

1 Derivatives

Before defining the derivative of a function, let's begin with two motivating examples.

Example: Driving Imagine motoring along down highway 61 leaving Minnesota on the way to New Orleans; though lost in listening to music, still mindful of the speedometer and odometer, both prominently placed on the dashboard of the car.

The speedometer reads 60 miles per hour, what is the odometer doing? Besides recording total distance traveled, it is incrementing dutifully every hour by 60 miles. Why? Well, the well-known formula relating distance, time and rate of travel is

$$\text{distance} = \text{rate} \times \text{time}.$$

If the rate is a constant 60 miles/hour, then in one hour the distance traveled is 60 miles.

Of course, the odometer isn't just incrementing once per hour, it is incrementing once every 1/10th of a mile. How much time does that take? Well, we would need to solve $1/10 = 60 \cdot t$ which means $t = 1/600$ hours, better known as once every 6 seconds.

Using some mathematical notation, would give $x(t) = v \cdot t$, where x is position at time t , v is the velocity and t the time traveled in hours. A simple graph of the first three hours of travel would show:

```
| using CalculusWithJulia
| using Plots
| position(t) = 60 * t
| plot(position, 0, 3)
```

```
| Plot{Plots.PlotlyBackend() n=1}
```

```
| plotly()
| plot(sin, 0, 2pi)
```

```
| Plot{Plots.PlotlyBackend() n=1}
```

Oh no, we hit traffic. In the next 30 minutes we only traveled 15 miles. We were so busy looking out for traffic, the speedometer was not checked. What would the average speed have been? Though in the 30 minutes, the displayed speed may have varied, the *average speed* would simply be the change in distance over the change in time, or $\Delta x / \Delta t$. That is

```
| 15/(1/2)
```

```
| 30.0
```

Now suppose that after 6 hours of travel the GPS in the car gives us a readout of distance traveled as a function of time. The graph looks like this:

```
|Plot{Plots.PlotlyBackend() n=1}
```

We can see with some effort that the slope is steady for the first three hours, is slightly less between 3 and 3.5 hours, then is a bit steeper for the next half hour. After that, it is flat for the about half an hour, then the slope continues on with same value as in the first 3 hours. What does that say about our speed during our trip?

Based on the graph, what was the average speed over the first three hours? Well, we traveled 180 miles, and took 3 hours:

```
|180/3
```

```
|60.0
```

What about the next half hour? Squinting shows the amount traveled was 15 miles (195 - 180) and it took 1/2 an hour:

```
|15/(1/2)
```

```
|30.0
```

And the half hour after that? The average speed is found from the distance traveled, 37.5 miles, divided by the time, 1/2 hour:

```
|37.5 / (1/2)
```

```
|75.0
```

Okay, so there was some speeding involved.

The next half hour the car did not move. What was the average speed? Well the change in position was 0, but the time was 1/2 hour, so the average was 0.

Perhaps a graph of the speed is a bit more clear. We can do this based on the above:

```
|function speed(t)
|    if      0    < t <= 3
|        60
|    elseif 3    < t <= 3.5
|        30
|    elseif 3.5  < t <= 4
|        75
|    elseif 4    < t <= 4.5
|        0
```

```

else
    60
end
end
plot(speed, 0, 6)

```

```
Plot{Plots.PlotlyBackend() n=1}
```

The jumps, as discussed before, are artifacts of the graphing algorithm. What is interesting, is we could have derived the graph of `speed` from that of `x` by just finding the slopes of the line segments, and we could have derived the graph of `x` from that of `speed`, just using the simple formula relating distance, rate, and time.

We were pretty loose with some key terms. There is a distinction between "speed" and "velocity", this being the speed is the absolute value of velocity. Velocity incorporates a direction as well as a magnitude. Similarly, distance traveled and change in position are not the same thing when there is back tracking involved. The total distance traveled is computed with the speed, the change in position is computed with the velocity. When there is no change of sign, it is a bit more natural, perhaps, to use the language of speed and distance.

Example: Galileo's ball and ramp experiment One of history's most famous experiments was performed by [Galileo](#) where he rolled balls down inclined ramps, making note of distance traveled with respect to time. As Galileo had no ultra-accurate measuring device, he needed to slow movement down by controlling the angle of the ramp. With this, he could measure units of distance per units of time. (Click through to *Galileo and Perspective Dauben*.)

Suppose that no matter what the incline was, Galileo observed that in units of the distance traveled in the first second that the distance traveled between subsequent seconds was 3 times, then 5 times, then 7 times, ... This table summarizes.

```

Error: MethodError: no method matching parse(::Int64)
Closest candidates are:
  parse(!Matched::IO; flavor) at /Users/julia/buildbot/worker/package_macos64/build/usr/share/julia/stdlib/v1.5/Markdown/src/parse/parse.jl:91
  parse(!Matched::IO, !Matched::Markdown.MD; breaking) at /Users/julia/buildbot/worker/package_macos64/build/usr/share/julia/stdlib/v1.5/Markdown/src/parse/parse.jl:88
  parse(!Matched::IO, !Matched::Markdown.MD, !Matched::Markdown.Config; breaking) at /Users/julia/buildbot/worker/package_macos64/build/usr/share/julia/stdlib/v1.5/Markdown/src/parse/parse.jl:79
...

```

A graph of distance versus time could be found by interpolating between the measured points:

```

ts = [0,1,2,3,4, 5]
xs = [0,1,4,9,16,25]
plot(ts, xs)

```

```
|Plot{Plots.PlotlyBackend() n=1}
```

The graph looks almost quadratic. What would the following questions have yielded?

- What is the average speed between 0 and 3?

```
| (9-0) / (3-0) # (xs[4] - xs[1]) / (ts[4] - ts[1])
```

```
| 3.0
```

- What is the average speed between 2 and 3?

```
| (9-4) / (3-2) # (xs[4] - xs[3]) / (ts[4] - ts[3])
```

```
| 5.0
```

From the graph, we can tell that the slope of the line connecting (2, 4) and (3, 9) will be greater than that connecting (0, 0) and (3, 9). In fact, given the shape of the graph (concave up), the line connecting (0, 0) with any point will have a slope less than or equal to any of the line segments.

The average speed between k and $k + 1$ for this graph is:

```
| xs[2]-xs[1], xs[3] - xs[2], xs[4] - xs[3], xs[5] - xs[4]
```

```
| (1, 3, 5, 7)
```

We see it increments by 2. The acceleration is the rate of change of speed. We see the rate of change of speed is constant, as the speed increments by 2 each time unit.

Based on this - and given Galileo's insight - it appears the acceleration will be a constant and the position as a function of time will be quadratic.

1.1 The slope of the secant line

In the above examples, we see that the average speed is computed using the slope formula. This can be generalized for any univariate function $f(x)$:

The average rate of change between a and b is $(f(b) - f(a))/(b - a)$. It is typical to express this as $\Delta y / \Delta x$, where Δ means "change".

Geometrically, this is the slope of the line connecting the points $(a, f(a))$ and $(b, f(b))$. This line is called a [secant](#) line, which is just a line intersecting two specified points on a curve.

Rather than parameterize this problem using a and b , we let c and $c + h$ represent the two values for x , then the secant-line-slope formula becomes

$$m = \frac{f(c + h) - f(c)}{h}.$$

1.2 The slope of the tangent line

The slope of the secant line represents the average rate of change over a given period, h . What if this rate is so variable, that it makes sense to take smaller and smaller periods h ? In fact, what if h goes to 0?

XXX can not include ‘.gif’ file here

The graphic suggests that the slopes of the secant line converge to the slope of a “tangent” line. That is, for a given c , this limit exists:

$$\lim_{h \rightarrow 0} \frac{f(c + h) - f(c)}{h}.$$

We’ll define the tangent line at $(c, f(c))$ to be the line through the point with the slope from the limit above - provided that limit exists. Informally, the tangent line is the line through the point that best approximates the function.

XXX can not include ‘.gif’ file here

The tangent line is not just a line that intersects the graph in one point, nor does it need only intersect the line in just one point.

This last point was certainly not obvious at first. [Barrow](#), who had Newton as a pupil and was the first to sketch a proof of part of the Fundamental Theorem of Calculus, understood a tangent line to be a line that intersects a curve at only one point.

Example What is the slope of the tangent line to $f(x) = \sin(x)$ at $c = 0$?

We need to compute the limit $(\sin(c + h) - \sin(c))/h$ which is the limit as h goes to 0 of $\sin(h)/h$. We know this to be 1.

```
f(x) = sin(x)
c = 0
tl(x) = f(c) + 1 * (x - c)
plot(f, -pi/2, pi/2)
plot!(tl, -pi/2, pi/2)
```

```
|Plot{Plots.PlotlyBackend() n=2}
```

1.3 The derivative

The limit of the slope of the secant line gives an operation: for each c in the domain of f there is a number (the slope of the tangent line) or it does not exist. That is, there is derived function from f . Call this function the *derivative* of f . There are many notations for this, here we use the "prime" notation:

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

The limit above is identical, only it uses x instead of c to emphasize that we are thinking of function now, and not just a value at a point.

1.3.1 Some basic derivatives

- **The power rule.** What is the derivative of the monomial $f(x) = x^n$? We need to look at $(x+h)^n - x^n$ for positive, integer-value n . Let's look at a case, $n = 5$

```
@vars x h real=true
n = 5
ex = expand((x+h)^n - x^n)
```

$$h^5 + 5h^4x + 10h^3x^2 + 10h^2x^3 + 5hx^4$$

All terms have an h in them, so we cancel it out:

```
|cancel(ex/h, h)
```

$$h^4 + 5h^3x + 10h^2x^2 + 10hx^3 + 5x^4$$

We see the lone term $5x^4$ without an h , so as we let h go to 0, this will be the limit. That is, $f'(x) = 5x^4$.

For general integer-value, positive n , the binomial theorem gives an expansion $(x+h)^n = x^n + nx^{n-1} \cdot h^1 + n \cdot (n-1)x^{n-2} \cdot h^2 + \dots$. Subtracting x^n then dividing by h leaves just the term nx^{n-1} without a power of h , so the limit, in general, is just this term. That is $[x^n]' = nx^{n-1}$.

It isn't a special case, but when $n = 0$, we also have the above formula applies, as x^0 is the constant 1, and all constant functions will have a derivative of 0 at all x . We will see that in general, the power rule applies for any n where x^n is defined.

- What is the derivative of $f(x) = \sin(x)$? We know that $f'(0) = 1$ by an earlier example, here we solve in general.

We need to consider the difference $\sin(x+h) - \sin(x)$:

```
| ex = sympy.expand_trig(sin(x+h) - sin(x)) # expand_trig is not exposed in `SymPy`
```

$$\sin(h) \cos(x) + \sin(x) \cos(h) - \sin(x)$$

That used the formula $\sin(x+h) = \sin(x)\cos(h) + \sin(h)\cos(x)$.

We could then rearrange the secant line slope formula to become:

$$\cos(x) \cdot \frac{\sin(h)}{h} + \sin(x) \cdot \frac{\cos(h) - 1}{h}$$

and take a limit. If the answer isn't clear, we can let **SymPy** do this work:

```
| limit((sin(x+h) - sin(x))/h, h=>0)
```

$$\cos(x)$$

- Let's see what the derivative of $\log(x)$ is. We have

$$\frac{\log(x+h) - \log(x)}{h} = \frac{1}{h} \log\left(\frac{x+h}{x}\right) = \log((1+h/x)^{1/h}).$$

Earlier we saw that the limit as u goes to 0 of $f(u) = (1+u)^{1/u}$ is e . Re-expressing the above we can get $1/x \cdot \log(f(h/x))$. The limit as h goes to 0 of this is found from the composition rules for limits: as $\lim_{h \rightarrow 0} f(h/x) = e$, and since $\log(e)$ is 1 we get this expression has a limit of $1/x$.

We verify through:

```
| limit((log(x+h) - log(x))/h, h=>0)
```

$$\frac{1}{x}$$

There are several different **notations** for derivatives. Some are historical, some just add flexibility. We use the prime notation of Lagrange: $f'(x)$, u' and $[\text{expr}]'$, where the first emphasizes that the derivative is a function with a value at x , the second emphasizes the derivative operates on functions, the last emphasize that we are taking the derivative of some expression (the "rule" part of an unnamed function).

There are many other notations:

- The Leibniz notation uses the infinitesimals: dy/dx to relate to $\Delta y/\Delta x$. This notation is very common, and especially useful when more than one variable is involved. **SymPy** uses Leibniz notation in some of its output, expressing somethings such as:

$$f'(x) = \frac{d}{d\xi}(f(\xi))\Big|_{\xi=x}.$$

The notation - | - on the right-hand side separates the tasks of finding the derivative and evaluating the derivative at a specific value.

- Euler used D for the operator $D(f)$. This notation is appropriated by an operator in the `CalculusWithJulia` package. This was initially used by [Argobast](#).
- Newton used a "dot" $\dot{x}(t)$, which is still widely used in physics to indicate a derivative in time.

1.4 Rules of derivatives

We could proceed in a similar manner to find other derivatives, but let's not. If we have a function $f(x) = x^5 \sin(x)$, it would be nice to leverage our previous work on the derivatives of $f(x) = x^5$ and $g(x) = \sin(x)$, rather than derive an answer from scratch.

As with limits and continuity, it proves very useful to consider rules that make the process of finding derivatives of combinations of functions a matter of combining derivatives of the individual functions in some manner.

1.4.1 Sum rule

Let's consider $k(x) = a \cdot f(x) + b \cdot g(x)$, what is its derivative? That is, in terms of f , g and their derivatives, can we express $k'(x)$?

We can rearrange $(k(x+h) - k(x))$ as follows:

$$(a \cdot f(x+h) + b \cdot g(x+h)) - (a \cdot f(x) + b \cdot g(x)) = a \cdot (f(x+h) - f(x)) + b \cdot (g(x+h) - g(x)).$$

Dividing by h , we see that this becomes

$$a \cdot \frac{f(x+h) - f(x)}{h} + b \cdot \frac{g(x+h) - g(x)}{h} \rightarrow a \cdot f'(x) + b \cdot g'(x).$$

This holds two rules: the derivative of a constant times a function is the constant times the derivative of the function; and the derivative of a sum of functions is the sum of the derivative of the functions.

1.4.2 Product rule

Other rules can be similarly derived. `SymPy` can give us them as well. Here we define to symbolic functions `u` and `v` and let `SymPy` derive a formula for the derivative of a product of functions:

```
u,v = SymFunction("u,v") # make symbolic functions
f(x) = u(x) * v(x)
limit((f(x+h) - f(x))/h, h=>0)
```


$$u(x) \left. \frac{d}{d\xi_1} v(\xi_1) \right|_{\xi_1=x} + v(x) \left. \frac{d}{d\xi_1} u(\xi_1) \right|_{\xi_1=x}$$

The output uses some new notation to represent that the derivative of $u(x) \cdot v(x)$ is the u times the derivative of v plus v times the derivative of u . A common shorthand is $[uv]' = u'v + uv'$.

1.4.3 Quotient rule

The derivative of $f(x) = u(x)/v(x)$ - a ratio of functions - can be similarly computed. The result will be $[u/v]' = (u'v - uv')/v^2$.

1.4.4 Chain rule

Finally, the derivative of a composition of functions can be computed. This gives a rule called the *chain rule*. Before deriving, let's give a slight motivation.

Consider the output of a factory for some widget. It depends on two steps: an initial manufacturing step and a finishing step. The number of employees is important in how much is initially manufactured. Suppose x is the number of employees and $g(x)$ is the amount initially manufactured. Adding more employees increases the amount made by the made-up rule $g(x) = \sqrt{x}$. The finishing step depends on how much is made by the employees. If y is the amount made, then $f(y)$ is the number of widgets finished. Suppose for some reason that $f(y) = y^2$.

How many widgets are made as a function of employees? The composition $u(x) = f(g(x))$ would provide that.

What is the effect of adding employees on the rate of output of widgets? In this specific case we know the answer, as $(f \circ g)(x) = x$, so the answer is just the rate is 1.

In general, we want to express $\Delta f / \Delta x$ in a form so that we can take a limit.

But what do we know? We know $\Delta g / \Delta x$ and $\Delta f / \Delta y$. Using $y = g(x)$, this suggests that we might have luck with the right side of this equation:

$$\frac{\Delta f}{\Delta x} = \frac{\Delta f}{\Delta y} \cdot \frac{\Delta y}{\Delta x}.$$

Interpreting this, we get the *average* rate of change in the composition can be thought of as a product: The *average* rate of change of the initial step ($\Delta y / \Delta x$) times the *average* rate of the change of the second step evaluated not at x , but at y , $\Delta f / \Delta y$.

Re-expressing using derivative notation with h would be:

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

The left hand side will converge to the derivative of $u(x)$ or $[f(g(x))]'$.

The right most part of the right side would have a limit $g'(x)$, were we to let h go to 0.

It isn't obvious, but the left part of the right side has the limit $f'(g(x))$. This would be clear if *only* $g(x+h) = g(x) + h$, for then the expression would be exactly the limit expression with

$c = g(x)$. But, alas, except to some hopeful students and some special cases, it is definitely not the case in general that $g(x + h) = g(x) + h$ - that right parentheses actually means something. However, it is *nearly* the case that $g(x + h) = g(x) + kh$ for some k and this can be used to formulate a proof (one of the two detailed [here](#) and [here](#)).

We can verify this using SymPy:

```
|limit(( u(v(x+h)) - u(v(x)) ) / (v(x+h) - v(x)), h=>0)
```

$$\frac{d}{dv(x)}u(v(x))$$

Combined, we would end up with:

The chain rule: $[f(g(x))]' = f'(g(x)) \cdot g'(x)$. That is the derivative of the outer function evaluated at the inner function times the derivative of the inner function.

To see that this works in our specific case, we assume the general power rule that $[x^n]' = nx^{n-1}$ to get: $g'(x) = (1/2)x^{-1/2}$, $f'(x) = 2x$, and $f'(g(x)) = 2(\sqrt{x})$. Together, the product is:

$$2\sqrt{x} \cdot (1/2)1/\sqrt{x} = 1$$

Which is the same as the derivative of x found by first evaluating the composition.

Proof of the Chain Rule A function is *differentiable* at a if the following limit exists $\lim_{h \rightarrow 0} (f(a+h) - f(a))/h$. Reexpressing this as: $f(a+h) - f(a) - f'(a)h = \epsilon_f(h)h$ where as $h \rightarrow 0$, $\epsilon_f(h) \rightarrow 0$. Then, we have:

$$g(a+h) = g(a) + g'(a)h + \epsilon_g(h)h = g(a) + h',$$

Where $h' = (g'(a) + \epsilon_g(h))h \rightarrow 0$ as $h \rightarrow 0$ will be used to simplify the following:

$$f(g(a+h)) - f(g(a)) = f(g(a) + g'(a)h + \epsilon_g(h)h) - f(g(a)) \quad (1)$$

$$= f(g(a)) + f'(g(a))(g'(a)h + \epsilon_g(h)h) + \epsilon_f(h')(h') - f(g(a)) \quad (2)$$

$$= f'(g(a))g'(a)h + f'(g(a))(\epsilon_g(h)h) + \epsilon_f(h')(h'). \quad (3)$$

Rearranging:

$$f(g(a+h)) - f(g(a)) - f'(g(a))g'(a)h = f'(g(a))\epsilon_g(h)h + \epsilon_f(h')(h') = (f'(g(a))\epsilon_g(h) + \epsilon_f(h')((g'(a) + \epsilon_g(h))h))$$

where $\epsilon(h)$ combines the above terms which go to zero as $h \rightarrow 0$ into one. This is the alternative definition of the derivative, showing $(f \circ g)'(a) = f'(g(a))g'(a)$ when g is differentiable at a and f is differentiable at $g(a)$.

More examples

- Find the derivative of $x^5 \cdot \sin(x)$.

This is a product of functions, using $[u \cdot v]' = u'v + uv'$ we get:

$$5x^4 \cdot \sin(x) + x^5 \cdot \cos(x)$$

- Find the derivative of $x^5 / \sin(x)$.

This is a quotient of functions. Using $[u/v]' = (u'v - uv')/v^2$ we get

$$(5x^4 \cdot \sin(x) - x^5 \cdot \cos(x)) / (\sin(x))^2.$$

- Find the derivative of $\sin(x^5)$. This is a composition of functions $u(v(x))$ with $v(x) = x^5$. The chain rule says find the derivative of u ($\cos(x)$) and evaluate at $v(x)$ ($\cos(x^5)$) then multiply by the derivative of v :

$$\cos(x^5) \cdot 5x^4.$$

- Similarly, but differently, find the derivative of $\sin(x)^5$. Now $v(x) = \sin(x)$, so the derivative of $u(x)$ ($5x^4$) evaluated at $v(x)$ is $5(\sin(x))^4$ so multiplying by v' gives:

$$5(\sin(x))^4 \cdot \cos(x)$$

We can verify these with **SymPy**. Rather than take a limit, we will use **SymPy**'s **diff** function to compute derivatives.

```
| diff(x^5 * sin(x))
```

$$x^5 \cos(x) + 5x^4 \sin(x)$$

```
| diff(x^5/sin(x))
```

$$-\frac{x^5 \cos(x)}{\sin^2(x)} + \frac{5x^4}{\sin(x)}$$

```
| diff(sin(x^5))
```

$$5x^4 \cos(x^5)$$

and finally,

| `diff(sin(x)^5)`

$$5 \sin^4(x) \cos(x)$$

The `diff` function can be called as `diff(ex)` when there is just one free variable, as in the above examples; as `diff(ex, var)` when there are parameters in the expression; or, as `diff(f)`, where `f` is the name of a univariate function.

- Let's see that the derivative of e^x is just itself.

$$\frac{e^{x+h} - e^x}{h} = \frac{e^x \cdot (e^h - 1)}{h}.$$

If we know that $\lim_{h \rightarrow 0} (e^h - 1)/h = 1$, we get $[e^x]' = e^x$, that is it is a function satisfying $f' = f$.

- Suppose we knew that $\log(x)$ had derivative of $1/x$, but didn't know the derivative of e^x . From their inverse relation, we have: $x = \log(e^x)$, so taking derivatives of both sides would yield:

$$1 = \left(\frac{1}{e^x}\right) \cdot [e^x]'$$

Or solving, $[e^x]' = e^x$. This is a general strategy to find the derivative of an *inverse* function.

- The general product rule: For any n - not just integer values - we can re-express x^n using e : $x^n = e^{n \log(x)}$. Now the chain rule can be applied:

$$[x^n]' = [e^{n \log(x)}]' = e^{n \log(x)} \cdot \left(n \frac{1}{x}\right) = nx^n \cdot \frac{1}{x} = nx^{n-1}.$$

Examples

- Find the derivative of $f(x) = x^3(1-x)^2$ using either the power rule or the sum rule.

The power rule expresses $f = u \cdot v$. With $u(x) = x^3$ and $v(x) = (1-x)^2$ we get:

$$u'(x) = 3x^2, \quad v'(x) = 2 \cdot (1-x)^1 \cdot (-1),$$

the last by the chain rule. Combining with $u'v + uv'$ we get: $f'(x) = (3x^2) \cdot (1-x)^2 + x^3 \cdot (-2) \cdot (1-x)$.

Otherwise, the polynomial can be expanded to give $f(x) = x^5 - 2x^4 + x^3$ which has derivative $f'(x) = 5x^4 - 8x^3 + 3x^2$.

- Find the derivative of $f(x) = (x-1)(x-2)(x-3)(x-4)(x-5)$.