CLARK UNIVERSITY 1887

The chain rule, part 1

Math 131 Multivariate Calculus

D Joyce, Spring 2014

The chain rule. First the one you know. We'll start with the chain rule that you already know from ordinary functions of one variable. It tells you how to find the derivative of the composition $A \xrightarrow{f\circ g} C$ of two functions, $\mathbf{R} \xrightarrow{f} \mathbf{R}$ and $\mathbf{R} \xrightarrow{g} \mathbf{R}$. This composition is defined by $(f\circ g)(x) = f(g(x))$ and illustrated by the commutative diagram

$$\begin{tikzcd} \mathbf{R} \arrow[r, "g"] \arrow[dr, "f\circ g"'] & \mathbf{R} \arrow[d, "f"] \\ & \mathbf{R} \end{tikzcd}$$

When we generalize this, we'll use a slightly different notation. We'll let the independent variable be t instead of x, but we'll use x instead of g, so x is a function of t. Thus, $f \circ x$ is defined by $(f \circ x)(t) = f(x(t))$.

$$\begin{tikzcd} \mathbf{R} \arrow[r, "x"] \arrow[dr, "f\circ x"'] & \mathbf{R} \arrow[d, "f"] \\ & \mathbf{R} \end{tikzcd}$$

Using these symbols for the functions and using the prime notation for derivatives, the chain rule says

$$(f \circ x)'(t) = f'(x(t))\, x'(t).$$

We'll have partial derivatives, so the chain rule in Leibniz' notation is a better place to start. In that notation, it says

$$\frac{df}{dt} = \frac{df}{dx}\,\frac{dx}{dt}.$$

How we're going to generalize the chain rule to several variables. We'll do it in stages starting today and finishing next time. First, we're going to turn x into a vector. Then, f is a scalar-valued function $\mathbf{R}^m \xrightarrow{f} \mathbf{R}$, and $\mathbf{x}$ is actually a function $\mathbf{R} \xrightarrow{\mathbf{x}} \mathbf{R}^m$. Then, $f \circ \mathbf{x} : \mathbf{R} \to \mathbf{R}$ is defined by $(f \circ \mathbf{x})(t) = f(\mathbf{x}(t))$.

$$\begin{tikzcd} \mathbf{R} \arrow[r, "\mathbf{x}"] \arrow[dr, "f\circ \mathbf{x}"'] & \mathbf{R}^m \arrow[d, "f"] \\ & \mathbf{R} \end{tikzcd}$$

The question is, what's the derivative $\dfrac{df}{dt}$ in this generalization?

After we've answered that question, we'll make t into a vector $\mathbf{t}$ in $\mathbf{R}^n$. Then $\mathbf{x}$ won't be just a function $\mathbf{R} \xrightarrow{\mathbf{x}} \mathbf{R}^m$, but a function $\mathbf{R}^n \xrightarrow{\mathbf{x}} \mathbf{R}^m$, and $f \circ \mathbf{x} : \mathbf{R}^n \to \mathbf{R}$ is defined as $(f \circ \mathbf{x})(\mathbf{t}) = f(\mathbf{x}(\mathbf{t}))$.

$$\begin{tikzcd} \mathbf{R}^n \arrow[r, "\mathbf{x}"] \arrow[dr, "f\circ \mathbf{x}"'] & \mathbf{R}^m \arrow[d, "f"] \\ & \mathbf{R} \end{tikzcd}$$

Finally, we'll make f into a vector-valued function $\mathbf{R}^n \xrightarrow{\mathbf{f}} \mathbf{R}^p$ to get full generality. Then $\mathbf{f} \circ \mathbf{x} : \mathbf{R}^n \to \mathbf{R}^p$ is defined by $(\mathbf{f} \circ \mathbf{x})(\mathbf{t}) = \mathbf{f}(\mathbf{x}(\mathbf{t}))$.

$$\begin{tikzcd} \mathbf{R}^n \arrow[r, "\mathbf{x}"] \arrow[dr, "\mathbf{f}\circ \mathbf{x}"'] & \mathbf{R}^m \arrow[d, "\mathbf{f}"] \\ & \mathbf{R}^p \end{tikzcd}$$

The first step in the case $m = 2$. To get a better start on this, we'll let $m = 2$. Then we've got a function $\mathbf{x} : \mathbf{R} \to \mathbf{R}^2$. It's made out of two coordinate functions

$$\mathbf{x}(t) = (x, y)(t) = (x(t), y(t)),$$

where x and y are real-valued functions of t. Then $f : \mathbf{R}^2 \to \mathbf{R}$ is evaluated at $(x(t), y(t))$ to get the composition $f(x(t), y(t))$. And we want to compute the derivative $\dfrac{df}{dt}$ in terms of the derivatives of f (which are the two partial derivatives $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$) and the derivatives of x and y, namely, $\dfrac{dx}{dt}$ and $\dfrac{dy}{dt}$.

Example 1. Let $f(x, y) = x^3 + xy + y^4$, and let $x(t) = \cos t$ and $y(t) = \sin t$. Then the composition is

$$f(x(t), y(t)) = \cos^3 t + \cos t \, \sin t + \sin^4 t.$$

Now, to compute $\dfrac{df}{dt}$, just differentiate the equation $f(x, y) = x^3 + xy + y^4$, but keep in mind that x and y are both functions of t. You get

$$\frac{df}{dt} = 3x^2 \frac{dx}{dt} + \frac{dx}{dt} y + x \frac{dy}{dt} + 4y^3 \frac{dy}{dt}.$$

Now, since

$$\frac{\partial f}{\partial x} = 3x^2 + y$$

while

$$\frac{\partial f}{\partial y} = x + 4y^3,$$

we can see that $\dfrac{df}{dt}$ can be rewritten as

$$\frac{df}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt}.$$

That last equation is the chain rule in this generalization. The formal proof depends on the ordinary definition of derivative and the usual properties of limits, but as this is a form of the chain rule, the proof has a lot of details.

The first step for general m. Of course, this generalizes to

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1} \frac{dx_1}{dt} + \frac{\partial f}{\partial x_2} \frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_m} \frac{dx_m}{dt}$$

for general m, not just for $m = 2$. But let's see how that can be written more concisely using vector notation, dot products, and matrix multiplication. Note that he gradient of f is

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_m} \right),$$

while the derivative of $\mathbf{x}$ is

$$\left(\frac{dx_1}{dt}, \frac{dx_2}{dt}, \ldots, \frac{dx_m}{dt} \right)$$

which we can denote either $D\mathbf{x}$ or $\mathbf{x}'$. Then the chain rule is the dot product of these two vectors

$$\frac{df}{dt} = (\nabla f) \cdot \mathbf{x}'.$$

When treated as matrices, we can write

$$Df = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_m} \end{bmatrix}$$

and

$$D\mathbf{x} = \begin{bmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \cdots \\ \frac{dx_m}{dt} \end{bmatrix}$$

so we can also write the chain rule as a matrix multiplication of these last two matrices

$$\frac{df}{dt} = Df \; D\mathbf{x}.$$

Continued in part 2.

Math 131 Home Page at
`http://math.clarku.edu/~djoyce/ma131/`