Supplement on the *Chain Rule*

Prepared by Jonathan Goodman for Calculus I, fall 2004, Courant Institute of Mathematical Sciences, NYU.

Though scientists and engineers often regard the chain rule as simple algebra (see *Supplement on Differentials*), mathematicians regard it as a way to find the derivative of a composite function. A *function* is a rule or procedure that associates a number, called the *value*, with any number, called the *argument*, in the *domain* of the function. Usually, different arguments are associated with different values. For example, the function $f(x) = \sqrt{x}$ associates the value $\sqrt{x}$ to the argument x, provided that x is in the domain of f (real numbers that are not negative). We often think of a function as a box, with arguments entering on the left and values leaving on the right.

We write $f(x)$ for the value of the function f with argument x. It is possible that we get the argument of f in some complicated way. For example, if we want to take the square root of $2t + 3$, we could write $f(2t + 3)$. This means: start with t, multiply by 2 then add 3, then use the result as the argument of f. In this $t = 4$ would give $2t + 3 = 11$ and $f(2t + 3) = f(11) = \sqrt{11} \approx 3.32$.

The mathematicians' abstract discussion of this situation uses the idea of a *composite* function. If f and g are two functions, the composite function, $h = f \circ g$, is gotten by using the value of g as the argument of f. If the argument of g is called x, this may be written $h(x) = f(g(x))$, which again indicates that we use the value g as the argument of f. It is possible to make more complicated compositions. For example, three functions, f, g, and u, can be composed to make $f \circ g \circ u$, which means: start with the argument of u, compute the value of u, then use that result as the argument of g and compute the value of g, then use that g result as the argument of f. Be careful to remember that the expression $f \circ g \circ u$ does not mean "first f, then g, then u", but the opposite. The American Physicist Richard Feynman joked that this might be because the person who invented that notation was a Jew[1].

The following interpretation of the derivative makes the chain rule easy to derive and understand. Suppose we have a function f, fix an "nominal" or "baseline" argument, x, and consider what happens when we use nearby numbers $x + \Delta x$ as arguments to f. The change in f when we go from x to $x + \Delta x$ is $\Delta f = f(x + \Delta x) - f(x)$. When Δx is very small, we have

$$\frac{\Delta f}{\Delta x} \approx f'(x) \; . \tag{1}$$

If we multiply both sides by Δx, we get $\Delta f \approx f'(x)h$. This says: the change in the value of f is (roughly) proportional to the change in the argument of f, with f' being the constant of proportionality. You get the change in the value from the change in the argument by multiplying by the derivative.

[1] Hebrew is read from right to left, unlike most European languages. Feynman himself was a Jew and knew the real origin of the notation.

The formula (1), when dressed up in more mathematical formality, is another way to *define* the derivative. If we know a number, A, so that $\Delta f \approx A\Delta x$ for small Δx, with the approximation getting better as Δx gets smaller, then $A = f'(x)$. For example, suppose x represents the time you've been travelling and f represents the distance travelled. Saying that you get the change in the distance travelled (in miles) by multiplying the change in time (in minutes) by .6 means that the speed df/dx is about .6 miles/minute, or about 36 miles/hour.

Now suppose the argument of f is the value of g. If we change the argument of g by Δx, we change the value of g by $\Delta g \approx g'h$. The change in f is roughly f' times the change in the argument of f. Now, however, the change in the argument of f is the change in g (because g *is* the argument of f). Thus

$$\Delta f \approx f'\Delta g \approx f'g'\Delta x \; . \tag{2}$$

This is how we come to multiply the derivative of f by the derivative of g. If $h(x) = f(g(x))$, then the change in the value of h, which is the same as the change in the value of f, is roughly

$$\Delta h \approx f'g'\Delta x \; . \tag{3}$$

As we said before, this can be used to identify $f'g'$ as the derivative of h.