

Lecture 16: The chain rule

Nathan Pflueger

16 October 2013

1 Introduction

Today we will add one more rule to our toolbox. This rule concerns functions that are expressed as *compositions* of functions. The idea of a composition is: you can sometimes interpret one function as a sequence of two steps. The chain rule allows you to differentiate the function by differentiating the two steps individually and multiplying the results. This rule will allow us to compute a great deal more derivatives, especially when it is used in conjunction with other rules.

The reference for today is Stewart §3.5.

2 The chain rule

The basic idea that underlies the chain rule is: *the faster the inputs of a function change, the faster its outputs will change*. So for example, if $f(x)$ is one function, and $f(2x)$ is another, then the “inputs to f ” in the second function are moving twice as fast as the “inputs to f ” in the first. So its derivative is magnified by a factor of 2: $\frac{d}{dx}f(2x) = 2f'(2x)$.

The chain rule generalizes this principle. There are two standard ways to write it, which are named after the two mathematicians usually credited with inventing calculus.

The chain rule (Newton notation)	The chain rule (Leibniz notation)
$(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$	$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$

Here, the symbol \circ means “composition” (NOT multiplication). It means: feed the outputs from one function into the other. So the function $f \circ g(x)$ is just the same thing as $f(g(x))$.

In the Leibniz notation, the symbol y should refer to something which is a function of x , and the symbol z should refer to something that is a function of y (and therefore also a function of x).

At first glance, it is not at all obvious how these two statements are related. To show how they both work, I will illustrate them both to compute the derivative of $\sin(2x)$.

Newton notation	Leibniz notation
Let $f(x) = \sin x$ and $g(x) = 2x$. Then $f \circ g(x) = \sin(2x)$. So $(\sin(2x))' = f'(g(x)) \cdot g'(x)$ $= \cos(2x) \cdot 2$ $= 2 \cos(2x)$	Let $z = \sin(2x)$ and let $y = 2x$. Then $z = \sin(y)$. So $\frac{d}{dx} \sin(2x) = \frac{d \sin(y)}{dy} \frac{d(2x)}{dx}$ $= \cos(y) \cdot 2 = \cos(2x) \cdot 2$ $= 2 \cos(2x)$

The idea is the same in both cases: when you have a *composite function* (that is, a function formed by plugging the output of one function into the input of another), you can *pretend the inner function is a variable* and differentiate with respect to it. Then you must multiply the result by the rate of change of the inner function. The idea is that the term $f'(g(x))$ (in Newton notation) or the term $\frac{dz}{dy}$ (in Leibniz notation)

tells how quickly the output changes *per unit change in the input to the outer function*, and then the terms $g'(x)$ and $\frac{dy}{dx}$ tell how quickly the inputs to the outer function change *per unit change in x* .

I think you will probably find the Newton notation easier to apply initially, but I find the Leibniz notation more intuitively helpful in the long term. In fact, for the first century or so after calculus was invented, the British preferred Newton's notation while the French and Germans preferred Leibniz's notation; it turned out that Leibniz's notation was more practical in leading to further advances, and French scientific knowledge advanced somewhat faster during this time¹. Now of course, we can set patriotism aside and use the two notations interchangeably, according to which is more useful at any given time.

As an example of how to use the chain rule (in Newton notation this time), consider the following problem.

Example 2.1. Suppose that you know the following information about two functions $f(x)$ and $g(x)$. Determine $(g \circ f)'(1)$.

x	$f(x)$	$g(x)$	$f'(x)$	$g'(x)$
0	0	3	12	3
1	2	1	-6	2
2	1	7	0	7

Solution. By the chain rule, $(g \circ f)'(1) = g'(f(1)) \cdot f'(1)$. By the value in the table, $f(1) = 2$, so this is the same as $g'(2) \cdot f'(1)$. By the values in the table, this is $7 \cdot (-6) = \boxed{-42}$.

3 First examples

I will illustrate the chain rule by differentiating the following eight functions.

1. $(2x + 1)^7$
2. $\sin(5x)$
3. $\sqrt{7x + 1}$
4. $(x^2 + 1)^7$
5. $\sqrt{1 - x^2}$
6. $\sqrt{e^x + 2}$
7. $\sin(e^x)$
8. $(e^x + 1)^6$

These can be differentiated as follows. I will use Leibniz notation in this section, since I personally prefer it. Note that in homework and exams, you do not need to show as many steps as I do here – over time you will get used to skipping some of the more obvious parts (I will also begin to omit some steps in my notes as well).

$$\begin{aligned}
 \frac{d}{dx}(2x + 1)^7 &= \left[\frac{d}{d(2x + 1)}(2x + 1)^7 \right] \frac{d(2x + 1)}{dx} \\
 &= 7(2x + 1)^6 \cdot 2 \\
 &= 14(2x + 1)^6
 \end{aligned}$$

¹For a discussion, see Philip E. B. Jourdain's *The Nature of Mathematics*, chapter 5.

$$\begin{aligned}
\frac{d}{dx} \sin(5x) &= \frac{d \sin(5x)}{d(5x)} \frac{d(5x)}{dx} \\
&= \cos(5x) \cdot 5 \\
&= 5 \cos(5x)
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx} \sqrt{7x+1} &= \frac{d\sqrt{7x+1}}{d(7x+1)} \frac{d(7x+1)}{dx} \\
&= \frac{1}{2\sqrt{7x+1}} \cdot 7 \\
&= \frac{7}{2\sqrt{7x+1}}
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx} (x^2+1)^7 &= \frac{d(x^2+1)^7}{d(x^2+1)} \frac{d(x^2+1)}{dx} \\
&= 7(x^2+1)^6 (2x) \\
&= 14x(x^2+1)^6
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx} \sqrt{1-x^2} &= \frac{d\sqrt{1-x^2}}{d(1-x^2)} \frac{d(1-x^2)}{dx} \\
&= \frac{1}{2\sqrt{1-x^2}} (-2x) \\
&= -\frac{x}{\sqrt{1-x^2}}
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx} \sqrt{e^x+2} &= \frac{d\sqrt{e^x+2}}{d(e^x+2)} \frac{d(e^x+2)}{dx} \\
&= \frac{1}{2\sqrt{e^x+2}} \cdot e^x \\
&= \frac{e^x}{2\sqrt{e^x+2}}
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx} \sin(e^x) &= \frac{d(\sin(e^x))}{de^x} \frac{de^x}{dx} \\
&= \cos(e^x) \cdot e^x \\
&= e^x \cos(e^x)
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dx} (e^x+1)^6 &= \frac{d(e^x+1)^6}{d(e^x+1)} \frac{d(e^x+1)}{dx} \\
&= 6(e^x+1)^5 \cdot e^x \\
&= 6e^x (e^x+1)^5
\end{aligned}$$