Computing Derivatives

Having established the derivative of some basic functions from first principles, we seek a machinery to compute derivatives of a larger class of functions based on this previous knowledge. The goal of the lectures outlined below is to indicate a possibly enlightening approach to the main methods used to do this in the course.

The Basic Rules. *Given two functions f and g, how can we combine them to produce another function?* Students should quickly recall the most basic operations such as adding, multiplying and dividing one function by another.

What happens when we add two continuous functions together? Is the resulting function also continuous? What happens when we add two differentiable functions together? Is the resulting function also differentiable? As is often the case in mathematics, the best thing to do is to consider a very simple case first, say $f(x) = g(x) = x$. Adding these functions together yields a function $h(x) = f(x) + g(x)$ and, computing its derivative from first principles, we observe that $h(x)$ *is* differentiable with derivative $h'(x) = f'(x)+g'(x)$. *Was this an accident? What was special about f and g in our computation of the limit?* Only the fact that they were differentiable, we've discovered a rule! (It may be psychologically instructive to redo the computation of the derivative with arbitrary functions and to highlight that the additive rule for derivatives follows from the corresponding limit rule.)

What happens when we multiply two differentiable functions together? Is the resulting function also differentiable? Many students won't have previously encountered the product rule so they may expect that it follows the same pattern as the additive rule we just discovered. Once again, it is most enlightening to consider a simple case first, say $f(x) = c$ and $g(x) = x$. Multiplying them together yields the function $h(x) = f(x) \cdot g(x) = c \cdot x$ and, computing the derivative from first principles, we observe that h is differentiable with $h'(x) = f(x) \cdot g'(x) = c \cdot 1$. *Why is this result peculiar?* This should seem a bit weird since our formula is asymmetric even though $f(x)g(x) = g(x)f(x)$. *What was special about f and g in this computation?* Students should eventually realize that we crucially used the fact that $f(x) = c$ was a constant function but that there wasn't anything special about the function $g(x)$: we've discovered the "constant multiple rule". (It may be, once again, psychologically instructive to redo the computation of the derivative keeping $f(x) = c$ fixed and letting $g(x)$ be arbitrary.)

Is this rule satisfactory? What would you explore next in your quest for a "product rule"? Evaluating the derivative of a product of arbitrary functions from first principles seems a bit daunting but there are some products of functions that we can handle. For instance, we have probably already observed that the derivative of $f(x) = x^2$ was $f'(x) = 2x$ and we could similarly work out from the definition that $g(x) = x^3$ has derivative $g'(x) = 3x^2$ while $h(x) = x^4$ has derivative $h'(x) = 4x^3$.

Given the function $p(x) = x^n$, would you expect it to be differentiable and, if so, what would you expect as its derivative? Students should be able to guess the appropriate formula and be willing to accept it on faith (disbelievers can always be referred to the binomial formula). *How is this telling us something about the derivatives of products of functions?* Students should be encouraged to realize that a function like $f(x) = x^5$ can be interpreted as a product in many ways such as $f(x) = x \cdot x^4 = x^2 \cdot x^3$. Writing our function as a product $f(x) = p(x) \cdot q(x)$, we would expect from our additive rule and our constant multiple rule that a formula for the derivative $f'(x)$ might involve sums and products of $p(x)$, $p'(x)$, $q(x)$ and $q'(x)$. Working out a few cases and laying out the data suggestively on the blackboard should lead to an accurate guess.

For instance, if we think of $f(x) = x^5$ as being the product $x^2 \cdot x^3$, then our data consists of

$$p(x) = x^2, \quad q(x) = x^3$$
$$p'(x) = 2x, \text{ and } \quad q'(x) = 3x^2.$$

How can you combine these pieces to yield $f'(x) = 5x^4$? A more dramatic example can also help: if we think of $f(x) = x^{24}$ as being the product $x^5 \cdot x^{19}$, then our data consists of

$$p(x) = x^5, \quad q(x) = x^{19}$$
$$p'(x) = 5x^4, \text{ and } \quad q'(x) = 19x^{18}.$$

How can you combine these pieces to yield $f'(x) = 24x^{23}$? Determining what the product rule looks like in this way makes it feel much less "random" from the student's point of view. Moreover, they will appreciate that it is a lot less daunting to try and evaluate the derivative of a product of arbitrary functions using the limit definition once they know what they are looking for (and they should be encouraged to try it out on their own to "prove" their formula is always correct).

Having gone through all this trouble, the students should be ready to accept that similar considerations would have lead them to discover the quotient rule as well. To conclude, it might be interesting to point out that, armed with these differentiation rules and knowing only the derivative of constant and linear functions, they can now handle the derivatives of all quotients of polynomial functions!

An Interesting Function. At this point in time, the functions whose derivatives we can compute from the definition and the rules we discovered generate our universe of differentiable functions. So far, this consists essentially of functions we can build by adding, subtracting, multiplying and dividing polynomials. *How could we go about expanding this universe?* We have seen that differentiation is well behaved with respect to addition of functions and, in particular, that the sum of two polynomials is once again differentiable. Proceeding naively, one might therefore think that an infinite sum (whatever that means) of polynomials could be differentiable as well.

For instance, recall that we introduced e as the magical number for which $e^x = 1 + x + \frac{x^2}{2} + \ldots + \frac{x^k}{k!} + \ldots$ and we convinced ourselves that this was plausible (we still don't know how to make this precise but we'll hold that thought for now) by plotting its successive polynomial approximations:

(1) $p_1(x) = 1 + x$,
(2) $p_2(x) = 1 + x + \frac{1}{2}x^2$,
(3) $p_3(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3$,
(4) $p_4(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{12}x^4$, and
(5) $p_5(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{12}x^4 + \frac{1}{60}x^5$.

Do you think $f(x) = e^x$ should be a differentiable function? Observing that $p_i'(x) = p_{i-1}(x)$, what would you expect $f'(x)$ to be?

Holding those thoughts for now, we will attempt to answer the following more general question: *If $f(x) = b^x$ for $0 < b \neq 1$, can we use our rules to determine whether $f'(x)$ exists and, if so, compute it?* When there's nothing better to go on, we start from the definition and see how far we can get:

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h\to 0} \frac{b^{x+h} - b^x}{h} = \lim_{h\to 0} \frac{b^x b^h - b^x}{h} = \lim_{h\to 0} \frac{b^x(b^h - 1)}{h} = b^x \cdot \left(\lim_{h\to 0} \frac{b^h - 1}{h} \right). \tag{1}$$

What happened in the last step? What can you say about the bracketed quantity? Can you spot the derivative? Rewriting the bracketed quantity, we see that it is nothing but

$$\lim_{h\to 0} \frac{b^h - 1}{h} = \lim_{h\to 0} \frac{b^{0+h} - b^0}{h} \tag{2}$$

which would coincides with $f'(0)$. Here, it can be left as an interesting exercise for the students to convince themselves that for $0 < b \neq 1$ this limit always exists and therefore $f'(0)$ has a well defined meaning. Putting this all together, we finally see that if $f(x) = b^x$, then $f'(x) = b^x \cdot f'(0)$. *What would happen if $f'(0)$ was somehow equal to 1? Can we find a base b for which this is the case?* It turns out there is a unique number for which this happens and we call it "e". Students tend to find such definitions a bit weird and should be encouraged to consider the analogy with π that is the unique number which, when multiplied by the diameter of a given circle, yields its circumference. *How would you argue that the number e exists?*

The Chain Rule. Understanding the subtleties behind the chain rule will be tricky for most students. In order to bring these subtleties to their attention, it can be good to start with a somewhat controversial question. Consider, for instance, the following expression:

$$\lim_{h\to 0} \frac{\sin(2x+h)-\sin(2x)}{h}. \tag{3}$$

Does the limit exist? If so, can you evaluate it? Most students will realize that this is a question about derivatives but they will usually be unsure what derivative they are considering. Many will be tempted to guess that the answer is $\cos(x)$. With a bit of help, the class should eventually come to realize that what they are after is the derivative of $\sin(2x)$ but most students will then stubbornly argue that the answer is in fact $\cos(2x)$. This is a great learning opportunity!

In order to illustrate what's going on, it is worthwhile to take a step back and break $\sin(2x)$ down as the composition of two functions $f \circ g\,(x)$ where $g(x) = 2x$ and $f(y) = \sin(y)$. At this point, it should be explicitly pointed out (Keeping in mind that most students are still shaky on function composition!) that, while we do know how to compute the derivatives of f and g, we do not yet have a mechanism at our disposition to use the resulting data to compute the derivative of $f \circ g$. This is where the chain rule comes to the rescue.

It can be tempting to introduce the chain rule in Leibniz notation but this is a disservice to the students since it obfuscates the underlying composition. One might state it as follows instead:

If g is differentiable at a and f is differentiable at $g(a)$, then $f \circ g$ is differentiable at a and its derivative is given by

$$(f \circ g)'(a) = f'(g(a)) \cdot g'(a). \tag{4}$$

What does the chain rule tell us about the derivative of $\sin(2x)$*? Why do you think this is called the chain rule?* Any answer is good here but perhaps it is because it allows us to handle the derivative of a "chain" of functions

$$x \xrightarrow{g} g(x) \xrightarrow{f} f \circ g\,(x). \tag{5}$$

What are the similarities and differences between the chain rule and the previous rules we have encountered? As before, it allows us to break up the derivative of a "complicated function" $f \circ g$ into data obtained from its "simpler" components f and g. The most prominent difference between this rule and the previous ones (besides the fact that the formula is different) is perhaps that we no longer evaluate the "pieces" of the formula at the same points.

Once students have gathered that to use the chain rule they only need to compute two derivatives and evaluate them at the appropriate points, the most tricky part for them is to identify what they should consider as their "outer" function f and their "inner" function g. The only way around this is through many class-driven examples, getting students to collectively choose what f and g ought to be. Using colour chalk here is crucial to illustrate what's going on. Having shown them all the crazy derivatives they can now compute, one might conclude with the following questions: *Which differentiation tool do you find to be the most useful, powerful, versatile? How much have we enlarged our universe of known differentiable functions? Can we find the derivative of* $\ln(y)$*?*

Implicit Differentiation. Students are typically well accustomed to thinking about the graph of a function as a subset of $\mathbb{R}^2$. What might not have occurred to them though is that *any* equation in two variables defines, or "carves out", a (possibly empty) subset of the plane corresponding to its set of solutions. This is best explored through a sequence of pictorial examples (great opportunity for computer graphics) such as $x = 1$, $y = x$, $x^2+y^2 = 1$, $y = x^2$, $y^2 = x$, $(y-1)(x-2) = 0$, $(x^2+y^2)^3 = 4x^2y^2$, $x^2 + 1 = 0$, etc. Another fun family of examples comes from elliptic curves where we can plot the zero locus of $y^2 = x^3 + ax + b$ for various values of a and b. It should be highlighted that these solution sets can make crazy patterns in the plane that *may or may not* represent a function. Nevertheless, they can still usually be described piece-by-piece using many hidden *implicitly* defined functions.

The first key thing to observe is that, when we consider an equation such as $y^2 = x$, there is no a-priori reason to think about x as an independent variable and y as a dependent variable (or vice-versa). It is by making such choices that we discover the various implicitly defined functions lurking behind an equation. For instance, if we think of x as the independent variable and $y = y(x)$ as depending on x, then we are considering the equation

$$y(x)^2 = x. \tag{6}$$

Is this a function? Trying to write out $y(x)$ explicitly as a function of x, we are forced to take a square root and consider two different cases:

$$y_1(x) = \sqrt{x} \quad \text{and} \quad y_2(x) = -\sqrt{x}. \tag{7}$$

The subset of the plane carved out by $y^2 = x$ can be described as the union of the graphs of y_1 and y_2 (this is best illustrated by using a different colour for each graph), two hidden implicitly defined functions. On the other hand, had we chosen to consider y as the independent variable and $x = x(y)$ as depending on y, then we could have described the subset carved out by $y^2 = x$ as the graph of the single function $x(y) = y^2$; our chosen point of view will always depend on the context. *Where have we seen this kind of shifting point of view before?* Students should come to realize that they often perform such mental acrobatics when they choose to consider demand as a function of price or price as function of demand. Indeed, in economics, "everything is always a functions of everything".

Having realized that there are many functions lurking behind an equation, we might start to wonder about their derivatives (especially in the context of a supply and demand equation, using a concrete economics example all along wouldn't be a bad idea). For instance, if we consider the equation of the unit circle $x^2+y^2=1$ and choose to consider y as our dependent variable and x as our independent variable, then there are functions of x implicitly defined by

$$x^2 + y(x)^2 = 1. \tag{8}$$

Rewriting the equation in this fashion might seem like overkill but students get so confused about this topic one shouldn't be afraid to spell everything out. Solving for $y(x)$ *explicitly*, we find two branches given by

$$y_1(x) = \sqrt{1-x^2} \text{ and } y_2(x) = -\sqrt{1-x^2}. \tag{9}$$

Now, we could compute the derivative of each branch separately but it turns out a special technique called *implicit differentiation* allows us to treat both cases simultaneously. The main idea underlying this technique is that we can treat differentiation as an "operator" that eats (differentiable) functions and spits out their derivative. Thinking in this way, the operator "$\frac{d}{dx}$" can be applied to both sides of our equation (it could be highlighted to students that this is a point of view where the Leibniz notation is quite useful) as follows

$$\frac{d}{dx}\left(x^2 + y(x)^2\right) = \frac{d}{dx}\left(1\right). \tag{10}$$

Using our differentiation rules (*which ones?*) we can simplify this expression as

$$\frac{d}{dx}\left(x^2\right) + \frac{d}{dx}\left(y(x)^2\right) = \frac{d}{dx}\left(1\right) \tag{11}$$

and then

$$2x + 2\,y(x)\cdot\frac{d}{dx}y(x) = 0. \tag{12}$$

Why did we use the chain rule when computing the derivative of $y(x)^2$ but not when computing the derivative of x^2? Solving for $\frac{d}{dx}y(x) = y'(x)$ we then find that

$$y'(x) = \frac{-x}{y(x)}. \tag{13}$$

What crucial properties of our equation did we use along the way? It should be highlighted that when we write $\frac{d}{dx}y(x)$ or $y'(x)$ we are *assuming* the implicitly defined functions being considered are differentiable at x. It should also be pointed out that when we divided by $y(x)$ we were, as usual, assuming that it was not equal to zero! In our case, we happen to know exactly what the implicitly defined functions are: $y_1(x) = \sqrt{1-x^2}$ and $y_2(x) = -\sqrt{1-x^2}$. *Where are these implicitly defined* *functions differentiable?* Everywhere except at $x = \pm 1$ which is precisely where they are equal to zero! *Was this an accident?*

As a sanity check, it is worthwhile to compute the derivatives of $y_1(x)$ and $y_2(x)$ explicitly to see that they are special cases of our formula. This also helps to emphasize that the role of $y(x)$ in the formula changes depending on the particular implicitly defined function we are considering. *Which of the two techniques was most labour-intensive?* It should be indicated to students by a more complicated concrete example, say $y^2 - y^3 = x^2$, that the latter technique may not only be more labour-intensive but actually impossible to carry out in practise!

To summarize, implicit differentiation is kind of like a western: you shoot first (differentiate both sides and solve for the derivative you seek) and ask questions later (make sure what you did makes sense). More precisely, given an equation in variables $\square$ and $\triangle$ where it is hard to solve for $\triangle$ explicitly in terms of $\square$, you can try to find the derivative of implicitly defined functions $\triangle(\square)$ as follows:

(1) Rewrite your equation making the choice of your dependent "$\triangle$" and independent "$\square$" variables clear, i.e., replacing "$\triangle$" by "$\triangle(\square)$".
(2) Apply the operator $\frac{d}{d\square}$ to both sides of the equation without forgetting to use the chain rule whenever necessary.
(3) Solve for $\frac{d}{d\square}(\triangle(\square)) = \triangle'(\square)$ keeping track of potentially problematic steps.
(4) Make sure to rule out all values of $\square$ where one of your steps would break down (e.g. divisions by zero).