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 be 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 it's 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)
$(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$	$\dfrac{dz}{dx} = \dfrac{dz}{dy} \cdot \dfrac{dy}{dx}$

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

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 tuned out that Leibniz's notation was more practical in leading to further advances, and French scientific knowledge advanced somewhat faster during this time[1]. 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).

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

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

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

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

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

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

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

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

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

4 Differentiating exponential functions

The chain rule gives us the necessary tool to differentiate arbitrary exponential functions. Remember that we chose the number e specifically to be the number such that $\frac{d}{dx}e^x = e^x$. This one fact, plus the chain rule, allows us to differentiate any exponential function.

For example, take $f(x) = 2^x$. Then using the laws of exponential functions, this can also be rewritten $f(x) = (e^{\ln 2})^x = e^{x \ln 2}$. This is a composition of two function: the chain rule says that its derivative will be $e^{x \ln 2} \cdot \ln 2 = \ln 2 e^{x \ln 2} = \ln 2 \cdot 2^x$. The same idea works for all exponentials to give the following fact.

$$\boxed{\tfrac{d}{dx} b^x = \ln(b) \cdot b^x}$$

Here, b is a constant value.

One thing that this fact reveals is that the number e is totally inescapable in calculus: even if you don't want to write your exponential functions in terms of the base e, you still must introduce the idea of a natural logarithm (and therefore the idea of the number e) to differentiate exponential functions.

5 Examples with multiple rules

In the following examples, we can differentiate the given functions with the help of the chain rule, but the chain rule must be used in conjunction with some of the other rules we have seen in class.

In this section, I will start to be a little more terse when applying the chain rule, rather than spelling all steps out in full as in the last section.

Example 5.1. Differentiate $f(x) = \tan(x \cdot 2^x)$.

Solution. This function is a composite of two functions: $\tan x$ and $x \cdot 2^x$. The first can be differentiated using the quotient rule, as we've seen: the result is $\sec^2 x$. The second can be differentiated using the product rule (and the fact mentioned in the previous section). The result is:

$$\begin{aligned}
\frac{d}{dx}\tan(x \cdot 2^x) &= \sec^2(x \cdot 2^x) \cdot \frac{d}{dx}(x \cdot 2^x) \text{ (chain rule)} \\
&= \sec^2(x \cdot 2^x) \cdot \left(\frac{dx}{dx}2^x + x\frac{d2^x}{dx}\right) \text{ (product rule)} \\
&= \sec^2(x \cdot 2^x)\left(2^x + x \cdot \ln 2 \cdot 2^x\right) \text{ (previous section)} \\
&= \sec^2(x \cdot 2^x) \cdot 2^x \cdot (1 + x \ln 2)
\end{aligned}$$

Example 5.2. Differentiate $f(x) = \sin\left(\frac{x}{x+1}\right)$.

Solution. Here we need to apply the chain rule and the quotient rule in sequence.

$$\begin{aligned}
\frac{d}{dx}\sin\left(\frac{x}{x+1}\right) &= \cos\left(\frac{x}{x+1}\right) \cdot \frac{d}{dx}\left(\frac{x}{x+1}\right) \text{ (chain rule)} \\
&= \cos\left(\frac{x}{x+1}\right) \cdot \frac{\frac{dx}{dx} \cdot (x+1) - x \cdot \frac{d}{dx}(x+1)}{(x+1)^2} \text{ (quotient rule)} \\
&= \cos\left(\frac{x}{x+1}\right) \cdot \frac{(x+1) - x}{(x+1)^2} \\
&= \frac{\cos\left(\frac{x}{x+1}\right)}{(x+1)^2}
\end{aligned}$$

Example 5.3. Differentiate $f(x) = \sqrt{\cos(x^2)}$.
Solution. This problem requires the chain rule to be applied twice in sequence.

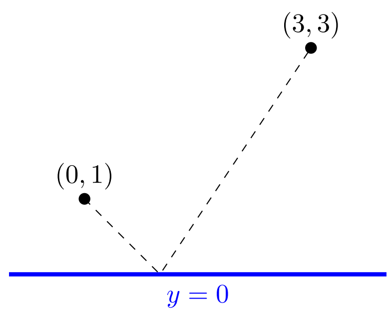
$$\begin{aligned}
\frac{d}{dx}\sqrt{\cos(x^2)} &= \frac{1}{2\sqrt{\cos(x^2)}} \cdot \frac{d}{dx}\cos(x^2) \text{ (chain rule)} \\
&= \frac{1}{2\sqrt{\cos(x^2)}} \cdot (-\sin(x^2))\frac{d}{dx}x^2 \text{ (chain rule again)} \\
&= \frac{1}{2\sqrt{\cos(x^2)}} \cdot (-\sin(x^2)) \cdot (2x) \\
&= -\frac{x\sin(x^2)}{\sqrt{\cos(x^2)}}
\end{aligned}$$

6 Optimization problems

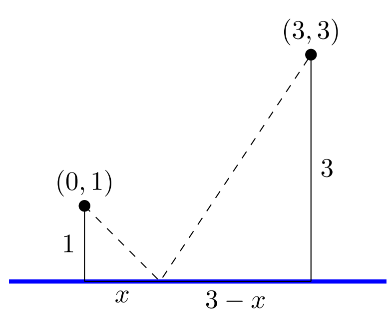
Here are a couple examples where the chain rule can be used in the solution of an optimization problem.

Example 6.1. Suppose that a horse begins at the point 1 mile north of a river, which runs east to west. She wishes to walk to the river to drink, and then to walk to the stable, which is 3 miles east and 2 miles north of where she stands. What is the shortest possible distance that she could walk to do this?

Solution. The situation is depicted visually in the following diagram.

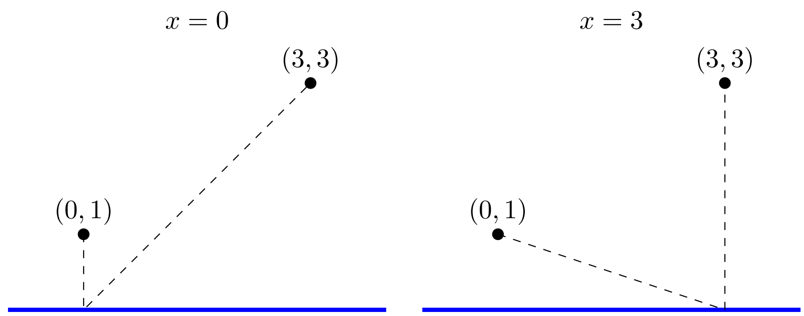


The dashed line shows two possible paths that the horse could take. She gets to choose where she will come to the river, but then should walk in a straight line to that point and then back to the stable.To find the best possible point to reach the river, introduce the variable x to be the x-coordinate of the place she arrives at the river. Then her path can be broken into the sum of the hypotenuses of two right triangles, as shown.



Total distance $= \sqrt{1 + x^2} + \sqrt{3^2 + (3 - x)^2}$

There are two extreme cases: $x = 0$ (when she goes straight to the river) and $x = 3$ (when she comes straight back from the river) shown below. Clearly no negative value of x could be better, since both legs of the trip would be longer; nor could any $x > 3$ be better than $x = 3$ for the same reason. So the best value of x must be in the interval $[0, 3]$.



So let's now optimize the function $f(x) = \sqrt{1+x^2} + \sqrt{9+(3-x)^2}$ on the interval $[0, 3]$. The first step is to differentiate it to find the critical points. We can do this with the chain rule.

$$\begin{aligned}
f'(x) &= \frac{1}{2\sqrt{1+x^2}}\frac{d}{dx}(1+x^2) + \frac{1}{2\sqrt{9+(3-x)^2}}\frac{d}{dx}(9+(3-x)^2) \\
&= \frac{2x}{2\sqrt{1+x^2}} + \frac{1}{2\sqrt{9+(3-x)^2}}2\cdot(3-x)\frac{d}{dx}(3-x) \\
&= \frac{x}{\sqrt{1+x^2}} - \frac{3-x}{\sqrt{9+(3-x)^2}}
\end{aligned}$$

To find the critical numbers, set this equal to 0 and solve.

$$\begin{aligned}
0 &= \frac{x}{\sqrt{1+x^2}} - \frac{3-x}{\sqrt{9+(3-x)^2}} \\
\frac{3-x}{\sqrt{9+(3-x)^2}} &= \frac{x}{\sqrt{1+x^2}} \\
\frac{(3-x)^2}{9+(3-x)^2} &= \frac{x^2}{1+x^2} \text{ (by squaring both sides)} \\
(3-x)^2(1+x^2) &= (9+(3-x)^2)x^2 \text{ (cross-multiplying)} \\
(3-x)^2 + (3-x)^2x^2 &= 9x^2 + (3-x)^2x^2 \text{ (distributing)} \\
(3-x)^2 &= 9x^2 \text{ (canceling like terms on both sides)} \\
(3-x) &= 3x \text{ (square root of both sides)} \\
3 &= 4x \\
3/4 &= x
\end{aligned}$$

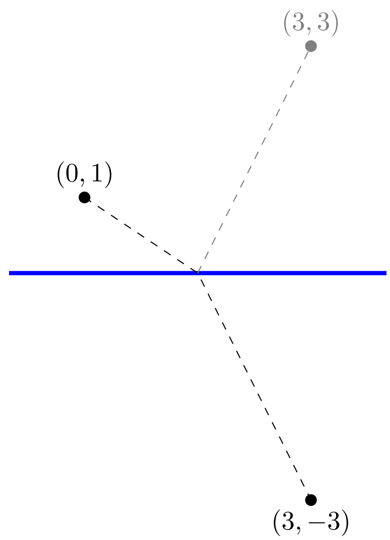
So there is one critical number: $x = 3/4$. Now check the critical number and the two endpoints to find the absolute minimum.

$$\begin{aligned}
f(0) &= 1 + 3\sqrt{2} \\
&\approx 5.24 \\
f(3/4) &= 5 \\
f(3) &= \sqrt{10} + 3 \\
&\approx 6.16
\end{aligned}$$

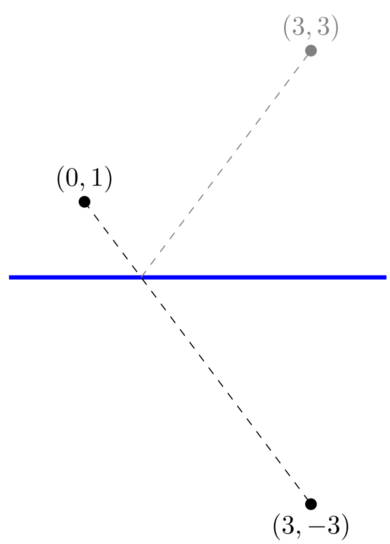
So the absolute minimum is $\boxed{5\text{ miles}}$, and occurs at $x = 3/4$.

Clever solution. This problem also has a classic, much more clever solution. Of course, a solution like this doesn't generalize well to more complicated situations, but you might enjoy seeing it.

The idea that you simply pretend that the stable is on the other side of the river. Then walking to the river and back to the stable is just like walking across the river; just reflect the path across the river.

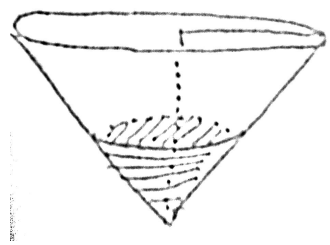


But now it is obvious what the faster path is: she should just walk in a straight line! So the shortest possible distance is the distance from $(0,1)$ to $(3,-3)$, which is $\sqrt{3^2+4^2} = 5$.



Remark. Although this problem is obviously contrived and silly, it is mathematically identical to an important problem in physics: that of a photon bouncing off a mirror. One of the basic principles of physics (discovered by Lagrange, I believe) is that light always takes the shortest possible time to go between two points. So when it must go from one point to another by way of a mirror, it takes the path that minimizes distance. In a sense, it solves the problem above (of course, the photon doesn't "solve" any equations; it just obeys mathematical laws as if it did). The same sort of problem becomes very important in studying the properties of lenses: in that setting, the relevant fact is that light travels more slowly through glass than through air. So to determine how light will bend as it passes through a lens is like asking the best path for a horse to take through a lens-shaped patch of tar, if it wants to get to the other side as quickly as possible.

Example 6.2. A cone-shaped coffee filter is 5cm tall, and has radius 5cm at the top. Suppose that it is filling up with liquid, and that height of the liquid in the filter is given by $h(t) = \frac{5}{1+t}$, where t is measured in seconds. What is the rate of change of the volume at time $t = 4$ seconds?



Solution. The liquid occupies a space shaped like a cone, with volume $\frac{1}{3}\pi r^2 h$, where r is the radius at the top and h is the height. In this case, the proportions of this cone are the same as the proportions of the filter as a whole, so the radius will be equal to the height: $r = h$. So the volume of liquid will be $\frac{1}{3}\pi h^3$. Now, h is in fact a function of t: $h(t) = \frac{5}{1+t}$, so we can write the volume also as a function of t: $V(t) = \frac{1}{3}\pi h(t)^3$. So we can find the rate of change of the volume using the chain rule.

$$
\begin{aligned}
V'(t) &= \frac{dV}{dh} \cdot \frac{dh}{dt} \\
&= \frac{1}{3}\pi \cdot 3h(t)^2 \cdot \frac{dh}{dt} \\
&= \pi \cdot h(t)^2 \cdot \frac{d}{dt}\frac{5}{1+t} \\
&= \pi \cdot h(t)^2 \cdot \left(-\frac{5}{(1+t)^2}\right) \\
&= -\frac{5\pi}{(1+t)^2} h(t)^2
\end{aligned}
$$

So in the specific case we are interested in, $V'(4) = -\frac{5\pi}{5^2} \cdot h(4)^2 = -\frac{5\pi}{5^2} \cdot \frac{5}{5} = \boxed{-\frac{\pi}{5}.}$

7 Appendix: The chain rule and linear approximation

As usual, this appendix is not part of the course material; it's included just in case of interest.

An alternative way to formulate the chain rule is: *the linear approximation of a composition is the composition of the linear approximations.* This formulation turns out to be the one that generalizes best to other situations (especially in multivariable calculus). To my mind, it is also he most intuitive way to think about it, although this may not be apparent the first time you learn the topic.

This formulation also happens to be the right strategy to use if you actually want to write down a proof of the chain rule.

To see why this is so, consider the composite function $f \circ g(x)$. Then its linear approximation around a given constant c is given as follows.

$$(f \circ g)(x) \approx f \circ g(c) + (f \circ g)'(c) \cdot (x - c) \tag{1}$$

Now, the linear approximation of g around c is:

$$g(x) \approx g(c) + g'(c)(x - c) \tag{2}$$

Now, consider the linear approximation of $f(x)$, not around the input c, but rather around the input $g(c)$ (that is the input that actually gets plugged into the function f):

$$f(x) \approx f(g(c)) + f'(g(c))(x - g(c)) \tag{3}$$

Now look what happens when you combine these last two approximations. They say that:

$$\begin{aligned} f(g(x)) &\approx f(g(c)) + f'(g(c)\,(g(x) - g(c)) \\ &\approx f(g(c)) + f'(g(c))\,(g(c) + g'(c)(x - c) - g(c)) \\ &\approx f(g(c)) + f'(g(c))g'(c)(x - c) \end{aligned}$$

The fact that this is the same as the linear approximation of $f \circ g(x)$ is just the same thing as $(f \circ g)'(x) = f'(g(c))g'(c)$.