Lecture Notes, M261-004, The Chain Rule

Sept 24, 2008

We talk today about the chain rule. The chain rule tells us how to take the derivative of a composition of functions. The chain rule becomes considerably more complicated for functions of multiple variables because we need to deal with the different dimensions. The chain rule is much easier to remember if we think in terms of the definition of derivative. We will look at the following concepts:

- Review: The Definition of Derivative
- Keeping Track of Dimensions
- Tree Diagrams
- The Chain Rule Formulas
- Implicit Differentiation

1 Review: The Definition of Derivative

In yesterday's problem session, we did a problem where we knew only the value of a function at a given point and its gradient at that same point. We used this information to approximate the value of the function at a nearby point. This gets at what the real meaning of the derivative is. It is the "slope" information for a linear function that approximates the function we are taking the derivative of. In the case of one variable to one variable, we have

$$f(x+h) \approx f(x) + hf'(x)$$

In the case of a vector valued function, we have

$$\mathbf{r}(t+h) \approx \mathbf{r}(t) + h\mathbf{r}'(t)$$

In the case of a function of two variables (for example), we have

$$f(x+h_1, y+h_2) \approx f(x,y) + \nabla f(x,y) \cdot \langle h_1, h_2 \rangle = f(x,y) + \frac{\partial f}{\partial x} h_1 + \frac{\partial f}{\partial y} h_2$$

In all of these cases we have a function of h, (h is a vector for functions of multiple variables). This function is given by a single point (the function value) plus h times the derivative. (We need to be careful how we define that multiplication in higher dimensions.)

Let's also review the chain rule for functions of one variable:

$$\frac{d}{dt}f(g(t)) = f'(g(t))g'(t)$$

And so we have

$$f(g(t+h)) \approx f(g(t)) + hf'(g(t))g'(t)$$

Note that we get two slopes here – f' and g' – and multiply them together to get the slope of our approximating line. This is always how the chain rule works, the multiplication simply takes a different form to accomodate for higher dimensions.

2 Keeping Track of Dimensions

Suppose we have a function of three variables $f(x, y, z)$ and a three dimensional vector-valued function $\mathbf{r}(t)$. Then the function obtained by composition $f(\mathbf{r}(t))$ is a function taking one variable to one variable since $\mathbf{r}(t)$ takes one variable to three variables and f takes three variables to one variable. The derivative of the composition should therefore be a single number. We know that the derivative of $\mathbf{r}(t)$ is a vector, and that the "derivative" of f – the gradient is also a vector. Following the logic before, it would seem that we need to "multiply" these two vectors together to get a scalar in order to find the derivative of our composition. This means we have to take the dot product. So

$$\frac{d}{dt}f(\mathbf{r}(t)) = \nabla f \cdot \mathbf{r}'(t)$$

and

$$f(\mathbf{r}(t+h)) \approx f(\mathbf{r}(t)) + h\nabla f \cdot \mathbf{r}'(t)$$

We can verify that this is the case with the example from yesterday:

$$f(x, y) = 10\sin(2\pi x)\cos(4\pi y), \mathbf{r}(t) = t^2\mathbf{i} + t^3\mathbf{j}$$

We have

$$f(\mathbf{r}(t)) = 10\sin(2\pi t^2)\cos(4\pi t^3)$$

so

$$\frac{d}{dt}f(\mathbf{r}(t)) = 40t\pi\cos(2\pi t^2)\cos(4\pi t^3) - 120\pi t^2\sin(2\pi t^2)\sin(4\pi t^3)$$

Now

$$\nabla f = \langle 20\pi\cos(2\pi x)\cos(4\pi y), -40\pi\sin(2\pi x)\sin(4\pi y)\rangle$$

and

$$\mathbf{r}'(t) = \left\langle 2t, 3t^2 \right\rangle$$

We can see that our result for the derivative is

$$\nabla f \cdot \mathbf{r}'(t)$$

after substituting t^2 for x and t^3 for y.

Suppose we have a function of three variables $w = f(x, y, z)$ where each of x, y and z are given as functions of the variables r and s:

$$x = x(r, s), \quad y = y(r, s), \quad z = z(r, s)$$

In this case, we can think of a function that takes the two variables r and s to the three variables x, y and z. Then w takes these two variables to one variable. So the composition takes two variables r and s to one variable. The derivative will be a two-dimensional gradient vector. We know that f has a three-dimensional gradient vector. The "derivative" of the function taking r and s to x, y and z needs to be something that we can multiply by a three-dimensional vector to get a two-dimensional vector. The correct object is a 3×2 matrix:

$$\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial s} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial s} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial s} \end{pmatrix}$$

Notice that the three rows are just the gradients of x, y and z with respect to r and s. To get the gradient of the composition, we need to multiply this matrix by the gradient of f:

$$\left\langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right\rangle \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial s} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial s} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial s} \end{pmatrix} = \left\langle \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial r}, \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial s} \right\rangle$$

3 Tree Diagrams

Another way to remember the chain rule that is recommended in the book is the use of tree diagrams, where we draw nodes for all of the independent and dependent variables and lines indicating which variables influence other variables. Each of these lines represents a derivative that needs to go into the definition of the chain rule.

4 The Chain Rule Formulas

However you choose to remember them, here are the chain rule formulas in this chapter:

Theorem 1. *If $w = f(x, y)$ has continuous partial derivative and x and y are given as functions of t, then the derivative of the composite function $w(t) = f(x(t), y(t))$ is given by*

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

Theorem 2. *If $w = f(x, y, z)$ has continuous partial derivative and x, y and z are given as functions of t, then the derivative of the composite function $w(t) = f(x(t), y(t), z(t))$ is given by*

$$\frac{dw}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}$$

Theorem 3. *If $w = f(x, y, z)$, $x = g(r, s)$, $y = h(r, s)$ and $z = k(r, s)$, then the partials of w with respect to r and s are given by*

$$\frac{\partial w}{\partial r} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial r} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial r}$$

and

$$\frac{\partial w}{\partial s} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial s}$$

5 Implicit Differentiation

The chain rule gives us a formula for implicit differentiation. Suppose y is implicitly defined as a function of x, $y = h(x)$ by the formula

$$F(x, y) = 0$$

If we consider the function $w(x) = F(x, h(x))$, then we see that $w = 0$ for all x, and so has derivative 0. Note that the derivative of the vector-valued function tracing out the path is $\langle 1, h'(x)\rangle$. So

$$\begin{aligned} w'(x) &= \nabla F \cdot \langle 1, h'(x)\rangle \\ 0 &= \frac{\partial F}{\partial x} + h'(x)\frac{\partial F}{\partial y} \end{aligned}$$

so

$$h'(x) = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}$$

Example 1. *Find* $\frac{dy}{dx}$ *if* $y^2 - x^2 - \sin(xy) = 0$.