Mathematics 102 — Fall 1999

The formal rules of calculus

So far we have calculated the derivative of each function we have looked at all over again from scratch, applying what is essentially the definition of the derivative as a limit of secant slopes. In this chapter we shall see that a very small number of general rules, together with just a few applications of the techniques we have seen so far, will allow us to calculate by algebraic manipulation the derivative of almost any function we can write down.

The three basic rules

There are just three basic rules for calculating derivatives which imply all the others, given that we know the derivatives of a few simple functions.

- **The sum rule.** *Suppose $f(x)$ and $g(x)$ are two functions whose derivatives we know. Then the sum of these two functions is defined by*

  $$(f+g)(x) = f(x) + g(x) \; .$$

  The derivative of the sum is just the sum of the two derivatives:

  $$(f+g)'(x) = f'(x) + g'(x) \; .$$

- **The product rule.** *Suppose $f(x)$ and $g(x)$ are two functions whose derivatives we know. Then the product of these two functions is defined by*

  $$(f \cdot g)(x) = f(x) \cdot g(x) \; .$$

  The derivative of the product is given by this formula:

  $$(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x) \; .$$

- **The composition rule.** *Suppose $f(x)$ and $g(x)$ are two functions whose derivatives we know. Then the composition of these two functions is defined by*

  $$(f \circ g)(x) = f(g(x)) \; .$$

  The derivative of the composition is given by this formula:

  $$(f \circ g)'(x) = f'(g(x))g'(x) \; .$$

The third rule is by far the most complicated, since it is probably not even clear at first what it means. We shall discuss this in detail in a moment. It is also called the **chain rule** since the composition of two functions is obtained in a sense by linking the two functions in a chain of calculations. It is important to keep in mind that the order in which we chain them together is important.

We can now proceed in one of two ways: (1) We can try to understand why these rules are true, or (2) we can see how they are applied in practice and how they can be extended to make up other rules, without worrying about where they come from. We shall follow the second path, although why they are true is very closely related to the basic meaning of the derivative. We shall take up this question somewhere else. In any event, we shall see that *these rules are very, very powerful.*

Before we go on, however, we shall make a remark about their history. Much of what we have said so far, (in earlier chapters) was known, although perhaps in slightly different terminology, to Isaac Newton and a few of his predecessors. Newton himself undoubtedly knew the rules themselves, and used them constantly. But his notation was exceedingly clumsy. Often, as many of our previous arguments did, his methods relied on explicit geometric reasoning. And it was not easy at all to apply them in practical problems. It was Leibniz, who all his life pursued the idea of 'automatic thinking,' who was largely responsible for our modern emphasis on the importance of simple formal rules and convenient notation in calculus. It was he who pointed out, among other things, the importance of of functions, which he might be said to have invented. That is to say, he was the first to think of all functions in a similar way–as a general class of things—and we think it is fair to say that he used our definition of a function as a rule. He also gave us the name ‘calculus’ for the techniques of calculation he formulated so clearly, in conscious analogy with the more primitive rules of arithmetic to which the term had previously been applied. (‘Calculus’ was originally a Latin word for the small pebbles used in an abacus.)

Now, the contributions of Leibniz over and above those of Newton were almost purely formal. No one doubts that in the hands of Newton Leibniz’ formalism would have been of no importance. But for mathematicians with less talent, it was a tremendous relief to be able to use calculus without demonstrating every time you used it that you understood it more or less at all levels.

We know almost exactly what the importance of Leibniz’ contribution was, because even within a few years of Leibniz’ work mathematicians on the mainland of Europe had adopted his style of calculus without pain, and during the next century used it to solve an almost unbelievable range of problems in mathematics and in science more generally. On the other hand, in England for all of that time Leibniz’ ideas were shunned as un-English, and using them was considered an insult to the memory of Newton. During the Eighteenth century, while the Continent was building astronomy and physics, the English were trying to understand no more than what Newton had written. It is a sad story. Unless you happen to be from one of those other countries, of course. Or perhaps Scottish.

A few simple consequences

In this section we shall look at a few of the simplest examples of how to apply the first two rules. In the next we shall look at the chain rule. But even in this one we shall see how the basic rules can be extended to include a number of other important formulas for derivatives.

The basic rules tell us only how to find some new derivatives if we already know some others. In order to use them, we have to have something to start with. As this course proceeds, we shall accumulate and occasionally modify a relatively small list or dictionary of some very simple functions whose derivatives must be calculated from basic principles. Our initial list will include just two easy ones:

- *If $f(x)$ is a constant, then $f'(x) = 0$.*
- *If $f(x) = x$ then $f'(x) = 1$.*

Both of these ought to be immediately apparent, since a constant function has a horizontal graph with slope 0, and the graph $y = x$ is a straight line with constant slope 1. Later on we shall expand our dictionary of ‘simple’ functions whose derivatives we have to calculate from basic principles—to include principally exponential functions and trigonometric functions—but even the two above will get us a long way. So what we are going to do now is start with the three basic rules and these two examples and see what we can deduce.

Example 1. Suppose $f(x) = x + 1$. Then the sum rule tells us that $f'(x) = 1$.

Example 2. Suppose $f(x) = cx$ where c is a constant. Then $f'(x) = c'x + cx' = c$.

Example 3. Suppose $f(x) = x^2$. We can write it as the product of x and x. Therefore we can apply the product rule to get

$$f'(x) = (x \cdot x)' = x\,x' + x'\,x = x + x = 2x\ .$$

Example 4. If $f(x) = x^2 + x$ then $f'(x) = 2x + 1$.

Example 5. Suppose $f(x) = x^3$. We can write it as the product of x and x^2, so we can calculate

$$f'(x) = (x \cdot x^2)' = x'\,x^2 + x\,(x^2)' = x^2 + x \cdot 2x = 3x^2\ .$$

Example 6. This sort of thing can be continued. It looks very much as though

- *The derivative of x^n is always nx^{n-1}, where n is a positive integer.*

If we know it to be true for n, then we can look at x^{n+1}, the product of x and x^n. We have

$$f'(x) = (x \cdot x^n)' = x\,x^n + x\,(x^n)' = x^n + nx^n = (n+1)x^n \ .$$

We know the rule works for $n = 0$ and $n = 1$, and then this reasoning allows to conclude in turn for $n = 2, 3$, etc. and eventually any positive integer n.

Example 7. Suppose $f(x) = cg(x)$. Then $f'(x) = c'g(x) = cg'(x) = cg'(x)$.

Example 8. The sum rule repeated many times will tell us the rule for finding the derivative of a polynomial

$$c_0 + c_1 x + c_2 x^2 + \cdots$$

which we have sen before.

Example 9. Suppose $f(x) = \sqrt{x}$. There is no direct way to see what $f'(x)$ is, but we can proceed indirectly. We know that $f(x)^2 = (\sqrt{x})^2 = x$. This might be called the **defining equation** of the square root. Therefore we can write

$$(f(x)^2)' = f'(x)f(x) + f(x)f'(x) = x' = 1$$

and hence

$$2f'(x)f(x) = 1, \quad f'(x) = \frac{1}{2f(x)} = \frac{1}{2\sqrt{x}} \ .$$

Example 10. Now let $f(x) = 1/x$. Again, we can do an indirect calculation. The defining equation of the function $f(x) = 1/x$ is

$$x \cdot f(x) = x/x = 1 \ .$$

From this we get

$$\begin{aligned} x'\,f(x) + xf'(x) &= 0 \\ xf'(x) &= -f(x) \\ f'(x) &= -\frac{f(x)}{x} \\ &= -\frac{1}{x^2} \ . \end{aligned}$$

Example 11. Suppose more generally that $f(x) = 1/g(x)$. Then we start with a defining equation and continue:

$$\begin{aligned} f(x)g(x) &= 1 \\ f'(x)g(x) + f(x)g'(x) &= 0 \\ f'(x) &= -\frac{f(x)g'(x)}{g(x)} \\ &= -\frac{g'(x)}{g(x)^2} \ . \end{aligned}$$

Example 12. Suppose

$$f(x) = \frac{g(x)}{h(x)} \ .$$

Then

$$\begin{aligned} f'(x) &= g'(x)\left(\frac{1}{h(x)}\right) + g(x)\left(\frac{1}{h(x)}\right)' \\ &= \frac{g'(x)}{h(x)} - \frac{g(x)h'(x)}{h(x)^2} \\ &= \frac{g'(x)h(x) - g(x)h'(x)}{h(x)^2} \ . \end{aligned}$$

This is called the **quotient rule**. It should be considered as a fourth basic rule, to be memorized along with the original three basic ones.

Exercise 1. *Find the derivative of* $(x+1)/(x-1)$.

Exercise 2. *Find the derivative of* $1/(x^2+x+1)$.

Exercise 3. *Find the derivative of* $(x^2-1)/(x^2+1)$.

Exercise 4. *Find the derivative of* $1/\sqrt{x}$.

The chain rule

We have defined the composition of two functions by the formula

$$(f \circ g)(x) = f(g(x)) \ .$$

Let's see what this means by looking at an example.

Example 13. Let $h(x) = \sqrt{x+1}$. What do we have to do to evaluate this function, say at $x = 1$? Well, first we calculate $x + 1 = 2$, and then we calculate $\sqrt{1+1} = \sqrt{2}$. In other words, we calculate $h(1)$ in a two-step process. This is always a situation in which we are looking at a composition of functions. One very good way to think of a function is as a black box—or maybe even a sausage grinder! Numbers go in one side, and generally different numbers come out the other. The composition of two functions is what you get when you feed in to a second box the stuff that comes out of the first one. Here the first box takes in x and spits out $x + 1$. The second then takes this output and feeds it into the $\sqrt{\ }$ box, which then produces $\sqrt{x+1}$. In our formulation of the composition rule g is the first box and f is the second, because **first** we calculate $g(x)$ and **then** evaluate f at that. So:

$$\begin{aligned} g(\bullet) &= \bullet + 1 \\ f(\bullet) &= \sqrt{\bullet} \end{aligned}$$

because then we have

$$(f \circ g)(x) = f(g(x)) = f(x+1) = \sqrt{x+1} \ .$$

We don't use x here because we want to emphasize (once more) that *we are really thinking of the function as a rule for producing other numbers, not an expression.*

Example 14. Let $h(x) = g(x)^n$. This will be our second look at what composition means, since we evaluate $h(x)$ first calculating $g(x)$, and then evaluating the n-power of that. In other words, $h(x)$ is the composition of the two functions (a) taking $\bullet$ to $g(\bullet)$ and (b) $\bullet$ to $\bullet^n$. Again, we use $\bullet$ as a variable because it might be more confusing to use x twice in different roles. So if $f(x) = x^n$ then $h = f \circ g$. By the composition rule

$$h'(x) = f'(g(x))g'(x) = ng(x)^{n-1}g'(x)$$

since the derivative of x^n is nx^{n-1}.

Exercise 5. *Find the derivative of* $(x^2+x+1)^{20}$.

Exercise 6. *Suppose* $f(x) = g(x^n)$*, which we get by chaining these same two functions in the opposite order. Find a formula for* $f'(x)$.

Exercise 7. *What is the derivative of* $(x^2 + x + 1)/\sqrt{x^2 + 1}$*?*

Exercise 8. *Calculate the derivative of* $x^{1/n}$ *where* n *is a positive integer. Explain how the rules apply.*

Exercise 9. *Calculate the derivative of* $x^{m/n}$ *where* m *and* n *are positive integers. Explain how the rules apply.*

Exercise 10. *Calculate the derivative of* $x^{-m/n}$ *where* m *and* n *are positive integers. Explain how the rules apply.*

Exercise 11. *Find one single formula for* x^r *where* r *is any fraction. Explain how the rules apply.*

A new class of functions to go in the dictionary

We have computed at various points entries in this tabel:

$f(x)$	$f'(x)$
1	0
x	1
$\sqrt{x}$	$1/(2\sqrt{x})$
$1/x$	$-1/x^2$
$\sqrt[3]{x}$	$1/(3\sqrt[3]{x^2})$

These are all examples of a single general rule

x^r	rx^{r-1}

as you can check for yourself (with $r = -1, 1/2, 1/3$). This formula can be proven from things we know when r is a fraction $\pm m/n$, but deducing it for arbitrary r is not so simple. We shall do this later on when we are introduced to exponential functions. For now, however, we shall add it to our basic dictionary of relatively simple functions whose derivatives we have to memorize. In fact we can restrict our list to just two types: (1) $f(x)$ constant; (2) $f(x) = x^r$.

More about the formal rules

Let's summarize—all of the functions we normally deal with in this course can be obtained by a few simple methods from a small list of elementary ones. The elementary ones we know so far are (1) constants and (2) powers of x of the form x^r. This last includes negative powers such as $x^{-1} = 1/x$ as well as fractional ones like $\sqrt{x} = x^{1/2}$. Eventually we shall add to this list a few more functions like exponentials and logarithms, as well as trigonometric ones.

The basic ways to get new functions from old ones are to (1) add them, (2) multiply them, (3) combine them (or chain them together), and (4) divide them. In order to find the derivative of a function, you must figure out how it is obtained from elementary ones by these processes. It is important to realize right at the beginning that although we can tell you a sure-fire way to go about finding derivatives, it is not a simple process. You should not think yourself unintelligent if it does not look obvious to you. It is certainly better adapted for a computer than for most of you. But you should be able to handle cases of medium complexity without fear.

Let's do an example. Let

$$f(x) = \sqrt{x^2 + 1}\,.$$

Now as we try to calculate the derivative of this, we shall be dealing with simpler and simpler functions. Let's keep a list of jobs to do as we proceed. We shall number the functions as they are listed, too. We start by putting $\sqrt{x^2 + 1}$ on the list. We shall use $\bullet$ ('blob') for the variable to remind ourselves that we *must* think of functions as rules, not just expressions.

So the list starts with

1. $f_1(\bullet) = \sqrt{\bullet^2 + 1}$

Being on this list means we want to find its derivative.

This function is the composite of two others, namely

2. $f_2(\bullet) = \sqrt{\bullet}$

3. $f_3(\bullet) = \bullet^2 + 1$

That is to say, in order to evaluate $f_1(x)$ we **first** evaluate $f_3(x) = x^2 + 1$ and **then** evaluate $\sqrt{f_3(x)}$. We can keep track of things better if we modify the first item to read

1.

$$\begin{aligned} f_1(\bullet) &= \sqrt{\bullet^2 + 1} \\ &= (f_2 \circ f_3)(\bullet) \end{aligned}$$

The chain rule tells us that we can calculate $f_1'(\bullet)$ if we can calculate f_2' and f_3', so we add to the entry for f_1:

1.

$$\begin{aligned} f_1(\bullet) &= \sqrt{\bullet^2 + 1} \\ &= (f_2 \circ f_3)(\bullet) \\ f_1'(\bullet) &= f_2'(f_3(\bullet))f_3'(\bullet) \end{aligned}$$

The function in **2.** is the same as $\bullet^{1/2}$, and we know its derivative to be $(1/2)\bullet^{-1/2}$. We now make a new kind of entry to our list, whenever we find a derivative for one on the list:

2.

$$\begin{aligned} f_2(\bullet) &= \sqrt{\bullet} \\ f_2'(\bullet) &= (1/2)\,\bullet^{-1/2} \end{aligned}$$

As for **3.** we can apply the sum rule and power rule to get

3.

$$\begin{aligned} f_3(\bullet) &= \bullet^2 + 1 \\ f_3'(\bullet) &= 2\,\bullet \end{aligned}$$

Now we can go back to **1.** We get

$$\begin{aligned} f_1'(\bullet) &= f_2'(f_3(\bullet))f_3'(\bullet) \\ &= \left[(1/2)(f_3(\bullet))^{-1/2}\right]\left[\,2\ \bullet\,\right] \\ &= \left[(1/2)(\bullet^2 + 1)^{-1/2}\right]\left[\,2\ \bullet\,\right] \\ &= \frac{\bullet}{\sqrt{\bullet^2 + 1}} \end{aligned}$$

That's it! Of course this is a somewhat long-winded process. You will get faster at this sort of thing with practice. But as I have often suggested in this course, if forced to make a choice between speed and security, your best bet nearly always is to go for security.

Let's do another example in the same style. Let

$$f(x) = \sqrt{\frac{x}{x+1} + x^2} + \frac{1}{x^2}\ .$$

1.

$$\begin{aligned} f_1(\bullet) &= \sqrt{\frac{\bullet}{\bullet + 1} + \bullet^2} + \frac{1}{\bullet^2} \\ &= f_2(\bullet) + f_3(\bullet) \\ f_1'(\bullet) &= f_2'(\bullet) + f_3'(\bullet) \end{aligned}$$

2.

$$f_2(\bullet) = \sqrt{\frac{\bullet}{\bullet+1} + \bullet^2}$$
$$= (f_4 \circ f_5)(\bullet)$$
$$f_2'(\bullet) = f_4'(f_5(\bullet))f_5'(\bullet)$$

3.

$$f_3(\bullet) = \frac{1}{\bullet^2}$$
$$= \bullet^{-2}$$
$$f_3'(\bullet) = -2\bullet^{-3}$$

4.

$$f_4(\bullet) = \sqrt{\bullet}$$
$$f_4'(\bullet) = (1/2)\bullet^{-1/2}$$

5.

$$f_5(\bullet) = \frac{\bullet}{\bullet+1} + \bullet^2$$
$$= f_6(\bullet) + f_7(\bullet)$$
$$f_5'(\bullet) = f_6'(\bullet) + f_7'(\bullet)$$

6.

$$f_6(\bullet) = \frac{\bullet}{\bullet+1}$$
$$f_6'(\bullet) = \frac{1\cdot(\bullet+1) - 1\cdot\bullet}{(\bullet+1)^2}$$
$$= \frac{1}{(\bullet+1)^2}$$

7.

$$f_7(\bullet) = \bullet^2$$
$$f_7'(\bullet) = 2\bullet$$

And now we go back up the path. In **5.** we get

$$f_5'(\bullet) = \frac{1}{(\bullet+1)^2} + 2\bullet$$

and in **2.** we get

$$f_2'(\bullet) = f_4'(f_5(\bullet))f_5'(\bullet)$$
$$= (1/2)\ (f_5(\bullet))^{-1/2}\left(\frac{1}{(\bullet+1)^2} + 2\bullet\right)$$
$$= (1/2)\ \left(\frac{\bullet}{\bullet+1} + \bullet^2\right)^{-1/2}\left(\frac{1}{(\bullet+1)^2} + 2\bullet\right)$$

Finally in **1.**

$$f_1'(\bullet) = f_2'(\bullet) + f_3'(\bullet)$$
$$= (1/2)\ \left(\frac{\bullet}{\bullet+1} + \bullet^2\right)^{-1/2}\left(\frac{1}{(\bullet+1)^2} + 2\bullet\right) - 2\bullet^{-3}$$

Whew! This is complicated. *But it is the overall process which is complicated. Each single step is very simple.* And keep in mind that *most of the time just one step will be needed.*

Final remarks on this process

We can visualize this process in a diagram that some might find helpful. We shall look at the last example again. We start with

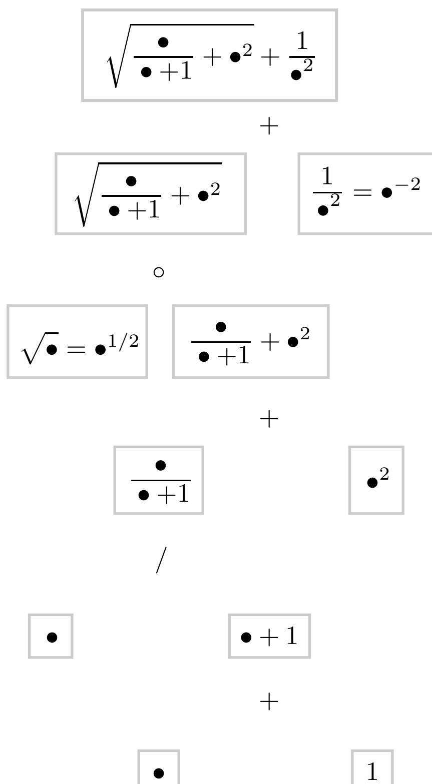
$$f_1(\bullet) = \sqrt{\frac{\bullet}{\bullet + 1} + \bullet^2} + \frac{1}{\bullet^2}\ .$$

We write it as the sum of two pieces $f_2 + f_3$ where

$$f_2(\bullet) = \sqrt{\frac{\bullet}{\bullet + 1} + \bullet^2}$$

$$f_3(\bullet) = \frac{1}{\bullet^2}$$

We then break down $f_2(\bullet)$ in turn, but don't have to break down $f_3(\bullet)$ because it is a power function. Etc. We can visualize what is going on by a diagram like this, which we shall explain in a moment.



The rule for this diagram is very simple. Each function is obtained from the two below it to left and right, according to the symbol immediately below it and in between them. If a function in the diagram is a power rule or a constant, it does not have to be broken down. Furthermore, each combination matches one of the basic rules, so moving up the tree we can always calculate derivatives at one level from derivatives at the level below it.

You will not be too surprised to know that this diagram matches exactly how a computer could be used to calculate derivatives.