PROFESSOR SMITH MATH 295 LECTURE NOTES

BY JOHN HOLLER

1. November 17: The Chain Rule

The chain rule tells us how to differentiate a composition of functions:

Theorem 1.1. *Consider a composition of functions* $\mathbb{R} \xrightarrow{g} \mathbb{R} \xrightarrow{f} \mathbb{R}$*. If* g *is differentiable at* a *and* f *is differentiable at* $g(a)$*, then* $f \circ g$ *is differentiable at* a *and*

$$(f \circ g)'(a) = f'(g(a))g'(a).$$

Example: Consider the function $h(x) = \sin^2 x$. This function is the composition of the functions $g(x) = \sin x$ and $f(y) = y^2$. Because both f and g are differentiable at every point of their domains, Theorem 1.1 tells us so is their composition. The chain rule tells us that for all x,

$$h'(x) = 2(\sin x)(\cos x).$$

The chain rule says that the derivative of a composition is the *product* of the derivatives of the two functions at the appropriate element of their domains. Why should this be true?

1.1. The Soul of a Differentiable function.

A linear function $L : \mathbb{R} \to \mathbb{R}$ is one that has a constant rate of change. Every linear function is described by a formula of the form

$$L(x) = m(x - a) + L(a),$$

where $m \in \mathbb{R}$ is the constant rate of change of L (or slope) and $a \in \mathbb{R}$ is any input value. The graph of a linear function L is a *line* of slope m passing through the point $(a, L(a))$. Of course, since a line is uniquely determined once we know its slope and one point on it, we can take a to be any input value in this expression. In high school it is common to take $a = 0$ and write b instead of $L(a)$. The formula becomes $L(x) = mx + b$, where b is the "y-intercept" of L. Although this is a nice compact way to write the formula for a linear function, it is not always the most convenient.

Consider a function f which is differentiable at a point a in its domain. The true intuitive meaning of this: in a sufficiently tiny neighborhood of a, infinitesimally close to the point a, the function f is, *deep down in its soul*, a *linear* function. When we draw the tangent line to the graph of f at a, we are drawing the graph of this linear function. This is how the founders of calculus imagined the true meaning of differentiability. So how to make this precise?

Date: Fall 2010.

First, take

$$f : \mathbb{R} \to \mathbb{R},$$

and look at the point a in the domain. The point is that when f is differentiable at a, the *linear function*

$$\begin{aligned} L_a(f) &: \mathbb{R} \to \mathbb{R} \\ x &\mapsto f'(a)(x-a) + f(a). \end{aligned}$$

approximates f very well in a neighborhood of a. This linear function has slope $f'(a)$, and sends a to $f(a)$. So the graph of this linear function is a line of slope $f'(a)$ passing through the point $(a, f(a))$—that is, the line tangent to the graph of f at a. Near a, the linear function $L_a(f)$ approximates f within any range of accuracy, as long as we shrink down to a sufficiently tiny neighborhood of a. The precise mathematical formulation of this idea is:

Proposition 1.2. *A function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at a point a if and only if*

$$f(x) = \Delta(x-a) + f(a) + E(x),$$

where Δ is some real number and $E(x)$ is a function satisfying $\lim_{x\to a} \frac{E(x)}{x-a} = 0$.

Here, $L(x) = \Delta(x-a) + f(a)$ is the linear function approximating f near a and the term $E(x)$ is the error term—for each x in the domain of f, $E(x)$ is the difference between the output of f and the output of its linear approximation at that x. (Of course Δ turns out to be $f'(a)$ but only after we know $f'(a)$ exists.) The function $E(x)$ is defined for all points in the domain of f, but of course, it may be very large if we input elements far from a, because the linear function $L_a(f)$ only reasonably approximates f in a small neighborhood of a.

When a function is approximated by a linear function, the error term $E(x)$ goes to zero as x approaches a. But the claim that also $\frac{E(x)}{x-a}$ approaches zero means that the error term *quickly* approaches zero. The estimates given by the linear function get more accurate as approach a. Formally, we have that if

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

then also

$$\lim_{x\to a} E(x) = 0.$$

Indeed,

$$\lim_{x\to a} E(x) = \lim_{x\to a} \frac{E(x)}{x-a}(x-a) = [\lim_{x\to a} \frac{E(x)}{x-a}][\lim_{x\to a} x-a] = 0 * 0 = 0.$$

So of course the statement that $\frac{E(x)}{x-a}$ approaches zero as inputs approach a means in particular that the error term approaches zero.

Proof of Proposition. First, suppose that f is differentiable at a. This means

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

Now we can define

$$E(x) = f(x) - [f'(a)(x-a) + f(a)].$$

We need only show that $\lim_{x\to a}\frac{E(x)}{x-a} = 0$, and the implication follows setting $\Delta = f'(a)$. This is easy:

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

which when expanded yields

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

as needed.

Conversely, say $f(x) = \Delta(x-a) + f(a) + E(x)$ where $\lim_{x\to a}\frac{E(x)}{x-a} = 0$. We have:

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

This shows the limit exists, so f is differentiable at a and its derivative is Δ. □

1.2. **Proof of the Chain Rule.** Remembering that intuitively, a differentiable function is like a linear function near a, let's try to understand why the chain rule is true first for linear functions.

Say we have a composition of linear functions

$$\mathbb{R}\xrightarrow{g}\mathbb{R}\qquad\qquad \mathbb{R}\xrightarrow{f}\mathbb{R}$$

$$x\mapsto mx+b;\quad y\mapsto ny+c.$$

So for their composition we simply substitute to achieve:

$$f\circ g(x) = n(mx+b)+c = nmx+nb+c.$$

Now, note simply that the rate of change (slope) of the composition $f\circ g$ is the product of the rates of change of f and g. In other words, the derivative of the composition is the product of the derivatives! This is exactly what the Chain Rule says, in this special case of linear functions!

Now since *all differentiable functions* are secretly linear at the infinitessimal level, the reason the chain rules holds is (at least intuitively) clear! To prove it precisely, we just approximate the functions by linear functions, and show that the error term goes to zero as x approaches a.

Proof of the Chain Rule. Using Proposition 1.2, we approximate g by a linear function near a and f by a linear function near $g(a)$:

$$(\heartsuit)\qquad g(x) = g'(a)(x-a)+g(a)+E_g(x)$$

$$(\diamondsuit)\qquad f(y) = f'(g(a))(y-g(a))+f(g(a))+E_f(y).$$

We now compute $(f\circ g)'(a) = \lim_{x\to a}\frac{f(g(x))-f(g(a))}{x-a}$. Substituting the formula $\diamondsuit$ for $f(y)$ evaluated at $y=g(x)$ we have

$$\lim_{x\to a}\frac{f'(g(a))(g(x)-g(a))+f(g(a))+E_f(g(x))-f(g(a))}{x-a},$$

which simplifies to

$$\lim_{x\to a}\frac{f'(g(a))(g(x)-g(a))+E_f(g(x))}{x-a}.$$

Now substituting ♡ for $g(x)$, we have

$$\lim_{x\to a}\frac{f'(g(a))((g'(a)(x-a)+g(a)+E_g(x))-g(a))+E_f(g(x))}{x-a},$$

which simplifies to

$$\lim_{x\to a}\frac{f'(g(a))((g'(a)(x-a)+E_g(x)))+E_f(g(x))}{x-a}$$

and then to

$$\lim_{x\to a}f'(g(a))g'(a)+\lim_{x\to a}\frac{f'(g(a))E_g(x)}{x-a}+\lim_{x\to a}\frac{E_f(g(x))}{x-a}.$$

Note that, substituting $y=g(x)$, and using the fact that $\lim_{x\to a}g(x)=g(a)$, we have

$$\lim_{x\to a}\frac{E_f(g(x))}{x-a}=\lim_{y\to g(a)}\frac{E_f(y)}{y-g(a)}.$$

Thus, both of the limits involving the error terms E_f and E_g go to zero, by the hypothesis on the error terms E_f and E_g provided by Proposition 1. Finally, we conclude that the limit

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

This shows that $f\circ g$ is differentiable at a and takes the form promised by the chain rule. □