be found as a limit of rates of change over shorter and shorter intervals. Newton was a little fuzzy about what limits were. Here's what he said in 1687. "The ultimate ratio of evanescent quantities ... [are] limits towards which the ratios of quantities decreasing without limit do always converge; and to which they approach nearer than by any given difference, but never go beyond, nor in effect attain to, till the quantities are diminished *in infinitum*."

He didn't actually define what limits were, nor did he state the properties that he expected limits to have.

Various mathematicians over the centuries clarified the concept, so by the 1800's it was pretty well understood. In the 1820's Cauchy propounded a method to compute a limit based on considering all small positive quantities ϵ (an error factor) and showing that the quantity approaching the limit eventually stays within ϵ of the limit.

That definition is what we'll develop at the beginning of chapter 2. Once we have that definition, we'll be able to find various properties of limits. When we finish chapter 2, we'll define derivatives using them in chapter 3.

Leibniz' alternative. Newton and Leibniz both developed the subject of calculus in the late 1600s. Both created rules for dealing with derivatives and integrals, rules that lead to the word "calculus" for the whole subject, but neither had a satisfactory understanding of the basis of their theory.

Leibniz rested his calculus on the concept of infinitesimals. Infinitesimals were supposed to be positive quantities less than any positive number. His theory required not just infinitesimals, but infinitely many orders of infinitesimals. He needed second-order infinitesimals infinitely smaller than any of the first-order infinitesimals, and third-order infinitesimals infinitely smaller than any of the second-order infinitesimals, and so forth. The foundations of Leibniz' infinitesimals were not logically justified until the middle of the 20th century.

What we're going to do is develop foundations of limits, the kind that Newton used only intuitively. That is going to take some time.

Math 120 Home Page at `http://math.clarku.edu/~ma120/`