CHAIN RULE

MATH 195, SECTION 59 (VIPUL NAIK)

Corresponding material in the book: Section 14.5.

What students should definitely get: The generic formulation of the chain rule, the particular cases of $1 \to 2 \to 1$ and $2 \to 2 \to 1$.

Executive summary

Words ...

(1) The general formulation of chain rule: consider a function with m inputs and n outputs, and another function with n inputs and p outputs. Composing these, we get a function with m inputs and p outputs. The m original inputs are termed *independent variables*, the n in-between things are termed *intermediate variables*, and the p final outputs are termed *dependent variables*.

For a given triple of independent variable t, intermediate variable x, and dependent variable u, the partial derivative of u with respect to t via x is defined as:

$$\frac{\partial u}{\partial x} \frac{\partial x}{\partial t}$$

The chain rule says that the partial derivative of u with respect to t is the sum, over all intermediate variables, over the partial derivatives via each intermediate variable.

(2) The $1 \to 2 \to 1$ and $2 \to 2 \to 1$ versions (see the lecture notes or the book).

(3) There is also a tree interpretation of this, where we make pathways based on the directions/paths of dependence. This is discussed in the book, not the lecture notes.

(4) The product rule for scalar functions can be proved using the chain rule. Other variants of the product rule can be proved using generalized formulations of the chain rule, which are beyond our current scope.

(5) Implicit differentiation can be understood in terms of the chain rule and partial derivatives.

1. The chain rule

1.1. 1 to 2 to 1 chain rule.

This simplest nontrivial chain rule is as follows: Consider two functions $x = x(t)$ and $y = y(t)$ of a single variable t, and consider a function $z = f(x, y)$ of two variables. We can compose these to get a function with one input and one output: $z = f(x(t), y(t))$. In other words, we have the composition:

$$t \overset{\langle x,y \rangle}{\mapsto} \langle x(t), y(t) \rangle \overset{f}{\mapsto} f(x(t), y(t))$$

We are composing a function from 1 variable to 2 variables and a function from 2 variables to 1 variable. Overall, we get a function from 1 variable to 1 variable. The chain rule states that:

$$\frac{d(f(x(t), y(t)))}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt}$$

We can think of this as follows: the first product is measuring the contribution to the derivative via changes in x, keeping y constant, and the second product is measuring the contribution to the derivative via changes in y, keeping x constant. (More on this interpretation a little later).

1.2. 1 to n to 1 chain rule. For the $1 \to n \to 1$ chain rule, we have n functions on one variable, and then compose them with a function of n variables to get a scalar function of one variable. The chain rule adds up the contributions of each variable.

In symbols: suppose we have function $x_1(t)$, $x_2(t)$, $\dots$, $x_n(t)$ and a function f of n variables. We can consider the function $t \mapsto f(x_1(t), x_2(t), \dots, x_n(t))$ and its derivative is as follows:

$$\frac{d}{dt}[f(x_1(t), x_2(t), \dots, x_n(t))] = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \frac{dx_i}{dt} \right)$$

1.3. 1 to n to m. If we are composing a $1 \to n$ function and a $n \to m$ function, we can reduce the chain rule to the previous chain rule, by simply looking at each coordinate of the final m-dimensional output and writing the corresponding $1 \to n \to 1$ rule.

1.4. m to n to 1, m to n to p. Suppose we have n functions, each having m inputs, and then we have a function of n inputs. Then, we can compose these and get a function with m inputs and 1 output. The chain rule for this looks the same as for $1 \to n \to 1$, except that now we have partial derivatives everywhere.

We explicitly write out the $2 \to 2 \to 1$ case. Suppose $z = f(x, y)$ is a differentiable function of x and y, where $z = g(s, t)$ and $y = h(s, t)$. Then, we have the following formulas:

$$\begin{aligned} \frac{\partial z}{\partial s} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \\ \frac{\partial z}{\partial t} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} \end{aligned}$$

We can combine all the ideas above to get the $m \to n \to p$ chain rule – the most generic version available.

The book has an interesting explanation in terms of tree diagrams. Please review this if you would like to improve upon your understanding of the above.

1.5. Terminology and conceptual formulation. For a $m \to n \to p$ chain rule, the starting m variables are termed the *independent variables*, the n variables in the middle are termed the *intermediate variables*, and the p variables at the end are termed the *dependent variables*.

Conceptually, the chain rule says that:

> The partial derivative of any dependent variable with respect to any independent variable is the sum over all intermediate variables of the product of (partial derivative of dependent variable with respect to intermediate variable) and (partial derivative of intermediate variable with respect to dependent variable).

Here's a longer version of the same explanation: given an independent variable t , an intermediate variable x , and a dependent variable u, the *derivative of u with respect to t via x* is the product:

$$\frac{\partial u}{\partial x}\frac{\partial x}{\partial t}$$

The partial derivative $\partial u/\partial x$ is to be understood as the partial derivative of u with respect to x *keeping all other intermediate variables constant* in the $n \to p$ function. The partial derivative $\partial x/\partial t$ is to be understood as the partial derivative of x with respect to t *keeping all the other independent variables constant* in the $m \to n$ function.

The derivative of u with respect to t is the *sum of all possible intermediates* of the derivative via each intermediate.

2. Deriving the product rule from the chain rule

2.1. Product rule for two scalar functions. Recall that the product rule says that:

$$\frac{d}{dt}[x(t)y(t)] = x(t)y'(t) + x'(t)y(t)$$

We now see how the product rule can be deduced using the chain rule and the fact that *constants can be pulled out of products* (in other words, the derivative of a constant times a function is the constant times the derivative of a function). Think of the function:

$$f(x, y) = xy$$

Then, the left side of the product rule is $(d/dt)[f(x(t), y(t))]$, and can thus be written as:

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

Now, $\partial(xy)/\partial x = y$ because the "constant" y can be pulled out of the derivative. Similarly, $\partial(xy)/\partial y = x$ because the "constant" x can be pulled out of the derivative. Plugging thesse in, we get the formula for the product rule.

Not only does this establish the product rule, it also gives a rough explanation for rules like the product rule which we've seen: the variant of the product rule for dot products, and the variant of the product rule for the product of a scalar and a vector function. The variant for the cross product is somewhat subtler and needs an even more generic perspective on the chain rule that is beyond our current scope.

2.2. **Product rule for more than two functions.** The general product rule says that:

$$\frac{d}{dt}[x_1(t)x_2(t)\dots x_n(t)] = x_1'(t)x_2(t)\dots x_n(t) + x_1(t)x_2'(t)\dots x_n(t) + \dots + x_1(t)x_2(t)\dots x_n'(t)$$

We can deduce this from the $1 \to n \to 1$ chain rule. The i^{th} of the summands on the right side is the partial derivative with respect to t via the i^{th} intermediate variable x_i.

3. Implicit differentiation explained

We now turn to unraveling the mystery of implicit differentiation, a topic that we learned way back in single variable calculus.

Here is how we thought of implicit differentiation. Suppose y is an implicit function of x given by a relational description of the form $F(x, y) = 0$, where it is not obvious how to isolate an expression for y in terms of x.

To find the derivative, we differentiate F with respect to x, treating y as an implicit function of x. This means that wherever we have to differentiate y, we just write dy/dx and leave it at that. After doing this differentiation, we regroup terms and compute dy/dx in terms of x and y.

We can now think of implicit differentiations as a special case of partial derivatives in the following sense. We treat x as the parameter and view x and y both as functions of x (with x being the identity function of itself). In this case, we have:

$$\frac{dF}{dx} = \frac{\partial F}{\partial x}\frac{dx}{dx} + \frac{\partial F}{\partial y}\frac{dy}{dx}$$

Here the x on the left side is the original x (viewed as the *independent variable*) and the x of partial differentiation on the right side is the *intermediate variable* x of the (x, y) pair, i.e., $\partial F/\partial x$ basically means we are differentiating with respect to the intermediate variable x treating the intermediate variable y constant in the $2 \to 1$ function, which differs from the actual differentiation with respect to x in the *composite* $1 \to 2 \to 1$ function.

Thus, if we start with $F(x, y) = 0$ and differentiate, we get $dF/dx = 0$, which gives:

$$\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{dy}{dx} = 0$$

We then rearrange and calculate:

$$\frac{dy}{dx} = -\frac{F_x}{F_y}$$