

## INTRODUCTION TO DERIVATIVES

MATH 152, SECTION 55 (VIPUL NAIK)

**Corresponding material in the book:** Sections 3.1, 3.2, 3.3, 3.5.

**Difficulty level:** Moderate if you remember derivatives to at least the AP Calculus level. Otherwise, hard. Some things are likely to be new.

**Covered in class?:** Yes, but we'll go *very quickly* over most of the stuff that you would have seen at the AP level, and will focus much more on the conceptual interpretation and the algebraic-verbal-graphical-numerical nexus.

**What students should definitely get:** The definition of derivative as a limit. The fact that the left-hand derivative equals the left-hand limit and the right-hand derivative equals the right-hand limit. Also, the derivative is the slope of the tangent to the graph of the function. Graphical interpretation of tangents and normals. Finding equations of tangents and normals. Leibniz and prime notation. The sum rule, difference rule, product rule, and quotient rule. The chain rule for composition. Differentiation polynomials and rational functions.

**What students should hopefully get:** The idea of derivative as an *instantaneous rate of change*, the notion of *difference quotient* and *slope of chord*, and the fact that the definitions of tangent that we use for circles don't apply for general curves. How to find tangents and normals to curves from points outside them. Subtle points regarding tangent lines.

### EXECUTIVE SUMMARY

#### 0.1. Derivatives: basics. Words ...

- (1) For a function  $f$ , we define the *difference quotient* between  $w$  and  $x$  as the quotient  $(f(w) - f(x))/(w - x)$ . It is also the slope of the line joining  $(x, f(x))$  and  $(w, f(w))$ . This line is called a *secant line*. The segment of the line between the points  $x$  and  $w$  is sometimes termed a *chord*.
- (2) The limit of the difference quotient is defined as the *derivative*. This is the slope of the *tangent line* through that point. In other words, we define  $f'(x) := \lim_{w \rightarrow x} \frac{f(w) - f(x)}{w - x}$ . This can also be defined as  $\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$ .
- (3) If the derivative of  $f$  at a point  $x$  exists, the function is termed *differentiable* at  $x$ .
- (4) If the derivative at a point exists, then the tangent line to the graph of the function exists and its slope equals the derivative. The tangent line is horizontal if the derivative is zero. Note that if the derivative exists, then the tangent line cannot be vertical.
- (5) Here are some misconceptions about tangent lines: (i) that the tangent line is the line perpendicular to the radius (this makes sense only for circles) (ii) that the tangent line does not intersect the curve at any other point (this is true for some curves but not for others) (iii) that any line other than the tangent line intersects the curve at at least one more point (this is always false – the vertical line through the point does not intersect the curve elsewhere, but is not the tangent line if the function is differentiable).
- (6) In the Leibniz notation, if  $y$  is functionally dependent on  $x$ , then  $\Delta y/\Delta x$  is the difference quotient – it is the quotient of the difference between the  $y$ -values corresponding to  $x$ -values. The limit of this, which is the derivative, is  $dy/dx$ .
- (7) The left-hand derivative of  $f$  is defined as the left-hand limit for the derivative expression. It is  $\lim_{h \rightarrow 0^-} \frac{f(x+h) - f(x)}{h}$ . The right-hand derivative is  $\lim_{h \rightarrow 0^+} \frac{f(x+h) - f(x)}{h}$ .
- (8) Higher derivatives are obtained by differentiating again and again. The *second* derivative is the derivative of the derivative. The  $n^{\text{th}}$  derivative is the function obtained by differentiating  $n$  times. In prime notation, the second derivative is denoted  $f''$ , the third derivative  $f'''$ , and the  $n^{\text{th}}$  derivative

for large  $n$  as  $f^{(n)}$ . In the Leibniz notation, the  $n^{\text{th}}$  derivative of  $y$  with respect to  $x$  is denoted  $d^n y/dx^n$ .

- (9) Derivative of sum equals sum of derivatives. Derivative of difference is difference of derivatives. Scalar multiples can be pulled out.
- (10) We have the *product rule* for differentiating products:  $(f \cdot g)' = f' \cdot g + f \cdot g'$ .
- (11) We have the *quotient rule* for differentiating quotients:  $(f/g)' = (g \cdot f' - f \cdot g')/g^2$ .
- (12) The derivative of  $x^n$  with respect to  $x$  is  $nx^{n-1}$ .
- (13) The derivative of sin is cos and the derivative of cos is  $-\sin$ .
- (14) The chain rule says that  $(f \circ g)' = (f' \circ g) \cdot g'$

Actions ...

- (1) We can differentiate any polynomial function of  $x$ , or a sum of powers (possibly negative powers or fractional powers), by differentiating each power with respect to  $x$ .
- (2) We can differentiate any rational function using the quotient rule and our knowledge of how to differentiate polynomials.
- (3) We can find the equation of the tangent line at a point by first finding the derivative, which is the slope, and then finding the point's coordinates (which requires evaluating the function) and then using the point-slope form.
- (4) Suppose  $g$  and  $h$  are everywhere differentiable functions. Suppose  $f$  is a function that is  $g$  to the left of a point  $a$  and  $h$  to the right of the point  $a$ , and suppose  $f(a) = g(a) = h(a)$ . Then, the left-hand derivative of  $f$  at  $a$  is  $g'(a)$  and the right-hand derivative of  $f$  at  $a$  is  $h'(a)$ .
- (5) The  $k^{\text{th}}$  derivative of a polynomial of degree  $n$  is a polynomial of degree  $n - k$ , if  $k \leq n$ , and is zero if  $k > n$ .
- (6) We can often use the sum rule, product rule, etc. to find the values of derivatives of functions constructed from other functions simply using the values of the functions and their derivatives at specific points. For instance,  $(f \cdot g)'$  at a specific point  $c$  can be determined by knowing  $f(c)$ ,  $g(c)$ ,  $f'(c)$ , and  $g'(c)$ .
- (7) Given a function  $f$  with some unknown constants in it (so a function that is not completely known) we can use information about the value of the function and its derivatives at specific points to determine those constant parameters.

## 0.2. Tangents and normals: geometry. Words...

- (1) The normal line to a curve at a point is the line perpendicular to the tangent line. Since the tangent line is the best linear approximation to the curve at the point, the normal line can be thought of as the line best approximating the perpendicular line to the curve.
- (2) The angle of intersection between two curves at a point of intersection is defined as the angle between the tangent lines to the curves at that point. If the slopes of the tangent lines are  $m_1$  and  $m_2$ , the angle is  $\pi/2$  if  $m_1 m_2 = -1$ . Otherwise, it is the angle  $\alpha$  such that  $\tan \alpha = |m_1 - m_2|/(|1 + m_1 m_2|)$ .
- (3) If the angle between two curves at a point of intersection is  $\pi/2$ , they are termed *orthogonal* at that point. If the curves are orthogonal at all points of intersection, they are termed *orthogonal curves*.
- (4) If the angle between two curves at a point of intersection is 0, that means they have the same tangent line. In this case, we say that the curves *touch* each other or are *tangent* to each other.

Actions...

- (1) The equation of the normal line to the graph of a function  $f$  at the point  $(x_0, f(x_0))$  is  $f'(x_0)(y - f(x_0)) + (x - x_0) = 0$ . The slope is  $-1/f'(x_0)$ .
- (2) To find the angle(s) of intersection between two curves, we first find the point(s) of intersection, then compute the value of derivative (or slope of tangent line) to both curves, and then finally plug that in the formula for the angle of intersection.
- (3) It is also possible to find all tangents to a given curve, or all normals to a given curve, that pass through a given point *not* on the curve. To do this, we set up the generic expression for a tangent line or normal line to the curve, and then plug into that generic expression the specific coordinates of the point, and solve. For instance, the generic equation for the tangent line to the graph of a

function  $f$  is  $y - f(x_1) = f'(x_1)(x - x_1)$  where  $(x_1, f(x_1))$  is the point of tangency. Plugging in the point  $(x, y)$  that we know the curve passes through, we can solve for  $x_1$ .

- (4) In many cases, it is possible to determine geometrically the number of tangents/normals passing through a point outside the curve. Also, in some cases, the algebraic equations may not be directly solvable, but we may be able to determine the number and approximate location of the solutions.

### 0.3. Deeper perspectives on derivatives. Words... (these points were all seen in the quiz on Chapter 3)

- (1) A continuous function that is everywhere differentiable need not be everywhere continuously differentiable.
- (2) If  $f$  and  $g$  are functions that are both continuously differentiable (i.e., they are differentiable and their derivatives are continuous functions), then  $f + g$ ,  $f - g$ ,  $f \cdot g$ , and  $f \circ g$  are all continuously differentiable.
- (3) If  $f$  and  $g$  are functions that are both  $k$  times differentiable (i.e., the  $k^{\text{th}}$  derivatives of the functions  $f$  and  $g$  exist), then  $f + g$ ,  $f - g$ ,  $f \cdot g$ , and  $f \circ g$  are also  $k$  times differentiable.
- (4) If  $f$  and  $g$  are functions that are both  $k$  times continuously differentiable (i.e., the  $k^{\text{th}}$  derivatives of both functions exist and are continuous) then  $f + g$ ,  $f - g$ , and  $f \cdot g$ , and  $f \circ g$  are also  $k$  times continuously differentiable.
- (5) If  $f$  is  $k$  times differentiable, for  $k \geq 2$ , then it is  $k - 1$  times continuously differentiable, i.e., the  $(k - 1)^{\text{th}}$  derivative of  $f$  is a continuous function.
- (6) If a function is *infinitely differentiable*, i.e., it has  $k^{\text{th}}$  derivatives for all  $k$ , then its  $k^{\text{th}}$  derivatives are continuous functions for all  $k$ .

## 1. BARE-BONES INTRODUCTION

**1.1. Down and dirty with derivatives.** The order in which I'll do things is to quickly introduce to you the mathematical formalism for derivatives, which is really quite simple, and give you a quick idea of the conceptual significance, and we'll return to the conceptual significance from time to time, as is necessary.

The derivative of a function at a point measures the *instantaneous rate of change* of the function at the point. In other words, it measures how quickly the function is changing. Or, it measures the velocity of the function. These are vague ideas, so let's take a closer look.

Let's say I am riding a bicycle on a road, and I encounter a sign saying "speed limit: 30 mph". Well, that sign basically says that the speed limit should be 30 mph, but what speed is it referring to? We know that speed is distance covered divided by time taken, but if I start out from home in the morning, get stuck in a traffic jam on my way, and then, once the road clears, ride at top speed to get to my destination, I might have taken five hours to just travel five miles. So my speed is 1 mile per hour. But maybe I *still* went over the speed limit, because maybe it was the case that at the point where I crossed the speed limit sign, I was going really really fast. So, what's really relevant isn't my *average* speed from the moment I began my ride to the moment I ended, but the speed at the particular *instant* that I crossed that sign. But how do we measure the speed at a particular instant?

Well, one thing I could do is maybe measure the time it takes me to get from the lamp post just before that sign to the lamp post just after that sign. So if that distance is  $a$  and the time it took me to travel that distance is  $b$ , then the speed is  $a/b$ . And that might be a more relevant indicator since it is the speed in a small interval of time around where I saw that sign. But it still does not tell the entire story, because it might have happened that I was going pretty fast between the lamp posts, but I slowed down for a fraction of a second while crossing that sign. So maybe I still didn't technically break the law because I was slow enough *at the instant* that I was crossing the sign.

What we're really trying to measure here is the change in my position, or the distance I travel, divided by the time, but we're trying to measure it for a smaller and smaller interval of time around that crucial time point when I cross the signpost. Smaller and smaller and smaller and smaller... this suggests that a limit lurks, and indeed it does.

Let's try to formalize this. Suppose the function giving my position in terms of time is  $f$ . So, for a time  $t$ , my position is given by  $f(t)$ . And let's say that the point in time when I crossed the sign post was the point  $c$ . That means that at time  $t = c$ , I crossed the signpost, which is located at the point  $f(c)$ .

Now, let's say I want to calculate my speed immediately after I cross the sign post. So I pick some  $t$  that's just a little larger than  $c$ , and I look at the speed from time  $c$  to time  $t$ . What I get is:

$$\frac{f(t) - f(c)}{t - c}$$

This gives the *average* speed from time  $c$  to time  $t$ . How do we narrow this down to  $t = c$ ? If we try to plug in  $t = c$  in the formula, we get something of the form  $0/0$ , which is not defined. So, what do we do?

This is where we need the crucial idea – we need the idea of a *limit*. So, the instantaneous speed just *after* crossing the signpost is defined as:

$$\lim_{t \rightarrow c^+} \frac{f(t) - f(c)}{t - c}$$

This is the limit of the *average speed* over the time interval  $[c, t]$ , as  $t$  approaches  $c$  from the right side. Similarly, the instantaneous speed just before crossing the signpost is:

$$\lim_{t \rightarrow c^-} \frac{f(c) - f(t)}{c - t} = \lim_{t \rightarrow c^-} \frac{f(t) - f(c)}{t - c}$$

Notice that both these are taking limits of the same expression, except that one is from the right and one is from the left. If both limits exist and are equal, then this is the instantaneous speed. And once we've calculated the instantaneous speed, we can figure out whether it is greater than what the signpost capped speed at.

**1.2. Understanding derivatives – formal definition.** The derivative of a function  $f$  at a point  $x$  is defined as the following limit, if it exists:

$$f'(x) := \lim_{w \rightarrow x} \frac{f(w) - f(x)}{w - x}$$

if this limit exists! If the limit exists, then we say that the function  $f$  is *differentiable* at the point  $x$  and the value of the derivative is  $f'(x)$ .

The *left-hand derivative* is defined as the *left-hand limit* for the above expression:

$$\lim_{w \rightarrow x^-} \frac{f(w) - f(x)}{w - x}$$

The *right-hand derivative* is defined as the *right-hand limit* for the above expression:

$$\lim_{w \rightarrow x^+} \frac{f(w) - f(x)}{w - x}$$

The expression whose limit we are trying to take in each of these cases is sometimes called a *difference quotient* – it is the quotient between the difference in the function values and the difference in the input values to the function.

So, in words:

The derivative of  $f$  at  $x$  = The limit, as  $w$  approaches  $x$ , of the difference quotient of  $f$  between  $w$  and  $x$

Now, what I've told you so far is about all that you need to know – the basics of the definition of derivative. But the typical way of thinking about derivatives uses a slightly different formulation of the definition, so I'll just talk about that. That's related to the idea of substituting variables.

**1.3. More conventional way of writing derivatives.** Recall that  $\lim_{w \rightarrow x} g(w) = \lim_{h \rightarrow 0} g(x + h)$ . This is one of those rules for substitution we use to evaluate some trigonometric limits. And we typically use this rule when writing down the definition of derivatives. So, with the  $h \rightarrow 0$  convention, we define:

$$f'(x) := \lim_{h \rightarrow 0} \frac{f(x + h) - f(x)}{h}$$

What this is saying is that what we're trying to measure is the quotient of the difference in the values of  $f$  for a small change in the value of the input for  $f$ , near  $x$ . And then we're taking the limit as the

increment gets smaller and smaller. So, with this notation, the left-hand derivative becomes the above limit with  $h$  approaching 0 from the negative side and the right-hand derivative becomes the above limit with  $h$  approaching 0 from the positive side. In symbols:

$$\text{LHD of } f \text{ at } x = \lim_{h \rightarrow 0^-} \frac{f(x+h) - f(x)}{h}$$

and:

$$\text{RHD of } f \text{ at } x = \lim_{h \rightarrow 0^+} \frac{f(x+h) - f(x)}{h}$$

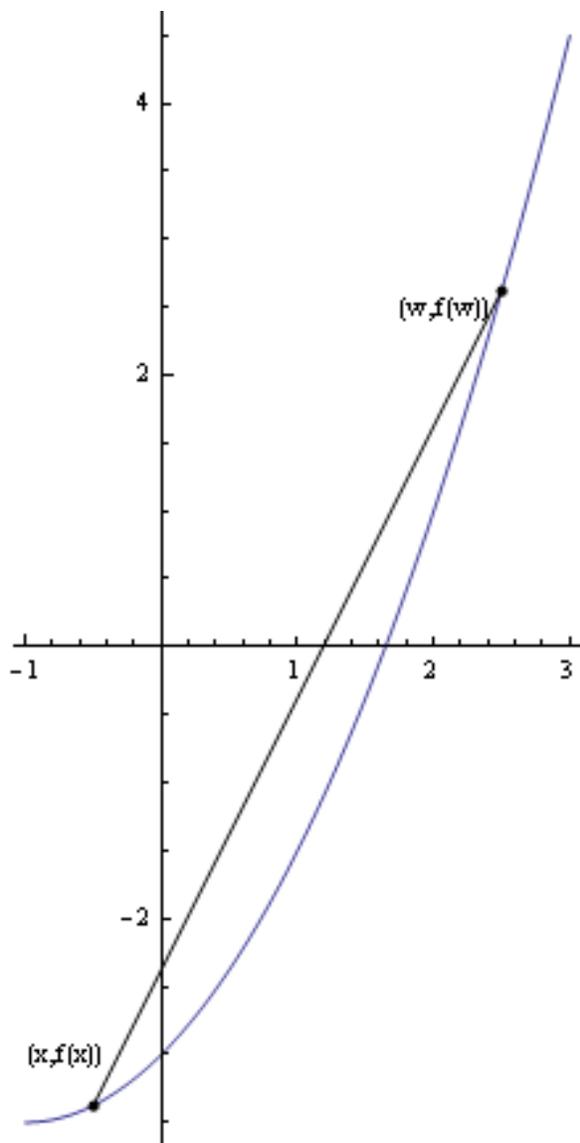
**1.4. Preview of graphical interpretation of derivatives.** Consider the graph of the function  $f$ , so we have a graph of  $y = f(x)$ . Then, say we're looking at two points  $x$  and  $w$ , and we're interested in the difference quotient:

$$\frac{f(w) - f(x)}{w - x}$$

What is this measuring? Well, the point  $(x, f(x))$  represents the point on the graph with its first coordinate  $x$ , and the point  $(w, f(w))$  represents the point on the graph with its first coordinate  $w$ . So, what we're doing is taking these two points, and taking the quotient of the difference in their  $y$ -coordinates by the difference in their  $x$ -coordinates. And, now this is the time to revive some of your coordinate geometry intuition, because this is essentially the *slope* of the line joining the points  $(x, f(x))$  and  $(w, f(w))$ . Basically, it is the *rise* over the *run* – the vertical change divided by the horizontal change.

So, one way of remembering this is that:

Difference quotient of  $f$  between points  $x$  and  $w = \text{Slope of line joining } (x, f(x)) \text{ and } (w, f(w))$

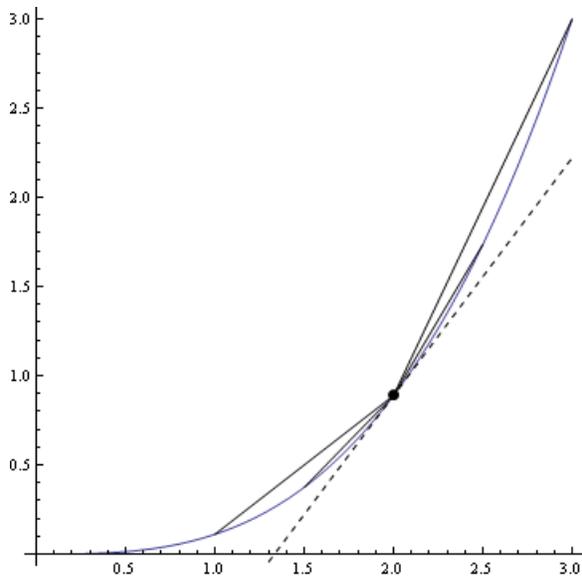


Also, note that the difference quotient is symmetric in  $x$  and  $w$ , in the sense that if we interchange the roles of the two points, the difference quotient is unaffected. So is the slope of the line joining the two points.

Now, let's try to use this understanding of the difference quotient as the slope of the line joining two points to understand what the derivative means geometrically. So, here's some terminology – this line joining two points is termed a *secant line*. The *line segment* between the two points is sometimes called a *chord* – just like you might have seen chords and secants for circles, we're now using the terminology for more general kinds of curves.

So, what we're trying to do is understand: as the point  $w$  gets closer and close to  $x$ , what happens to the slope of the secant line? Well, the pictures make it clear that what's happening is that the secant line is coming closer and closer to a line that *just touches* the function at the point  $(x, f(x))$  – what is called the *tangent line*. So, we get:

Derivative of  $f$  at  $x =$  Slope of tangent line to graph of  $f$  at  $(x, f(x))$



1.5. **More on the tangent line.** I'll have a lot to say on the tangent line and derivatives – our treatment of these issues has barely scratched the surface so far. But before proceeding too much, here's just one very important point I want to make.

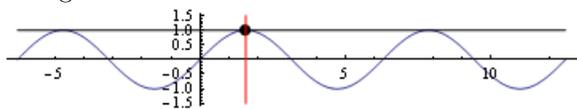
Some of you have done geometry and you've seen the concept of tangent lines to circles. You'll recall that the tangent line at a point on a circle can be defined in a number of ways: (i) as the line perpendicular to the line joining the point to the center of the circle (the latter is also called the radial line), and (ii) as the unique line that intersects the circle at exactly that one point.

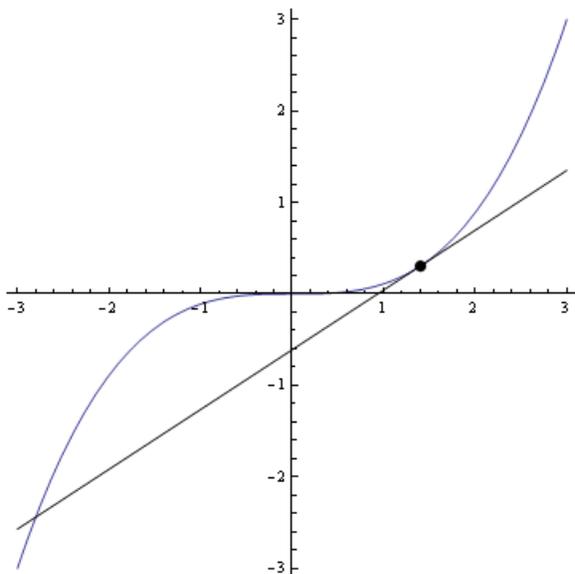
Now, it's obvious that you need to discard definition (i) when thinking about tangent lines to general curves, such as graphs of functions. That's because a curve in general has no center. It doesn't even have an inside or an outside. So definition (i) doesn't make sense.

But definition (ii) makes sense, right? Well, it makes sense, but it is wrong. For general curves, it is wrong. It's wrong because, first, the tangent line may well intersect the curve at other points, and second, there may be many non-tangent lines that also intersect the curve at just that one point.

For instance, consider the function  $f(x) := \sin x$ . We know the graph of this function. Now, our intuition of the tangent line should say that, at the point  $x = \pi/2$ , the tangent line should be the limit of chords between  $\pi/2$  and very close points, and those chords are becoming almost horizontal, so the tangent line should be horizontal. But – wow! The tangent line intersects the graph of the function at *infinitely* many other points. So banish from your mind the idea that the tangent line doesn't intersect the curve anywhere else.

And, there are other lines that are very much *not* tangent lines that intersect the curve at exactly that one point. For instance, you can make the vertical line through the point  $(\pi/2, 1)$ , so that's the line  $x = \pi/2$ , and that intersects the curve at exactly one point – but it's far from the tangent line! And you can make a lot of other lines – ones that are close to vertical, that intersect the curve at exactly that one point, but are not tangential.





So, thinking of the tangent line as the line that intersects the curve at exactly that one point is a flawed way of thinking about tangent lines. So what's the right way? Well, one way is to think of it as a limit of secant lines, which is what we discussed. But if you want a way of thinking that isn't really limits-based, think of it as the line *best line approximating the curve near the point*. It's the answer to the question: if we wanted to treat this function as a linear function near the point, what would that linear function look like? Or, you can go back to the Latin roots and note that tangent comes from *touch*, so the feeling of just touching, or just grazing, is the feeling you should have.

**1.6. Actual computations of derivatives.** Suppose we have  $f(x) := x^2$ . We want to calculate  $f'(2)$ . How do we do this?

Let's use the first definition, in terms of  $w$ , which gives:

$$f'(2) = \lim_{w \rightarrow 2} \frac{w^2 - 4}{w - 2} = \lim_{w \rightarrow 2} \frac{(w + 2)(w - 2)}{w - 2} = \lim_{w \rightarrow 2} (w + 2) = 4$$

So, there, we have it,  $f'(2) = 4$ .

Now this is the time to pause and think one level more abstractly. We have a recipe that allows us to calculate  $f'(x)$  for any specific value of  $x$ . But it seems painful to have to do the limit calculation for each  $x$ . So if I now want to find  $f'(1)$ , I have to do that calculation again replacing the 2 by 1. So, in order to save time, let's try to find  $f'(x)$  for a *general* value of  $x$ .

$$f'(x) = \lim_{w \rightarrow x} \frac{w^2 - x^2}{w - x} = \lim_{w \rightarrow x} \frac{(w + x)(w - x)}{w - x} = \lim_{w \rightarrow x} (w + x) = 2x$$

Just to illustrate, let me use the  $h \rightarrow 0$  formulation:

$$f'(x) = \lim_{h \rightarrow 0} \frac{(x + h)^2 - x^2}{h} = \lim_{h \rightarrow 0} \frac{h(2x + h)}{h} = \lim_{h \rightarrow 0} (2x + h) = 2x$$

So good, both of these give the same answer, so we now have a *general* formula that says that for  $f(x) = x^2$ ,  $f'(x) = 2x$ .

Note that this says that the function  $f$  is differentiable *everywhere*, which is a pretty strong statement, and the derivative is *itself* a continuous function, namely the linear function  $2x$ .

Let's see what this means graphically. The graph of  $f(x) := x^2$  looks like a parabola. It passes through  $(0, 0)$ . What the formula tells us is that  $f'(0) = 0$ . Which means that the tangent at 0 is horizontal, which it surely is. The formula tells us that  $f'(1) = 2$ , so the tangent line at 1 has its  $y$ -coordinate rising twice as fast as its  $x$ -coordinate, which is again sort of suggested by the graph.  $f'(2) = 4$ ,  $f'(3) = 6$ , and so on. Which means that the slope of the tangent line increases as  $x$  gets larger – again, suggested by the graph

– it’s getting steeper and steeper. And for  $x < 0$ ,  $f'(x)$  is negative, which is again pictorially clear because the slope of the tangent line is negative.

Let’s do one more thing: write down the equation of the tangent line at  $x = 3$ . Remember that if a line has slope  $m$  and passes through the point  $(x_0, y_0)$ , then the equation of the line is given by:

$$y - y_0 = m(x - x_0)$$

This is called the *point-slope form*.

So what do we know about the point  $x = 3$ ? For this point, the value of the function is 9, so one point that we for sure know is on the tangent line is the point  $(3, 9)$ . And the tangent line has slope  $2(3) = 6$ . So, plugging into the point-slope form, we get:

$$y - 9 = 6(x - 3)$$

And that’s it – that’s the equation of the line. You can rearrange stuff and rewrite this to get:

$$6x - y = 9$$

**1.7. Differentiable implies continuous.** Important fact: if  $f$  is differentiable at the point  $x$ , then  $f$  is continuous at the point  $x$ . Why? Below is a short proof.

Consider, for  $h \neq 0$ :

$$f(x + h) - f(x) = \frac{f(x + h) - f(x)}{h} \cdot h$$

Now, taking the limit as  $h \rightarrow 0$ , and using the fact that  $f$  is differentiable at  $x$ , we get:

$$\lim_{h \rightarrow 0} (f(x + h) - f(x)) = \lim_{h \rightarrow 0} \frac{f(x + h) - f(x)}{h} \cdot \lim_{h \rightarrow 0} h$$

The first limit on the right is  $f'(x)$ , and the second limit is 0, so the product is 0, and we get:

$$\lim_{h \rightarrow 0} (f(x + h) - f(x)) = 0$$

Since  $f(x)$  is a constant (independent of  $h$ ) it can be pulled out of the limit, and rearranging, we get:

$$\lim_{h \rightarrow 0} f(x + h) = f(x)$$

Which is precisely what it means to say that  $f$  is continuous at  $x$ .

What does this mean intuitively? Well, intuitively, for  $f$  to be differentiable at  $x$  means that the rate of change at  $x$  is finite. But if the function is jumping, or suddenly changing value, or behaving in any of those odd ways that we saw could occur at discontinuities, then the instantaneous rate of change wouldn’t make sense. It’s like if you apparate/teleport, you cannot really measure the *speed* at which you traveled. So, we can hope for a function to be differentiable at a point only if it’s continuous at the point.

**Aside: Differentiability and the  $\epsilon - \delta$  definition.** (Note: This is potentially confusing, so ignore it if it confuses you). The derivative of  $f$  at  $x = p$  measures the rate of change of  $f$  at  $c$ . Suppose the derivative value is  $\nu = f'(p)$ . Then, this means that, close to  $c$ , the graph looks like a straight line with slope  $\nu$  and passing through  $(p, f(p))$ .

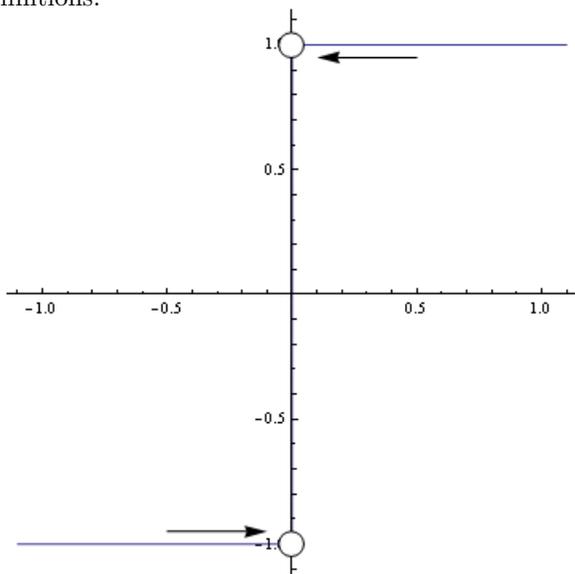
If it were exactly the straight line, then the strategy for an  $\epsilon - \delta$  proof would be to pick  $\delta = \epsilon/|\nu|$ . In practice, since the graph is not exactly a straight line, we need to pick a slightly different (usually smaller)  $\delta$  to work. Recall that for  $f$  a quadratic function  $ax^2 + bx + c$ , we chose  $\delta = \min\{1, \epsilon/(|a| + |2ap + b|)\}$ . In this case,  $\nu = 2ap + b$ . However, to make the proof go through, we need to pad an extra  $|a|$  in the denominator, and that extra padding is because the function isn’t quite linear.

**1.8. Derivatives of linear and constant functions.** You can just check this – we’ll deal more with this in the next section:

- (1) The derivative of a constant function is zero everywhere.
- (2) The derivative of a linear function  $f(x) := ax + b$  is everywhere equal to  $a$ .

**1.9. Handling piecewise definitions.** Let's think about functions defined piecewise. In other words, let us think about functions that change definition, i.e., that are governed by different definitions on different parts of the domain. How do we calculate the derivative of such a function?

The first rule to remember is that the derivative of a function at a point depends only on what happens at the point and very close to the point. So, if the point is contained in an open interval where one definition rules, then we can just differentiate the function using that definition. On the other hand, if the point is surrounded, very closely, by points where the function takes different definitions, we need to handle all these definitions.

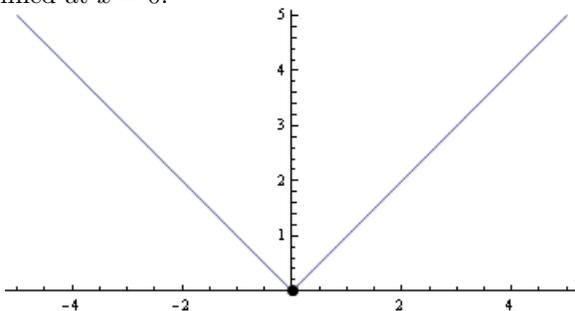


Let's begin with the signum function. Define  $f(x) = x/|x|$  for  $x \neq 0$ , and define  $f(0) = 0$ . So  $f$  is constant at  $-1$  for  $x < 0$ ,  $f(0) = 0$ , and  $f$  is constant at  $1$  for  $x > 0$ .

So, what can we say about the derivative of  $f$ ? Well, think about what's happening for  $x = -3$ , say. Around  $x = -3$ , the function is constant at  $-1$ . Yes, it's going to change its value *far in the future*, but that doesn't affect the derivative. The derivative is a *local measurement*. So we simply need to calculate the derivative of the function that's constant-valued to  $-1$ , for  $x = -3$ . You can apply the definition and check that the derivative is  $0$ , which makes sense, because constant functions are unchanging.

Similarly, at the point  $x = 4$ , the function is constant at  $1$  around the point, so the derivative is again  $0$ .

What about the point  $x = 0$ ? Here, you can see that the function isn't continuous, because it's jumping, so the derivative is not defined. So the derivative of  $f$  is the function that is  $0$  for all  $x \neq 0$ , and it isn't defined at  $x = 0$ .



Let's look at another function, this time  $g(x) := |x|$ . For  $x < 0$ , this function is  $g(x) = -x$ , and for  $x > 0$ , this function is  $g(x) = x$ . The crucial thing to note here is that  $g$  is continuous at  $0$ , so *both definitions* – the definition  $g(x) = x$  and the definition  $g(x) = -x$  apply at  $x = 0$ .

For  $x < 0$ , for instance,  $x = -3$ , we can treat the function as  $g(x) = -x$ , and differentiate this function, What we'll get is that  $g'(x) = -1$  for  $x < 0$ . And for  $x > 0$ , we can use  $g(x) = x$ , so the derivative, which we can again calculate, is  $g'(x) = 1$  for all  $x > 0$ .

But what about  $x = 0$ ? The function is continuous at  $x = 0$ , so we cannot do the cop-out that we did last time. So we need to think in terms of left and right: calculate the left-hand derivative and the right-hand derivative.

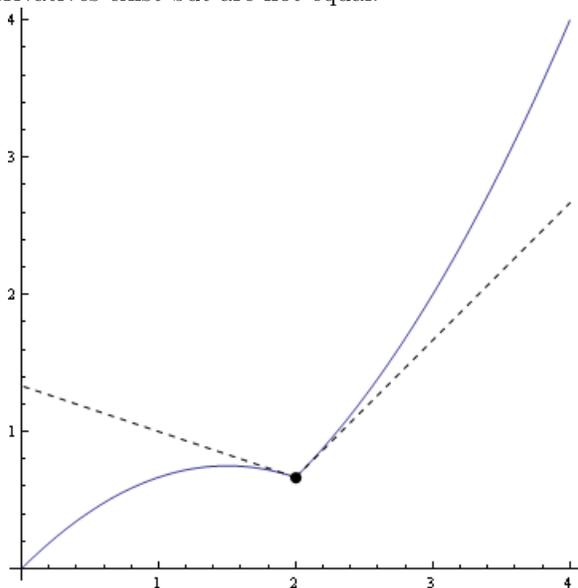
Now, the important idea here is that since the definition  $g(x) = -x$  also applies at  $x = 0$ , the formula for  $g'$  that we calculated for  $-x$  also applies to the *left-hand derivative* at 0. So the left-hand derivative at  $x = 0$  is  $-1$ , which again you can see from the graph. And the right-hand derivative at  $x = 0$  is  $+1$ , because the  $g(x) = x$  definition also applies at  $x = 0$ .

So, the function  $g'$  is  $-1$  for  $x$  negative, and  $+1$  for  $x$  positive, but it isn't defined for  $x = 0$ . However, the left-hand derivative at 0 is defined – it's  $-1$ . And so is the right-hand derivative – it's  $+1$ .

So, the upshot is: Suppose a function  $f$  has a definition  $f_1$  to the left of  $c$  and a definition  $f_2$  to the right of  $c$ , and both definitions give everywhere differentiable functions on all real numbers:

- (1)  $f$  is differentiable at all points other than  $c$ .
- (2) If  $f_1(c) = f_2(c)$  and these agree with  $f(c)$ , then  $f$  is continuous at  $c$ .
- (3) If  $f$  is continuous from the left at  $c$ , the left-hand derivative at  $c$  equals the value of the derivative  $f'_1$  evaluated at  $c$ .
- (4) If  $f$  is continuous from the right at  $c$ , the right-hand derivative at  $c$  equals the value of the derivative  $f'_2$  evaluated at  $c$ .
- (5) In particular, if  $f$  is continuous at  $c$ , and  $f'_1(c) = f'_2(c)$ , then  $f$  is differentiable at  $c$ .

Here is an example of a picture where the function is continuous but changes direction, so the one-sided derivatives exist but are not equal:



### 1.10. Three additional examples.

$$f(x) := \begin{cases} x^2, & x \leq 0 \\ x^3, & x > 0 \end{cases}$$

Here  $f_1(x) = x^2$  and  $f_2(x) = x^3$ . Both piece functions are differentiable everywhere. Note that  $f$  is continuous at 0, since  $f_1(0) = f_2(0) = 0$ . The left hand derivative at 0 is  $f'_1(x) = 2x$  evaluated at 0, giving 0. The right hand derivative at 0 is  $f'_2(x) = 3x^2$  evaluated at 0, giving 0. Since both one-sided derivatives agree,  $f$  is differentiable at 0 and  $f'(0) = 0$ .

Now consider the example:

$$f(x) := \begin{cases} x^2 + 1, & x \leq 0 \\ x^3, & x > 0 \end{cases}$$

This function is not continuous at 0 because the two functions  $x^2 + 1$  and  $x^3$  do not agree at 0. Note that since 0 is included in the left definition,  $f$  is left continuous at 0, and  $f(0) = 1$ . We can also calculate the left hand derivative: it is  $2x$  evaluated at 0, which is  $2 \cdot 0 = 0$ .

However, since the function is not continuous from the right at 0, we *cannot* calculate the right hand derivative by plugging in the derivative of  $x^3$  at 0. In fact, we should not feel the need to do so, because *the function is not right continuous at 0, so there cannot be a right hand derivative.*

Finally, consider the example:

$$f(x) := \begin{cases} x, & x \leq 0 \\ x^3, & x > 0 \end{cases}$$

This function is continuous at 0, and  $f(0) = 0$ . The left hand derivative at 0 is 1 and the right hand derivative is  $3(0)^2 = 0$ . Thus, the function is *not differentiable* at 0, though it has one-sided derivatives.

**1.11. Existence of tangent and existence of derivative.** Our discussion basically pointed to the fact that, for a function  $f$  defined around a point  $x$ , if the derivative of  $f$  at  $x$  exists, then there exists a tangent to the graph of  $f$  at the point  $(x, f(x))$  and the slope of that tangent line is  $f'(x)$ . There are a few subtleties related to this:

- (1) If both the left-hand derivative and the right-hand derivative at a point are defined, but are not equal, then the left-hand derivative is the tangent to the graph on the *left* side and the right-hand derivative is the tangent to the graph on the *right* side. There is no single tangent to the whole graph at the point.
- (2) In some cases, the tangent to the curve exists and is vertical. If the tangent exists and is vertical, then the derivative does not exist. In fact, a vertical tangent is the *only* situation where the tangent exists but the derivative does not exist.

An example of this is the function  $f(x) = x^{1/3}$  at the point  $x = 0$ . See Page 110 of the book for a more detailed discussion of this. We will get back to this function (and the general notion of vertical tangent) in far more gory detail in the near future.

## 2. RULES FOR COMPUTING DERIVATIVES

**2.1. Quick recap.** Before proceeding further, let's recall the definition of the derivative.

If  $f$  is a function defined around a point  $x$ , we define:

$$f'(x) := \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{w \rightarrow x} \frac{f(w) - f(x)}{w - x}$$

The left-hand derivative is defined as:

$$f'_l(x) := \lim_{h \rightarrow 0^-} \frac{f(x+h) - f(x)}{h} = \lim_{w \rightarrow x^-} \frac{f(w) - f(x)}{w - x}$$

The right-hand derivative is defined as:

$$f'_r(x) := \lim_{h \rightarrow 0^+} \frac{f(x+h) - f(x)}{h} = \lim_{w \rightarrow x^+} \frac{f(w) - f(x)}{w - x}$$

(The subscripts  $l$  and  $r$  are not standard notation, but are used for simplicity here).

**2.2. Differentiating constant and linear functions.** Suppose a function  $f$  is constant with constant value  $k$ . We want to show that the derivative of  $f$  is zero everywhere. Let's calculate the derivative of  $f$ .

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{k - k}{h} = \lim_{h \rightarrow 0} 0 = 0$$

So the derivative of a constant function is zero. And this makes sense because the derivative is the *rate of change* of the function. And a constant function is unchanging, so its derivative is zero.

Let's look at the derivative of a linear function  $f(x) := ax + b$ . Let's calculate the derivative:

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{a(x+h) + b - (ax + b)}{h} = \lim_{h \rightarrow 0} \frac{ah}{h} = a$$

So, for the function  $f(x) = ax + b$ , the derivative is a *constant* function whose value is equal to  $a$ . And this makes sense, because the graph of  $f(x) = ax + b$  is a straight line with slope  $a$ . And now remember that the derivative of a function equals the slope of the tangent line. But now, what's the tangent line to a straight line at a point? It is the line itself. So, the tangent line is the line itself, and its slope is  $a$ , which is what our calculations also show.

Remember that the tangent line is the *best linear approximation* to the curve, so if the curve is already a straight line, it coincides with the tangent line to it through any point.

**2.3. The sum rule for derivatives.** Suppose  $f$  and  $g$  are functions, both defined around a point  $x$ . The sum rule for derivatives states that if both  $f$  and  $g$  are differentiable at the point  $x$ , then  $f + g$  is also differentiable at the point  $x$ , and  $(f + g)'(x) = f'(x) + g'(x)$ . In words, *the sum of the derivatives equals the derivative of the sum*. This is the first part of Theorem 3.2.3. The proof is given below. It involves very simple manipulation. (You're not expected to know this proof, but this should be sort of manipulation that you eventually are comfortable reading and understanding).

So, what we have is:

$$(1) \quad f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

and:

$$(2) \quad g'(x) = \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h}$$

Now, we also have:

$$(f + g)'(x) = \lim_{h \rightarrow 0} \frac{(f + g)(x+h) - (f + g)(x)}{h} = \lim_{h \rightarrow 0} \frac{f(x+h) + g(x+h) - f(x) - g(x)}{h}$$

Simplifying further, we get:

$$(f + g)'(x) = \lim_{h \rightarrow 0} \frac{(f(x+h) - f(x)) + (g(x+h) - g(x))}{h} = \lim_{h \rightarrow 0} \left[ \left( \frac{f(x+h) - f(x)}{h} \right) + \left( \frac{g(x+h) - g(x)}{h} \right) \right]$$

Now, using the fact that the limit of the sum equals the sum of the limits, we can split the limit on the right to get:

$$(3) \quad (f + g)'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} + \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h}$$

Combining (1), (2), and (3), we obtain that  $(f + g)'(x) = f'(x) + g'(x)$ .

On a similar note, if  $\alpha$  is a constant, then the derivative of  $\alpha f$  is  $\alpha$  times the derivative of  $f$ . In other words:

$$(\alpha f)'(x) = \lim_{h \rightarrow 0} \frac{\alpha f(x+h) - \alpha f(x)}{h} = \lim_{h \rightarrow 0} \frac{\alpha(f(x+h) - f(x))}{h} = \alpha f'(x)$$

**2.4. Global application of these rules.** The rule for sums and scalar multiples apply at each point. In other words, if both  $f$  and  $g$  are differentiable at a point  $x$ , then  $f + g$  is differentiable at  $x$  and the derivative of  $f + g$  is the sum of the derivatives of  $f$  and  $g$  at the point  $x$ .

An easy corollary of this is that if  $f$  and  $g$  are everywhere differentiable functions, then  $f + g$  is also everywhere differentiable, and  $(f + g)' = f' + g'$ . In other words, what I've just done is applied the previous result at every point. And, so, the function  $(f + g)'$  is the sum of the functions  $f'$  and  $g'$ .

In words, *the derivative of the sum is the sum of the derivatives.*

**2.5. Rule for difference.** In a similar way that we handled sums and scalar multiples, we can handle differences, so we get  $(f - g)' = f' - g'$ .

In words, *the derivative of the difference is the difference of the derivatives.*

**2.6. Rule for product.** People less seasoned in calculus than you may now expect the rule for products to be: *the derivative of the product is the product of the derivatives.* But this is *wrong*.

The rule for the product is that, if  $f$  and  $g$  are differentiable at a point  $x$ , then the product  $f \cdot g$  is also differentiable at the point  $x$ , and:

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

We won't bother with a proof of the product rule, but it's an important result that you should know. In particular, again, when  $f$  and  $g$  are defined everywhere, we have:

$$(f \cdot g)' = f' \cdot g + f \cdot g'$$

**2.7. Rule for reciprocals and quotients.** If  $f$  is differentiable at a point  $x$ , and  $f(x) \neq 0$ , then the function  $(1/f)$  is also differentiable at the point  $x$ :

$$\left(\frac{1}{f}\right)'(x) = \frac{-f'(x)}{(f(x))^2}$$

If, at a point  $x$ ,  $f$  and  $g$  are both differentiable, and  $g(x) \neq 0$ , then:

$$\left(\frac{f}{g}\right)'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}$$

**2.8. Formula for differentiating a power.** Here's the formula for differentiating a power function: if  $f(x) = x^n$ , then  $f'(x) = nx^{n-1}$ . You can actually prove this formula for positive integer values of  $n$  using induction and the product rule (I've got a proof in the addendum at the end of this section). So, for instance, the derivative of  $x^2$  is  $2x$ . The derivative of  $x^5$  is  $5x^4$ . Basically, what you do is to take the thing in the exponent and pull it down as a coefficient and subtract 1 from it.

Okay, now there are plenty of things I want you to note here. First, note that the  $n$  in the exponent really should be a *constant*. So, for instance, we cannot use this rule to differentiate the function  $x^x$ , because in this case, the thing in the exponent is itself dependent on  $x$ .

The second thing you should note is that you are already familiar with some cases of this formula. For instance, consider the constant function  $x^0 = 1$ . The derivative of this is the function  $0x^{0-1} = 0x^{-1} = 0$ . Also, for the function  $x^1 = x$ , the derivative is 1, which is again something we saw. And for the function  $x^2$ , the derivative is  $2x^1 = 2x$ , which is what we saw last time.

The third thing you should think about is the scope and applicability of this formula. I just said that when  $n$  is a positive integer, we can show that for  $f(x) = x^n$ , we have  $f'(x) = nx^{n-1}$ . But in fact, this formula works not just for positive integers, but for all integers, and not just for all integers, even for non-integer exponents:

- (1) If  $f(x) = x^n$ ,  $x \neq 0$ , where  $n$  is a negative integer,  $f'(x) = nx^{n-1}$ .
- (2) If  $f(x) = x^r$ , where  $r$  is some real number, and  $x > 0$ , we have  $f'(x) = rx^{r-1}$ . We will study later what  $x^r$  means for arbitrary real  $r$  and positive real  $x$ .
- (3) The above formula also applies for  $x < 0$  where  $r$  is a rational number with denominator an odd integer (that's necessary to make sense of  $x^r$ ).

**2.9. Computing the derivative of a polynomial.** To differentiate a polynomial, what we do is use the rule for differentiating powers, and the rules for sums and scalar multiples.

For instance, to differentiate the polynomial  $f(x) := x^3 + 2x^2 + 5x + 7$ , what we do is differentiate each of the pieces. The derivative of the piece  $x^3$  is  $3x^2$ . The derivative of the piece  $2x^2$  is  $4x$ , the derivative of the piece  $5x$  is  $5$ , and the derivative of the piece  $7$  is  $0$ . So, the derivative is  $f'(x) = 3x^2 + 4x + 5$ .

Now, the way you do this is you write the polynomial, and with a bit of practice, which you did in high school, you can differentiate each individual piece mentally. And so you keep writing the derivatives one by one.

In particular, polynomial functions are everywhere differentiable, and the derivative of a polynomial function is another polynomial. Moreover, the degree of the derivative is one less than the degree of the original polynomial.

**2.10. Computing the derivative of a rational function.** What if we have a rational function  $Q(x) = f(x)/g(x)$ ?

Essentially, we use the formula for differentiating a quotient, which gives us:

$$Q'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}$$

Since both  $f$  and  $g$  are polynomials, we know how to differentiate them, so we are effectively done.

For instance, consider  $Q(x) = x^2/(x^3 - 1)$ . In this case:

$$Q'(x) = \frac{(x^3 - 1)(2x) - x^2(3x^2)}{(x^3 - 1)^2} = \frac{-2x - x^4}{(x^3 - 1)^2}$$

Note another important thing – since the denominator of  $Q'$  is the square of the denominator of  $Q$ , the points where  $Q'$  is not defined are the *same* as the point where  $Q$  is not defined. Thus, a rational function is differentiable wherever it is defined.

**Addendum: Proof by induction that the derivative of  $x^n$  is  $nx^{n-1}$ .** We prove this for  $n \geq 1$ .

*Base case for induction:* Here,  $f(x) = x^1 = x$ . We have:

$$f'(x) = \lim_{h \rightarrow 0} \frac{(x+h) - x}{h} = \lim_{h \rightarrow 0} 1 = 1 = 1x^0$$

Thus, for  $n = 1$ , we have  $f'(x) = nx^{n-1}$  where  $f(x) = x^n$ .

*Induction step:* Suppose the result is true for  $n = k$ . In other words, suppose it is true that for  $f(x) := x^k$ , we have  $f'(x) = kx^{k-1}$ . We want to show that for  $g(x) := x^{k+1}$ , we have  $g'(x) = (k+1)x^k$ .

Using the product rule for  $g(x) = xf(x)$ , we have:

$$g'(x) = 1f(x) + xf'(x)$$

Substituting the expression for  $f'(x)$  from the assumption that the statement is true for  $k$ , we get:

$$g'(x) = 1(x^k) + x(kx^{k-1}) = x^k + kx^k = (k+1)x^k$$

which is what we need to prove. This completes the induction step, and hence completes the proof.

### 3. GRAPHICAL INTERPRETATION OF DERIVATIVES

**3.1. Some review of coordinate geometry.** So far, our attention to coordinate geometry has been minimal: we discussed how to get the *equation* of the tangent line to the graph of a function at a point using the *point-slope form*, wherein we determine the point possibly through function evaluation and we determine the slope by computing the derivative of the function at the point.

Now, recall that the *slope* of a line equals  $\tan \theta$ , where  $\theta$  is the angle that the line makes with the  $x$ -axis, measured counter-clockwise from the  $x$ -axis. What are the kind of lines whose slope is not defined? These are the vertical lines, in which case  $\theta = \pi/2$ . Notice that since the slope is not defined, we see that if a function is differentiable at the point, then the tangent line cannot be vertical.

The slope of a line equals  $\tan \theta$ , where  $\theta$  is the angle that that line makes with the  $x$ -axis. And for some of you, thinking of that angle might be geometrically more useful than thinking of the slope as a number. We will study this in more detail a little later when we study the graphing of functions more intensely.

**Caveat: Axes need to be scaled the same way for geometry to work right.** When you use a graphing software such as Mathematica, or a graphing calculator for functions, you'll often find that the software or calculator automatically chooses different scalings for the two axes, so as to make the picture of the graph fit in nicely. If the axes are scaled differently, the geometry described here does not work. In particular, the derivative is no longer equal to  $\tan \theta$  where  $\theta$  is the angle with the horizontal; instead, we have to scale the derivative by the appropriate ratio between the scaling of the axes. Similarly, the angles of intersection between curves change and the notion of orthogonality is messed up.

However, the notion of tangency of curves discussed below does *not* get messed up.

**3.2. Perpendicular lines, angles between lines.** Two lines are said to be *orthogonal* or *perpendicular* if the angle between them is  $\pi/2$ . This means that if one of them makes an angle  $\theta_1$  with the  $x$ -axis and the other makes an angle  $\theta_2$  with the  $x$ -axis (both angles measured counter-clockwise from the  $x$ -axis) then  $|\theta_1 - \theta_2| = \pi/2$ . How do we characterize this in terms of slopes?

Well, we need to take a short detour here and calculate the formula for  $\tan(A-B)$ . Recall that  $\sin(A-B) = \sin A \cos B - \cos A \sin B$  and  $\cos(A-B) = \cos A \cos B + \sin A \sin B$ . When we take the quotient of these, we get:

$$\tan(A-B) = \frac{\sin A \cos B - \cos A \sin B}{\cos A \cos B + \sin A \sin B}$$

If both  $\tan A$  and  $\tan B$  are defined, this simplifies to:

$$\tan(A-B) = \frac{\tan A - \tan B}{1 + \tan A \tan B}$$

Now, what would happen if  $|A-B|$  is  $\pi/2$ ? In this case,  $\tan(A-B)$  should be undefined, which means that the denominator on the right side should be 0, which means that  $\tan A \tan B = -1$ .

Translating back to the language of slopes, we obtain that two lines (neither of which is vertical or horizontal) are perpendicular if the product of their slopes is  $-1$ .

More generally, given two lines with slopes  $m_1$  and  $m_2$ , we have the following formula for  $\tan \alpha$  where  $\alpha$  is the angle of intersection:

$$\tan \alpha = \left| \frac{m_1 - m_2}{1 + m_1 m_2} \right|$$

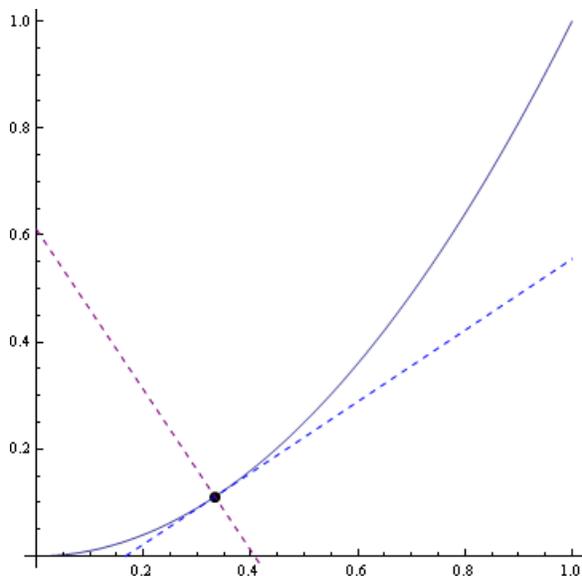
The absolute value means that we are not bothered about the direction (or *orientation* or *sense*) in which we are measuring the angle (if we care about the direction, we have to specify angle from which line to which line, measured clockwise or counter-clockwise).

**3.3. Normal lines.** Given a point on a curve, the *normal line* to the curve at that point is the line perpendicular to the tangent line to the curve at that point. This normal is the same normal you may have seen in *normal force* and other related ideas in classical mechanics.

Now, we know how to calculate the equation of the normal line, if we know the value of the derivative. let me spell this out with two cases:

- (1) If the derivative of a function at  $x = a$  is 0, then the tangent line is  $y = f(a)$  and the normal line is  $x = a$ .
- (2) If the derivative of a function at  $x = a$  is  $m \neq 0$ , then the tangent line is  $y - f(a) = m(x - a)$  and the normal line is  $y - f(a) = (-1/m)(x - a)$ . We got the slope of the normal line using the fact that the product of slopes of two perpendicular lines is  $-1$ .

Here is a picture of the usual case where the derivative is nonzero:



**3.4. Angle of intersection of curves.** Consider the graphs of the functions  $f(x) = x$  and  $g(x) = x^2$ . These curves intersect at two points:  $(0, 0)$  and  $(1, 1)$ . We would like to find the angles of intersection between the curves at these two points.

The *angle of intersection* between two curves at a given point is the angle between the tangent lines to the curve at that point. In particular, if  $\alpha$  is this angle of intersection, then for  $\alpha \neq \pi/2$ , we have:

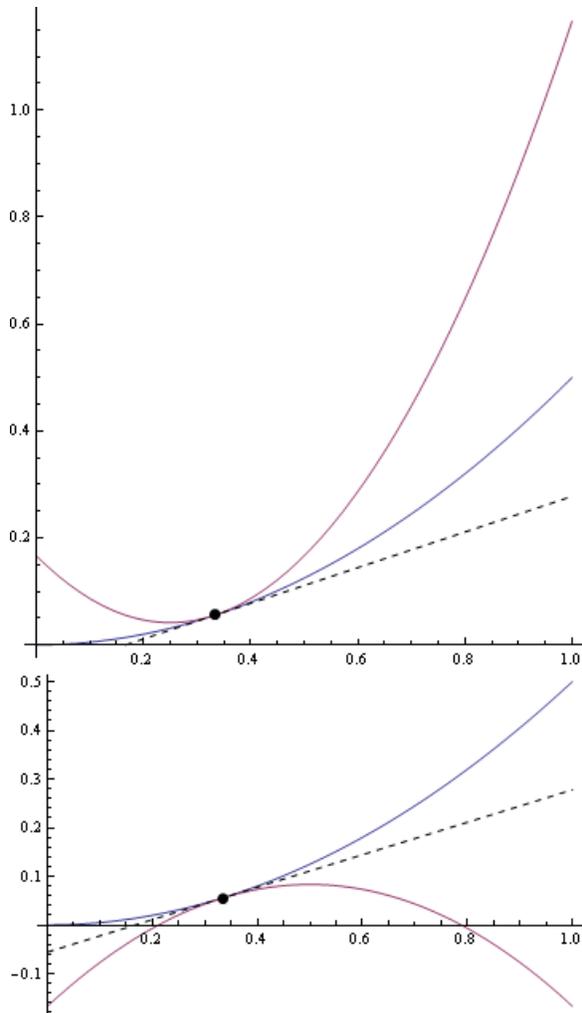
$$\tan \alpha = \left| \frac{f'(a) - g'(a)}{1 + f'(a)g'(a)} \right|$$

If  $f'(a)g'(a) = -1$ , then  $\alpha = \pi/2$ .

So let's calculate  $\tan \alpha$  at these two points. At the point  $(0, 0)$ , we have  $f'(0) = 1$ ,  $g'(0) = 0$ . Plugging into the formula, we obtain  $\tan \alpha = 1$ , so  $\alpha = \pi/4$ .

What about the point  $(1, 1)$ ? At this point,  $f'(1) = 1$  and  $g'(1) = 2$ . Thus, we obtain that  $\tan \alpha = 1/3$ . So,  $\alpha$  is the angle whose tangent is  $1/3$ . We don't know any particular angle  $\alpha$  satisfying this condition, but by the intermediate-value theorem, we can see that  $\alpha$  is somewhere between  $0$  and  $\pi/6$ . In the 153 course, we will look at inverse trigonometric functions, and when we do that, we will write  $\alpha$  as  $\arctan(1/3)$ . For now, you can just leave your answer as  $\tan \alpha = 1/3$ .

**3.5. Curves that are tangent and orthogonal at specific points.** Here are some pictures of tangent pairs of curves:

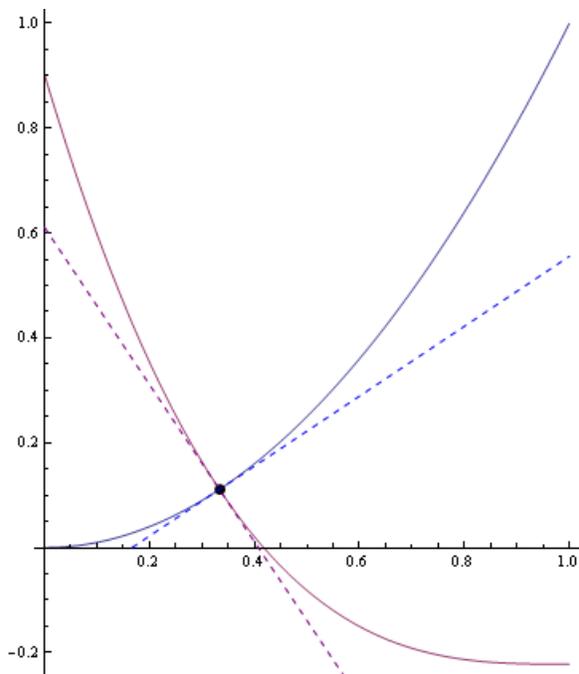


We say that two curves are *tangent* at a point if the angle of intersection between the curves at that point is 0. In other words, the curves share a tangent line. You can think of it as the two curves *touching* each other. [Draw pictures to explain]. Note that it could happen that they are kissing each other outward, as in this picture, or one is crossing the other, as in this picture.

If two curves are tangent at the point, that means that, *to the first order of approximation*, the curves behave very similarly close to that point. In other words, the best *linear* approximation to both curves close to that point is the same.

We say that two curves are *perpendicular* or *orthogonal* at a particular point of intersection if their tangent lines are perpendicular. If two curves are orthogonal at *every point of intersection*, we say that they are *orthogonal curves*. You'll see more on orthogonal curves in a homework problem.

Here is a graphical illustration of orthogonal curves:



**3.6. Geometric addendum: finding tangents and normals from other points.** Given the graph of a function  $y = f(x)$ , we can ask, given a point  $(x_0, y_0)$  in the plane (not necessarily on the graph) what tangents to the graph pass through this point. Similarly, we can ask what normals to the graph pass through this point.

Here's the general approach to these questions.

For the tangent line, we assume, say, that the tangent line is tangent at some point  $(x_1, y_1)$ . The equation of the tangent line is then:

$$y - y_1 = f'(x_1)(x - x_1)$$

Since  $y_1 = f(x_1)$ , we get:

$$y - f(x_1) = f'(x_1)(x - x_1)$$

Further, we know that the tangent line passes through  $(x_0, y_0)$ , so we plugging in, we get:

$$y_0 - f(x_1) = f'(x_1)(x_0 - x_1)$$

We now have an equation in the variable  $x_1$ . We solve for  $x_1$ . Note that when  $f$  is some specific function, both  $f(x_1)$  and  $f'(x_1)$  are expressions in  $x_1$  that we can simplify when solving the equation.

In some cases, we may not be able to solve the equation precisely, but can guarantee the existence of a solution and find a narrow interval containing this solution by using the intermediate value theorem.

A similar approach works for normal lines. We'll revisit this in the near future.

#### 4. LEIBNIZ NOTATION

**4.1. The  $d/dx$  notation.** Recall what we have done so far: we defined a notation of derivative, and then introduced a notation, the *prime notation*, for the derivative. So, if  $f$  is a function, the *derivative function* is denoted  $f'$ . And this is the function you obtain by differentiating  $f$ .

We now talk of a somewhat different notation that has its own advantages. And to understand the motivation behind this definition, we need to go back to how we think of derivative as the *rate of change*.

Suppose we have a function  $f$ . And, for simplicity, let's denote  $y = f(x)$ . So, what we want to do is study how, as the value of  $x$  changes, the value of  $f(x)$ , which is  $y$ , changes. So if for a point  $x_1$  we have  $f(x_1) = y_1$  and for a point  $x_2$  we have  $f(x_2) = y_2$ , we want to measure the difference between the  $y$ -values (which is

$y_2 - y_1$ ) and compare it with the difference between the  $x$ -values. And the quotient of the difference in the  $y$ -values and the difference in the  $x$ -values is what we called the *difference quotient*:

$$\text{Difference quotient} = \frac{y_2 - y_1}{x_2 - x_1}$$

And, this is also the slope of the line joining  $(x_1, y_1)$  and  $(x_2, y_2)$ .

Now, there's a slightly different way of writing this, which is typically used when the values of  $x$  are fairly close, and that is using the letter  $\Delta$ . And that says:

$$\text{Difference quotient} = \frac{\Delta y}{\Delta x}$$

Here  $\Delta y$  means "change in"  $y$  and  $\Delta x$  means "change in"  $x$ . So, the difference quotient is the ratio of the change in  $y$  to the change in  $x$ .

The *limiting* value of this quotient, as the  $x$ -values converge and the  $y$ -values converge, is called  $dy/dx$ . This is read "dee y dee x" or "dee y by dee x" and is also called the derivative of  $y$  with respect to  $x$ .

So, if  $y = f(x)$ , the function  $f'$  can also be written as  $dy/dx$ .

**4.2. Derivative as a function and derivative as a point.** The function  $f'$  can be evaluated at any point; so for instance, we write  $f'(3)$  to evaluate the derivative at 3. With the  $dy/dx$  notation, things are a little different. To evaluate  $dy/dx$  at a particular point, we write something like:

$$\left. \frac{dy}{dx} \right|_{x=3}$$

The bar followed by the  $x = 3$  means "evaluated at 3". In particular, it is *not* correct to write  $dy/d3$ . That doesn't make sense at all.

**4.3. Dependent and independent variable.** The Leibniz  $d/dx$  notation has a number of advantages over the prime notation. The first advantage is that instead of thinking in terms of functions, we now think in terms of two variables –  $x$  and  $y$ , and the relation between them. The fact that  $y$  can be expressed as a function of  $x$  becomes less relevant.

This is important because calculus is meant to address questions like: "When you change  $x$ , what happens to  $y$ ?" The explicit functional form of  $y$  in terms of  $x$  is only of secondary interest – what matters is that we have these two variables measuring the two quantities  $x$  and  $y$  and we want to determine how changes in the variable  $x$  influence changes in the variable  $y$ . We sometimes say that  $x$  is the *independent variable* and  $y$  is the *dependent variable* – because  $y$  depends on  $x$  via some (may be known, may be unknown) functional dependence.

**4.4.  $d/dx$  as an operator and using it.** The great thing about the  $d/dx$  notation is that you don't need to introduce separate letters to name your function. For instance, we can write:

$$\frac{d}{dx}(x^2 + 3x + 4)$$

No need to define  $f(x) := x^2 + 3x + 4$  and ask for  $f'(x)$  – we can directly write this stuff down.

This not only makes it easier to write down the first step, it also makes it easier to write and apply the rules for differentiation.

- (1) The sum rule becomes  $d(u + v)/dx = du/dx + dv/dx$ .
- (2) The product rule becomes  $d(uv)/dx = u(dv/dx) + v(du/dx)$ .
- (3) The scalar multiplication rule becomes  $d(\alpha u)/dx = \alpha(du/dx)$ .
- (4) The difference rule becomes  $d(u - v)/dx = du/dx - dv/dx$ .
- (5) The quotient rule becomes  $d((u/v))/dx = (v(du/dx) - u(dv/dx))/v^2$ .

The great thing about this notation is that we can write down partially calculated derivatives in intermediate steps, without naming new functions each time we break up the original function. For instance:

$$\frac{d}{dx} \left( \frac{x^3 + \sqrt{x} + 2}{x^2 + 1/(x+1)} \right)$$

We can write down the first step:

$$\left[ (x^2 + 1/(x+1)) \frac{d}{dx} (x^3 + \sqrt{x} + 2) - (x^3 + \sqrt{x} + 2) \frac{d}{dx} (x^2 + 1/(x+1)) \right] / (x^2 + 1/(x+1))^2$$

and then simplify the individual parts. If using the function notation, we would have to give names to both the numerator and denominator functions, but here we don't have to.

You should think of  $d/dx$  as an operator – the *differentiation operator* – that you can apply to expressions for functions.

## 5. HIGHER DERIVATIVES

**5.1. Higher derivatives and the multiple prime notation.** So far, we defined the derivative of a function at the point as the *rate of change* of the function at that point. The *second derivative* is the *rate of change of the rate of change*. In other words, the second derivative measures the rate at which the rate of change is changing. Or, it measures the rate at which the graph of the function is *turning* at the point, because a change in the derivative means a change in the direction of the graph.

So, this is just to remind you that higher derivatives are useful. Now, let's discuss the notation for higher derivatives.

In the prime notation, the second derivative of  $f$  is denoted  $f''$ . In other words,  $f''(x)$  is the derivative of the function  $f'$  evaluated at  $x$ . The third derivative is denoted  $f'''$ , the fourth derivative is denoted  $f''''$ . To simplify notation, we have the shorthand  $f^{(n)}$  for the  $n^{\text{th}}$  derivative, where by  $n^{\text{th}}$  derivative we mean the function obtained after applying the differentiation operator  $n$  times. So the first derivative  $f'$  can also be written as  $f^{(1)}$ , the second derivative  $f''$  can also be written as  $f^{(2)}$ , and so on. Typically, though, for derivatives up to the third derivative, we put the primes instead of the parenthesis ( $n$ ) notation.

Well, how do we compute these higher derivatives? Differentiate one step at a time. So, for instance, if  $f(x) = x^3 - 2x^2 + 3$ , then  $f'(x) = 3x^2 - 4x$ , so  $f''(x) = 6x - 4$ , and  $f'''(x) = 6$ . The fourth derivative of  $f$  is zero. And all *higher* derivatives are zero.

Similarly, if  $f(x) = 1/x$ , then  $f'(x) = -1/x^2$ ,  $f''(x) = 2/x^3$ ,  $f'''(x) = -6/x^4$ , and so on and so forth.

The first derivative measures the rate of change of the function, or the slope of the tangent line. The second derivative measures the rate at which the tangent line is turning, or the speed with which the graph is turning. The third derivative measures the rate at which this rate of turning itself is changing.

So, for the function  $f(x) = x^2$ , the first derivative is  $f'(x) = 2x$ , and the second derivative is 2. So, the first derivative is an increasing function, and the second derivative is constant, so the graph of the function is turning at a *constant rate*.

(The detailed discussion of the role of derivatives in terms of whether a function is increasing or decreasing will be carried out later – for now, you should focus on the computational aspects of derivatives).

**5.2. Higher derivatives in the Leibniz notation.** The Leibniz notation for higher derivatives is a bit awkward if you haven't seen it before.

Recall that the first derivative of  $y$  with respect to  $x$  is denoted  $dy/dx$ . Note that what I just called *first derivative* is what I have so far been calling *derivative* – when I just say derivative without an ordinal qualifier, I mean first derivative. So the first derivative is  $dy/dx$ . How would we write the second derivative?

Well, the way of thinking about it is that the first derivative is obtained by applying the  $d/dx$  operator to  $y$ , so the second derivative is obtained by applying the  $d/dx$  operator to  $dy/dx$ . So the second derivative is:

$$\frac{d}{dx} \left( \frac{dy}{dx} \right)$$

And that's perfectly correct, but it is long to write, so we can write this in shorthand as:

$$\frac{d^2 y}{(dx)^2}$$

Basically, we are (only notationally, not mathematically) multiplying the  $ds$  on top and multiplying the  $dx$ 's down below. There's a further simplification we do with the notation – we omit the parentheses in the denominator, to get:

$$\frac{d^2 y}{dx^2}$$

This is typically read out as “dee two y dee x two” though some people read it as “dee square y dee x square”.

And more generally, the  $n^{\text{th}}$  derivative is given by the shorthand:

$$\frac{d^n y}{dx^n}$$

Note that the  $dx^n$  in the denominator should be thought of as  $(dx)^n$  – *not* as  $d(x^n)$ , even though you omit parentheses.

This is typically read out as “dee n y dee x n”.

So for instance:

$$\frac{d^2}{dx^2} (x^3 + x + 1) = \frac{d}{dx} \left( \frac{d(x^3 + x + 1)}{dx} \right) = \frac{d}{dx} (3x^2 + 1) = 6x$$

## 6. CHAIN RULE

**6.1. The chain rule: statement and application to polynomials.** Now, the rules we have seen so far allow us to basically differentiate any rational function any number of times without ever using the limit definition – simply by applying the formulas.

Okay, how would you differentiate  $h(x) := (x^2 + 1)^5$ ? Well, in order to apply the formulas, you need to expand this out first, then use the termwise differentiation strategy. But taking the fifth power of  $x^2 + 1$  is a lot of work. So, we want a strategy to differentiate this without expanding.

This strategy is called the chain rule.

The chain rule states that if  $y$  is a function of  $v$  and  $v$  is a function of  $x$ , then:

$$\frac{dy}{dx} = \frac{dy}{dv} \frac{dv}{dx}$$

The intuition here may be that we can cancel the  $dv$ s – however, that’s not a rigorous reason since these are not really ratios but limits. But that’s definitely one way to remember the result.

In this case, we have  $y = (x^2 + 1)^5$ . What  $v$  can we choose? Well, let’s try  $v = x^2 + 1$ . Then  $y = v^5$ . So we have:

$$\frac{dy}{dx} = \frac{d(v^5)}{dv} \frac{d(x^2 + 1)}{dx}$$

Now, the first term on the right is  $5v^4$  and the second term on the right is  $2x$ , so the answer is:

$$\frac{dy}{dx} = (5v^4)(2x)$$

And  $v = x^2 + 1$ , so plugging that back in, we get:

$$\frac{dy}{dx} = 5(x^2 + 1)^4(2x) = 10x(x^2 + 1)^4$$

**6.2. Introspection: function composition.** What we really did was we had the function  $h(x) = (x^2 + 1)^5$ , and we decomposed  $h$  into two parts, the function  $g$  that sends  $x$  to  $x^2 + 1$ , and the function  $f$  that sends  $v$  to  $v^5$ . What we did was to write  $h = f \circ g$ , for two functions  $f$  and  $g$  that we can handle easily in terms of differentiation. Then, we used the chain rule to differentiate  $h$ , using what we know about differentiating  $f$  and  $g$ .

In functional notation, what the chain rule says equationally is that:

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

Going back to the  $y, v$  terminology, we have  $v = g(x)$ , and  $y = f(v)$ . And in that notation,  $dy/dv = f'(v) = f'(g(x))$ , while  $dv/dx = g'(x)$ . Which is precisely what we have here.

Notice that we apply  $f'$  not to  $x$  but to the value  $g(x)$ , which is what we called  $v$ , i.e., the value you get after applying one function but before applying the other one. But  $g'$  we apply to  $x$ .

Another way of writing this is:

$$\frac{d}{dx} [f(g(x))] = f'(g(x))g'(x)$$

This is a mix of the Leibniz notation and the prime notation.

**6.3. Precise statement of the chain rule.** What we said above is the correct equational expression for the chain rule, but let's just make the precise statement now with the assumptions.

If  $f, g$  are functions such that  $g$  is differentiable at  $x$  and  $f$  is differentiable at  $g(x)$ , then  $f \circ g$  is differentiable at  $x$  and:

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

Again, this statement is *at a point*, but if the differentiability assumptions hold globally, then the expression above holds globally as well, in which case we get:

$$(f \circ g)' = (f' \circ g) \cdot g'$$

The  $\cdot$  there denotes the pointwise product of functions.

## 7. ADDITIONAL FACTS AND SUBTLETIES

Most of these are things you will discover with experience as you do homework problems, but they're mentioned here just for handy reference. This is something you might want to read more carefully when you review these notes at a later stage.

**7.1. One-sided versions.** The situation with one-sided versions of the results for derivatives is very similar to that with limits and continuity. For all pointwise combination results, one-sided versions hold. Conceptually, each result for derivatives depends (in its proof) on the corresponding result for limits. Since the result on limits has a one-sided version, so does the corresponding result on derivatives. In words:

- (1) The left-hand derivative of the sum is the sum of the left-hand derivatives.
- (2) The right-hand derivative of the sum is the sum of the right-hand derivatives.
- (3) The left-hand derivative of a scalar multiple is the same scalar multiple of the left-hand derivative.
- (4) The right-hand derivative of a scalar multiple is the same scalar multiple of the right-hand derivative.
- (5) The analogue of the product rule for the left-hand derivative can be obtained if we replace the derivative in all three places in the product rule with left-hand derivative. Similarly for right-hand derivative.
- (6) Similar to the above for quotient rule.

But – *you saw it coming* – the naive one-sided analogue of the rule for composites fails, for the same reason as the one-sided analogue of the composition results for limits and continuity fail. We need the additional condition that the direction of approach of the intermediate expression is the same as that of the original domain variable.

**7.2. The derivative as a function: is it continuous, differentiable?** Suppose  $f$  is a function. For simplicity, we'll assume the domain of  $f$  to be a (possibly infinite) open interval  $I$  in  $\mathbb{R}$ . We're taking an open interval to avoid one-sided issues at boundary points. We say that  $f$  is *differentiable* on its domain if  $f'$  exists everywhere on  $I$ . If  $f$  is differentiable, what can we say about the properties of  $f'$ ?

Your first instinct may be to say that if  $f'$  is defined on an open interval, then it should be continuous on that interval. Indeed, in all the simple examples one can think of, the existence of the derivative *on an open interval* implies continuity of the derivative. However, this is not true as a general principle. Some points of note:

- (1) It is possible for the derivative to not be a continuous function. An example is the function  $g(x) := \begin{cases} x^2 \sin(1/x), & x \neq 0 \\ 0, & x = 0 \end{cases}$ . This function is differentiable everywhere, but the derivative at 0 is *not* the limit of the derivative near zero.

- (2) However, the derivative of a continuous function, if defined everywhere on an open interval, satisfies the intermediate value property. This is a fairly hard and not very intuitive theorem called Darboux's theorem, and you might see it if you take up the 203-204-205 analysis sequence. In this respect, it behaves in a manner very similar to a continuous function. In particular, any discontinuities of the derivative must be of the oscillatory kind on both the left and right side. In particular, *if* a one-sided limit *exists* for the derivative, it equals the value of the derivative.

The proof of (2) is well beyond the scope of this course. You don't even need to know the precise statement, and I'm including it here just in order to place the examples you've seen in context.

A fun discussion of the fact that derivatives need not be continuous but satisfy the intermediate value property and the implications of this fact can be found here:

<http://www.thebigquestions.com/2010/09/16/speed-math/>

**7.3. Higher differentiability.** We say that a function  $f$  on an open interval  $I$  is  $k$  times differentiable on  $I$  if the  $k^{\text{th}}$  derivative of  $f$  exists at all points of  $I$ . We say that  $f$  is  $k$  times continuously differentiable on  $I$  if the  $k^{\text{th}}$  derivative of  $f$  exists *and* is continuous on  $I$ .

Recall that differentiable implies continuous. Thus, if a function is twice differentiable, i.e., the first derivative is differentiable function, this implies that the first derivative is a continuous function. We thus see a chain of implications:

Continuous  $\Leftarrow$  Differentiable  $\Leftarrow$  Continuously differentiable  $\Leftarrow$  Twice differentiable  $\Leftarrow$  Twice continuously differentiable  $\Leftarrow$  . . .

In general:

$k$  times differentiable  $\Leftarrow$   $k$  times continuously differentiable  $\Leftarrow$   $k + 1$  times differentiable

We say that a function  $f$  is *infinitely differentiable* if it is  $k$  times differentiable for all  $k$ . The above implications show us that this is equivalent to saying that  $f$  is  $k$  times continuously differentiable for all  $k$ .

A  $k$  times continuously differentiable function is sometimes also called a  $C^k$ -function. (When  $k = 0$ , we get continuous functions, so continuous functions are sometimes called  $C^0$ -functions). An infinitely differentiable function is sometimes also called a  $C^\infty$ -function. We will not use the term in this course, though we will revisit it in 153 when studying power series, and you'll probably see it if you do more mathematics courses.

All these containments are strict. Examples along the lines of the  $x^n \sin(1/x)$  constructions can be used to show this.

**7.4. Carrying out higher differentiation.** Suppose functions  $f$  and  $g$  are both  $k$  times differentiable. Are there rules to find the  $k^{\text{th}}$  derivatives of  $f + g$ ,  $f - g$ ,  $f \cdot g$ , etc. directly in terms of the  $k^{\text{th}}$  derivatives of  $f$  and  $g$  respectively? For the sum, difference, and scalar multiples, the rules are simple:

$$\begin{aligned}(f + g)^{(k)} &= f^{(k)} + g^{(k)} \\ (f - g)^{(k)} &= f^{(k)} - g^{(k)} \\ (\alpha f)^{(k)} &= \alpha f^{(k)}\end{aligned}$$

Later on, we'll see that this bunch of rules can be expressed more briefly by saying that the operation of differentiating  $k$  times is a linear operator.

For products, the rule is more complicated. In fact, the general rule is somewhat like the binomial theorem. The situation with composites is also tricky. We will revisit both products and composites a little later. For now, all we care about are existence facts:

- If  $f$  and  $g$  are  $k$  times differentiable on an open  $I$ , so are  $f + g$ ,  $f - g$ , and  $f \cdot g$ . If  $g$  is not zero anywhere on  $I$ , then  $f/g$  is also  $k$  times differentiable on  $I$ .
- Ditto to the above, replacing " $k$  times differentiable" by " $k$  times continuously differentiable."
- If  $f$  and  $g$  are functions such that  $g$  is  $k$  times differentiable on an open interval  $I$  and  $f$  is  $k$  times differentiable on an open interval  $J$  containing the range of  $g$ , then  $f \circ g$  is  $k$  times differentiable on  $I$ .
- Ditto to the above, replacing " $k$  times differentiable" by " $k$  times continuously differentiable".

**7.5. Families of functions closed under differentiation.** Suppose  $\mathcal{F}$  is a collection of functions that is closed under addition, subtraction and scalar multiplication. We say in this case that  $\mathcal{F}$  is a *vector space* of functions. If, in addition,  $\mathcal{F}$  contains constant functions and is closed under multiplication, we say that  $\mathcal{F}$  is an *algebra* of functions. (You aren't responsible for learning this terminology, but it really helps make clear what we're going to be talking about shortly).

The vector space generated by a bunch of functions  $\mathcal{B}$  is basically the set of all functions we can get starting from  $\mathcal{B}$  by the processes of addition and scalar multiplication. If  $\mathcal{B}$  generates a vector space  $\mathcal{F}$  of functions, then we say that  $\mathcal{B}$  is a generating set for  $\mathcal{F}$ .

For instance, if we consider all the functions  $1, x, x^2, \dots, x^n, \dots$ , these generate the vector space of all polynomials: we can get to all polynomials by the processes of addition and scalar multiplication starting with these functions.

The algebra generated by a bunch  $\mathcal{B}$  of functions (which we assume includes constant functions) is the collection of functions  $\mathcal{A}$  that we obtain by starting with the functions in  $\mathcal{B}$  and the processes of addition, subtraction, multiplication, and scalar multiplication.

For instance, the algebra generated by the identity function (the function *xmapsto*) is the algebra of all polynomial functions.

The point of all this is as follows:

- (1) Suppose  $\mathcal{B}$  is a bunch of functions and  $\mathcal{F}$  is the vector space generated by  $\mathcal{B}$ . Then, if every function in  $\mathcal{B}$  is differentiable and the derivative of the function is in  $\mathcal{F}$ , then every function in  $\mathcal{F}$  is differentiable and has derivative in  $\mathcal{F}$ . As a corollary, every function in  $\mathcal{F}$  is infinitely differentiable and all its derivatives lie in  $\mathcal{F}$ .
- (2) Suppose  $\mathcal{B}$  is a bunch of functions and  $\mathcal{A}$  is the algebra generated by  $\mathcal{B}$ . Then, if every function in  $\mathcal{B}$  is differentiable and the derivative of the function is in  $\mathcal{A}$ , then every function in  $\mathcal{A}$  is differentiable and the derivative of the function is in  $\mathcal{A}$ . As a corollary, every function in  $\mathcal{A}$  is infinitely differentiable and all its derivatives lie in  $\mathcal{A}$ .

Let's illustrate point (2) (which is in some sense the more powerful statement) with the example of polynomial functions. The single function  $f(x) := x$  generates the algebra of all polynomial functions. The derivative of  $f$  is the function 1, which is also in the algebra of all polynomial functions. What point (2) is saying is that just this simple fact allows us to see that the derivative of *any* polynomial function is a polynomial function, and that polynomial functions are infinitely differentiable.

We'll see a similar trigonometric example in the near future. We'll also explore this way of thinking more as the occasion arises.

Another way of thinking of this is that each time we obtain a formula to differentiate a bunch of functions, we have a technique to differentiate *all* functions in the algebra generated by that bunch of functions. While this may seem unremarkable, the analogous statement is *not true at all* for other kinds of operators such as indefinite integration.