0.5 The Chain Rule

The chain rule is so important that I listed it as the only prerequisite for general relativity in the prelude. Here I want to explicitly point out how it should be interpreted.

THE CHAIN RULE

$$df(x_0, x_1, ..., x_n) = \sum_{i=0}^{n} \frac{\partial f}{\partial x_i} dx_i$$

The less open-minded coming from a mathematical background sometimes faint at the quantity df. In their church a derivative $\frac{df}{dx}$ is never a fraction, and the fraction-looking symbol is just a notation for a rigorously defined limit, with some epsilons and some deltas slaving behind the scenes.

We shouldn't restrict our views to such a narrow perspective.

The quantity $f(x_0, x_1, ..., x_n)$, or in short just f, is the dependent variable. Its value depends on the values of the set of the independent variables $\{x_0, x_1, ..., x_n\}$. If some of the independent variables change their values by a little, it's natural to expect that the value of f will also change by a little. For example, say $f(x) = x^2$. If x changes from 1 to 1.0001, then f changes from 1 to 1.00020001. If you don't trust me, use your calculator to make sure!

$d(blah)$ simply denotes a small change in the quantity *blah*. In the above example, $dx = 0.0001$, and $df = 0.00020001 \approx 0.0002$.

You might have noticed that $df = 2dx$, or in more familiar looks, $\frac{df}{dx} = 2$. If you calculate the derivative of $f(x) = x^2$ at our original point $x = 1$, you get $\left.\frac{df}{dx}\right|_{x=1} = 2x\Big|_{x=1} = 2$. As you can see, the derivative is simply a **ratio** between the small changes of the dependent variable and the independent variable. In this sense, the derivative really *is* a fraction, since it's just a ratio.

The derivative is the ratio between the small changes of the dependent variable and the independent variable. However, since independent variables (the xs) can be, well, independently varied, their small variations (another word for "small change") can be think of as known. In other words, the dx are usually known. **The derivative answers the question, if we change the independent variable by some small quantity $\boldsymbol{dx}$, what will the small change in the dependent variable $\boldsymbol{df}$ be?** The answer is the chain rule, $df = \frac{df}{dx}dx$. Really, it's nothing more than multiplying out the denominator of a fraction.

The usual math textbook chain rule is somewhat different: $\frac{df}{dt} = \frac{df}{dx}\frac{dx}{dt}$. This is again because of mathematicians' fear of a standalone df. In its essence this version isn't any different than our version, just divided out by another dt. The chain of logic just extends by a little bit: now f depends on x, and x itself depends on another variable t. If we change the value of t by some small amount dt, it will first change x by $dx = \frac{dx}{dt}dt$, and that change in x will cause a change in f as

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

The ratio between the change in f and the change in t is thus

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

The two versions are essentially the same. Our version $df = \frac{df}{dx}dx$ is a statement about the small changes themselves; the textbook version
$\frac{df}{dt} = \frac{df}{dx}\frac{dx}{dt}$ is a statement about the ratios.

If the dependent variable depends on many independent variables instead of just one, i.e. $f(x_0, x_1, ..., x_n)$ instead of just $f(x)$, the change in f is just adding the changes in f caused by the individual changes in $x_i, i = 0, 1, ..., n$.

The quantity $\frac{\partial f}{\partial x_i}dx_i$ is the small change in f caused only by a small change in x_i, with all the other x values held fix. This comes from the definition of partial derivatives: only one independent variable is being varied, and all the others are treated as constant. Thus,

$$df(x_0, x_1, ..., x_n) = \sum_{i=0}^{n} \frac{\partial f}{\partial x_i}dx_i$$

simply tallies up all the changes in f due to all the xs independently changing.

Explicitly, say we have two independent variables, i.e. $f(x, y)$. All this says is that if each independent variable is changed by a little bit by dx and dy respectively, then the resultant change df is the individual change brought by dx (with $dy = 0$, i.e. y held constant) plus the individual change brought by dy. This should be fairly intuitive: varying x and y together is the same as first varying x and leaving y unchanged, and then varying y with x unchanged. For example, changing (x,y) from (0,0) to (0.01, 0.01) can be viewed as first changing from (0,0) to (0.01, 0), then changing from (0.01, 0) to (0.01, 0.01).