The Chain Rule

These are some supplementary notes on the chain rule. They are an attempt to clarify a couple of confusing issues about the chain rule in a very intuitive manner. They are not mathematically rigorous and you should supplement them with one of the wide variety of excellent rigorous treatments out there (see, for example, Marsden and Tromba's "Vector Calculus").

Often it is the case that a function of interest depends on variables which in turn depend on other variables. For example, we might be interested in the function $f = f(x, y, z)$ and further it may be the case that z depends on an auxiliary parameter λ, so that $f = f(x, y, z(\lambda))$. It is useful to visualize these mathematical statements as a graphical network of dependencies. Figure 1 contains the example just mentioned. The links of the diagram indicate what depends on what

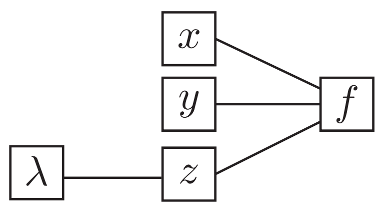

Figure 1: The network of dependencies amongst a function's arguments.

and the letters in the boxes indicate the name of that input. In particular, you can talk about shaking (or, more precisely, varying) one of the inputs in the network and see what happens. So, for example, shaking the y input clearly changes f and so it makes sense to talk about,

$$\frac{\partial f}{\partial y},$$

that is how f varies when you vary y, keeping everything else fixed.

Now, the chain rule become interesting when we shake the λ input. This causes the z input to shake, which in turn, causes f to change. This is the content of the chain rule and we write it mathematically as,

$$\frac{d}{d\lambda}\Big(f\Big) = \frac{\partial f}{\partial z}\frac{dz}{d\lambda}.$$

This is fairly straightforward. However, there are two subtleties that have arisen in section and lecture recently and I want to discuss these in particular.

Subtlety 1: The first one involves another case where, say, the function g depends on x, y and z and both y and z depend on λ so that $g = g(x, y(\lambda), z(\lambda))$. The network for g is shown in Figure 2. We can certainly shake the y and z inputs to g and we can even do this independently (that is, one at a time) and each of these variations gives rise to,

$$\frac{\partial g}{\partial y} \qquad \text{and} \qquad \frac{\partial g}{\partial z},$$

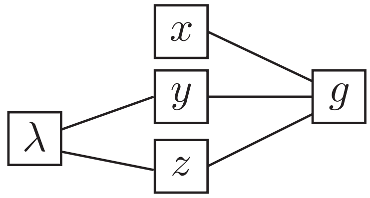

Figure 2: The network of dependencies amongst the function g's arguments.

respectively. However, shaking the λ input is a different story. When we do this both y and z necessarily vary. This doesn't mean that the way that z varies depends on the way that y varies(!), it only means that they both respond to λ being varied. This is captured by adding the two responses,

$$\frac{d}{d\lambda}\Big(g\Big) = \frac{\partial g}{\partial y}\frac{dy}{d\lambda} + \frac{\partial g}{\partial z}\frac{dz}{d\lambda}.$$

In particular, notice that if the λ dependence of y was turned off, like breaking the link between λ and y in our diagram, then the first term would vanish ($dy/d\lambda = 0$) and we would appropriately recover the analogous result to our previous example. This gives an intuitive way to think about the fact that the chain rule acts like a product within one slot and like a sum over different slots.

Subtlety 2: The second subtlety is more semantic in nature because it involves what we call things. Let's return to the first example with the function f and, now let's suppose that we're given the functional dependence of z on λ, say, $z(\lambda) = \lambda^2$. With this information in hand I might want to write,

$$f = f(x, y, z(\lambda)) = f(x, y, \lambda^2),$$

and we frequently, happily do this. But it can cause confusion. Sometimes, people are tempted by this notation to write down something like (I'm about to write something wrong),

$$\frac{d}{d\lambda}\Big(f\Big) = \frac{\partial f}{\partial \lambda^2}\frac{d\lambda^2}{d\lambda}, \qquad \text{(wrong equation!)}.$$

The trouble with this equation is that it confuses the functional dependence of the z argument of f, $z(\lambda) = \lambda^2$, with the *name*, z, of the argument of f. Instead, it is still true that,

$$\frac{d}{d\lambda}\Big(f\Big) = \frac{\partial f}{\partial z}\frac{dz}{d\lambda},$$

and the additional thing that we can say is what that second derivative is, namely,

$$\frac{d}{d\lambda}\Big(f\Big) = \frac{\partial f}{\partial z}\frac{dz}{d\lambda} = \frac{\partial f}{\partial z}\cdot(2\lambda).$$

It is with precisely this confusion in mind that mathematicians sometimes write partial derivatives with numerical subscripts, for example,

$$D_3 f(x, y, z) \equiv \frac{\partial f}{\partial z}.$$

In this notation our chain rule reads,

$$\frac{d}{d\lambda}\Big(f\Big) = (D_3 f)\frac{dz}{d\lambda}.$$

I don't plan to switch to this mathematician's notation for the simple reason that it's not very common. Much more important than notation is that you understand what is happening here and I think the example above helps to elucidate the situation.

Let me illustrate this subtlety with the two derivations of the Euler-Lagrange equations from lecture. I have received questions about each of these independently and they do a good job of showing how this topic can be confusing.

In lecture 5 (Sept. 7th) we derived the Euler-Lagrange equations using Euler's method. You can review the lecture notes for the full details. We were interested in $f(x, y(x), y'(x))$ and in particular we had found it useful to approximate the continuum derivative y' by the discrete difference,

$$y' \approx \frac{y_n - y_m}{\Delta x}.$$

This then led to the following step in the derivation,

$$\begin{aligned}
&\frac{\partial}{\partial y_n}\left(\underbrace{\cdots}_{\text{no } y_n\text{s}} + f\left(x_M, y_m, \frac{y_n - y_m}{\Delta x}\right)\Delta x + f\left(x_N, y_n, \frac{y_o - y_n}{\Delta x}\right)\Delta x + \underbrace{\cdots}_{\text{no } y_n\text{s}}\right) \\
&= \frac{\partial f}{\partial y'}\frac{\partial}{\partial y_n}\left(\frac{y_n - y_m}{\Delta x}\right)\Delta x + \frac{\partial f}{\partial y_n}\Delta x + \frac{\partial f}{\partial y'}\frac{\partial}{\partial y_n}\left(\frac{y_o - y_n}{\Delta x}\right)\Delta x.
\end{aligned}$$

The first and the third terms are a result of the second subtlety we are discussing here. The third slot of the function f is the y' slot and so, when we apply the chain rule we take the derivative with respect to this variable and multiply it by how this variable varies with y_n, i.e.

$$\frac{\partial}{\partial y_n}\left(\frac{y_n - y_m}{\Delta x}\right).$$

In our second derivation of the Euler-Lagrange equations, using Lagrange's method (see Lecture 6 (Sept. 9th), this subtlety also arose. This time it was with the following derivative,

$$\frac{\partial}{\partial \alpha}\Big(f(x, Y(x), Y'(x))\Big) = \frac{\partial}{\partial \alpha}\Big(f(x, y + \alpha\eta, y' + \alpha\eta')\Big).$$

To test your understanding of this subtlety, write out what you think this derivative is and then check it against your lecture notes.