Calculus Unraveled

Intuition, Proofs, and Python

Dr. Mike X Cohen

0.1 Front matter

This page contains some important details about the book that basically no one reads but somehow is always in the first page.

ISBN: 9798309397228, edition 1.

This book was written and formatted in LaTeX by Mike X Cohen.

0.2 Dedication

If you're reading this, then the book is dedicated to you. I wrote this book for *you*. Now turn the page and start learning calculus!

0.3 Forward

The past is immutable and the present is fleeting. Forward is the only direction.

Contents

Chapter 5

Fundamentals of differentiation

This might sound strange or difficult to believe, but there isn't anything new to learn in the rest of this book. Functions and limits are the basis for everything in calculus. For the rest of this book, everything "new" that you will learn is just different or unintuitive ways of combining limits and algebra.

In this chapter, you will learn how to differentiate individual functions, and then in the next few chapters, you'll learn the rules for differentiating combinations of functions, applications of differentiation, then on to integration.

One small point about grammar: You will come across the terms *derivatives* and *differentiation*. Derivatives are the slopes of a function at each value of x. A derivative is a noun. Differentiation is the process of obtaining the derivative. It's an algorithm you apply. A verb. So, the goal of differential calculus is to obtain the derivative of a function, and the way you get that derivative is through a technique called differentiation.

5.1 Slope of a line

You may be wondering why there is a section on slopes in a chapter on derivatives. Here's why: The derivative is the slope of a line. More specifically, the slope of a tiny line segment. If you are already comfortable with the math and interpretation of a slope, then that's great! You have a head start on learning about derivatives and how to calculate them. By the end of this section, you will have a deep intuition for how slopes of lines lead to derivatives, and then in the next section you'll learn the formal definition of a derivative (hint: it's what happens when you combine slopes of lines with limits to zero).

Calculating slope What is the slope of the line in Figure 5.1? See if you can calculate it just based on the information provided in the graph.

I hope you were able to determine that the slope is 5/4 (1.25 is an equally acceptable answer). How did you get that number? I guess you quantified the change in the y-axis by the change in the x-axis. The function traversed 5 units on the y-axis ($y = 1$ to $y = 6$) while traversing 4 units on the x-axis ($x = -1$ to $x = 3$). In fact, you calculated the following formula:

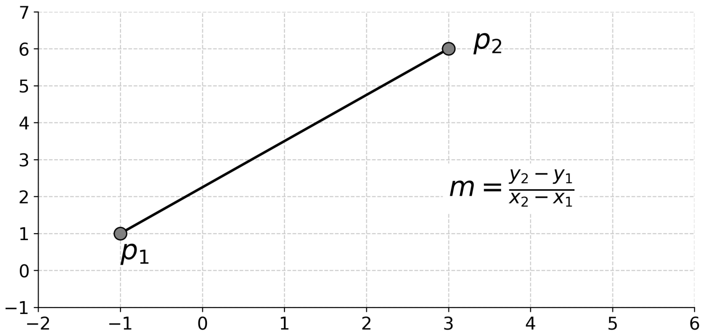

Figure 5.1: What is the slope of this line?

$$m = \frac{\Delta y}{\Delta x} \tag{5.1}$$

m is the common variable name to indicate the slope. It comes from the equation for a line, which you've seen in Chapter 3 when I introduced the difference between linear and nonlinear equations.

$$y = mx + b \tag{5.2}$$

As a reminder, m is the slope and b is the intercept, which is the value of the function when $x = 0$; in other words, where the function line crosses the y-axis. Importantly, m and b are separate parameters: you can have an infinite number of functions with the same slope but different intercepts (Figure 5.2). This also means that if I give you a value m and ask you to reconstruct the line, you wouldn't know the exact line to reconstruct; you would only know how to set the angle of that line. This is an extremely important concept in calculus, and in particular, when integrating functions.

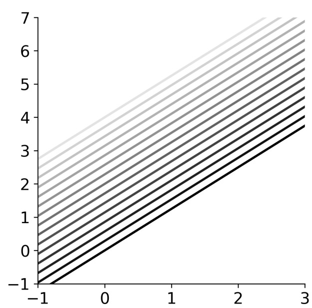

Figure 5.2: 15 lines that have identical slope ($m = 5/4$) and different intercepts.

One final point before moving on: The slope of a perfectly vertical line is undefined. That's clear from Equation 5.1: If the line moves straight up, then it doesn't change on the x-axis, which means the denominator of the slope equation is zero.

Two slopes Now consider Figure 5.3, which shows a piecewise function comprising two linear segments. What's the slope of the line depicting this piecewise function?

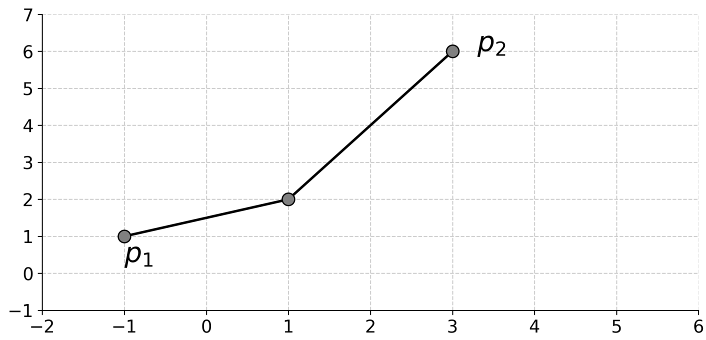

Figure 5.3: A piecewise function.

I hope you have a feeling of unease and confusion when looking at that line with the goal of calculating its slope. Sure, you could calculate the slope of a line that directly connects points p_1 and p_2 (the slope of that line would be 5/4), but that line is not actually drawn. The line that is actually drawn doesn't have *one* slope. It is a line comprising two segments, each with a different slope. Measuring one slope between points p_1 and p_2 isn't really wrong per se, but it misses important information about the function.

I am sure you agree that the most accurate approach here is to label that third point p_3 and then calculate two slopes for that line: the slope $m_{1\to3}$ and the slope $m_{3\to2}$. (Those slopes are, respectively, 1/2 and 2.)

Many slopes I am also sure that you see where this is going: The more segments a line has, the more slopes we can use to characterize that line. How many slopes would you need for the line in Figure 5.4?

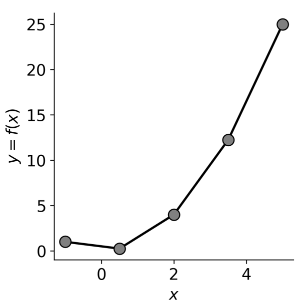

Figure 5.4: Five points describe a piecewise function comprising four segments.

The answer is four. That function has five points, and the slope is defined as the difference between successive points, and so that leads to four slopes. That's a general feature of computing discrete differences: A line defined by N points has $N-1$ segments with $N-1$ slopes, and could be defined as a piecewise function with $N-1$ pieces. This will be relevant for computing *discrete differences*, which is basically the digitized version of a derivative. It's also how you calculate the derivative of an empirical function that is not defined by a mathematical formula (for example, empirical time series data).

The examples I've shown so far are all straight line segments. But lots of functions are not represented by straight lines; they are represented by curves. How can we adapt this idea of multiple slopes for multiple line segments to characterize a curve like x^2? The answer is in the next section, but I encourage you to think of an answer based on the discussion above and on what you learned in the previous chapter.

5.2 Formal definition of derivative

The derivative is conceptually straightforward: calculate the slopes of small line segments. How small are those line segments? They are really *really* small. So tiny, in fact, that they are almost zero-length long. In fact, a derivative is the sequence of slopes of the line segments as the x-axis distance goes to zero:

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \tag{5.3}$$

I hope that equation makes sense and feels intuitive after reading the previous section. However, that's not the definition of a derivative that you normally see. The formal definition is as follows; before reading the text below, please take a moment to understand how the following equation expresses the same concept as in Equation 5.3.

$$\frac{dy}{dx} = \lim_{h \to 0} \left[\frac{f(x+h) - f(x)}{h} \right] \tag{5.4}$$

This definition is called the "limits-based definition" or the "incremental form" of the derivative. It is one of the most important equations in all of calculus, so I will explain each part of it in detail.

dy* and *dx

dx is the same as Δx but for really tiny steps. We use Δ for "big" intervals and d for "tiny" intervals; hence, Δx indicates slopes and dx indicates derivatives. What's the cut-off for "big" vs. "tiny"? Basically, in terms of calculus, any number that you can think of is "big", and any number that is so incredibly tiny that it's not really distinguishable from zero (though it is technically bigger than zero) is "tiny." In terms of Python implementation, `sympy` is capable of representing dx whereas anything done in `numpy` can be considered Δx.

You might be tempted to use the term "infinitesimal" (infinitesimal is $1/\infty$) — that word was historically used in calculus but it fell out of fashion and modern calculus is built on the idea of limits to zero instead of infinitesimals. You can think of dx as what happens when the step size of Δx goes to zero. Of course, dy is the same concept but reflects how much the function changes on the y-axis for each tiny line segment defined by dx. dy therefore depends on the definition of the function, whereas dx involves cutting up the x-axis grid to vanishingly small intervals.

h

This is a short-hand for Δx. h is easier and faster to write. Think of h as the vanishingly tiny discretization of the x-axis.

You can immediately see from Equation 5.4 that the plug-in method won't solve this limit problem. That's the difficult part of differentiation: finding the tricks that will solve this limit problem for each type of function. Once you solve the limit problem, you have the derivative.

f(x)

This is simply the function that we're trying to differentiate.

Δy refers specifically to two values on the y-axis. Replacing Δy with $f(x+\Delta x)-f(x)$ makes the equation more general (recall the discussion in Chapter 2 about the difference between y and $f(x)$). Therefore, as Δx shrinks down to dx, Δy shrinks down to dy.

How to pronounce *dy/dx*

There are several different notations for the derivative (Leibiz, Newton, Lagrange, Euler) that I will explain in Section 5.3. dy/dx is one of the most common notations, and is pronounced as "dee why by dee ex," "dee why dee ex," "dee by dee ex," or "the derivative of why with respect to ex." You will also see df/dx for function $f(x)$, or sometimes d/dx.

x is a commonly used variable name, but you will see other letters. For example, dy/dt might indicate the derivative over time, dy/ds might indicate the derivative over space, $dy/d\theta$ might indicate the derivative over angles, etc. Similarly, the derivative of function $r(\tau)$ would be written $dr/d\tau$.

5.2.1 Example with a linear function

The first example will involve using the derivative definition to confirm the slope of the line in Figure 5.1. The main goal is to warm you up to working with Equation 5.4.

The function that defines that line is $f(x) = \frac{5}{4}x + \frac{9}{4}$. Let's set $h = 1$ and we will pick a specific value of $x = 0$.[1] Here's how that looks:

$$m = \lim_{h\to 1}\left[\frac{f(x+h) - f(x)}{h}\right] \tag{5.5}$$

When h=1, the plug-in method solves this limit.

$$= \left[\frac{\left(\frac{5}{4}(0+1) + \frac{9}{4}\right) - \left(\frac{5}{4}0 + \frac{9}{4}\right)}{1}\right] \tag{5.6}$$

$$= \left(\frac{5}{4} + \frac{9}{4}\right) - \frac{9}{4} \tag{5.7}$$

$$= \frac{5}{4} \tag{5.8}$$

Notice that the intercept canceled. That's not a quirk of this particular example — the intercept will *always* cancel in the derivative formula because the intercept does not multiply an x, and therefore is unaffected by $x + h$. And that is why an infinite number of functions with different intercepts can have identical derivatives. Pin that concept in the back of your mind; it becomes important later in the book when learning about integrals and antiderivatives.

Perhaps you think this wasn't a good illustration of the derivative formula, because I set the limit as $h \to 1$ and picked one specific value for x. Fair point; I accept your criticism.

Let's try it again, with the limit to zero and using the variable x instead of a specific value of x:

$$\frac{dy}{dx} = \lim_{h\to 0}\left[\frac{f(x+h) - f(x)}{h}\right] \tag{5.9}$$

$$= \lim_{h\to 0}\left[\frac{\left(\frac{5}{4}(x+h) + \frac{9}{4}\right) - \left(\frac{5}{4}x + \frac{9}{4}\right)}{h}\right] \tag{5.10}$$

$$= \lim_{h\to 0}\left[\frac{\frac{5}{4}x + \frac{5}{4}h + \frac{9}{4} - \frac{5}{4}x - \frac{9}{4}}{h}\right] \tag{5.11}$$

[1] $x = 0$ is a convenient choice because it simplifies the arithmetic. You should try it again using $x = 1$ or $x = \pi^{\sqrt{e^{17}}}$ or any other number. It won't affect the final result in this example.

$$= \lim_{h\to 0}\left[\frac{\frac{5}{4}h}{h}\right] \tag{5.12}$$

$$= \frac{5}{4}\lim_{h\to 0}\left[\frac{h}{h}\right] = \frac{5}{4} \tag{5.13}$$

A few remarks:
(**1**) Notice that x is replaced by $x+h$ in the function definition (Equation 5.10). It's not so confusing in this example because there is only one x in the function, but this is a source of mistakes when differentiating, so be mindful of replacing x with $x+h$ in the first function term in the numerator.
(**2**) Even though we used the variable x instead of a specific value of x, it still canceled to give a final derivative of the function that equals the slope of the line. That does not happen for all functions; it is a unique property of straight lines.
(**3**) Equation 5.13 relied on the factorable property of limits.
(**4**) The plug-in method seems like it would have failed, but we ended up with h/h. That is *not* $0/0$ — remember we are taking the limit *as h approaches but does not exactly equal zero.* In other words, h is an exceedingly small number, and when divided by itself, equals 1.

5.2.2 Example with a quadratic function

A quadratic function is any 2^{nd} order polynomial, thus a function of the form $f(x) = ax^2 + bx + c$ where $a \neq 0$.

Now let's differentiate the function $f(x) = x^2$. The full solution is shown below, but I encourage you to try it on your own first: Start from Equation 5.4 and keep working with the equations until you can solve the limit and obtain a dy/dx for $f(x) = x^2$.

$$\frac{dy}{dx} = \lim_{h\to 0}\left[\frac{(x+h)^2 - x^2}{h}\right] \tag{5.14}$$

$$= \lim_{h\to 0}\left[\frac{x^2 + 2xh + h^2 - x^2}{h}\right] \tag{5.15}$$

$$= \lim_{h\to 0}\left[\frac{h(2x+h)}{h}\right] \tag{5.16}$$

$$= \lim_{h\to 0}\,[2x+h] \tag{5.17}$$

$$\frac{dy}{dx} = 2x \tag{5.18}$$

I hope you were able to derive $2x$ on your own. If you struggled with your work and looked at the answer above, then I now encourage you to read through my remarks below, then try it again on your own.

Remarks: (**1**) Again, it seems like the plug-in method would fail, but some algebraic expansions and simplifications allows us to cancel the h from the denominator.
(**2**) There are no problems when plugging in $h = 0$ outside the denominator (Equation 5.17).
(**3**) Even for this simple function, the algebra can get complicated with the substitutions and cancellations. Never rush through a differentiation problem, and never hesitate to check your work.
(**4**) Notice that the derivative here has an x, unlike the derivative for a straight line in the previous example.

5.2.3 Proof: Derivative of a constant is 0

Here is a fact:

$$f(x) = c \tag{5.19}$$

$$\frac{d}{dx} f(x) = 0 \tag{5.20}$$

In other words, the derivative of any constant is zero. I'm sure you agree that that is an intuitive fact: A function that is flat along the x-axis has a slope of zero. The proof of that claim is simple, and involves applying the derivative formula (Equation 5.4) to the function $f(x) = c$ and solving the limit. I'm going to show the proof below, but I encourage you to work through it on your own first.

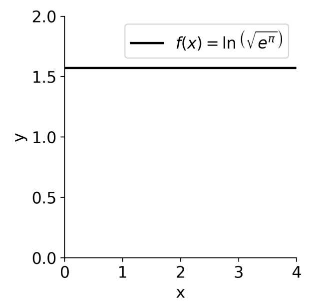

Figure 5.5: A constant function doesn't change on the *y*-axis, and therefore has a derivative of zero.

$$f(x) = c \tag{5.21}$$

$$\frac{df}{dx} f(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{5.22}$$

$$= \lim_{h \to 0} \frac{c - c}{h} \tag{5.23}$$

$$= \lim_{h \to 0} \frac{0}{h} = 0 \tag{5.24}$$

Because the function does not include an x term, there is nowhere to substitute h. Thus, the numerator cancels to zero. You might think that this produces a fraction of $0/0$ as $h \to 0$, but remember that h is not exactly zero; it's a number that is fantastically close to zero. In the expression $0/h$, we don't need to plug in $h = 0$ to see that the fraction equals zero for any value of h that is not exactly equal to zero.

Now you've seen a few examples of applying the differentiation formula. What do the results mean and how can you interpret them? That's the topic for the rest of this chapter — and the next five chapters.

But first, I have an assignment for you: Write down Equation 5.4 on a piece of paper many times. Maybe three times. Maybe seven times. That equation is one of the most important in calculus, and is worth committing to memory. And not just memorizing, but understanding the simple but powerful concept that the derivative of a function is how much the output of a function (y) changes when its input (x) changes by a tiny amount.

5.3 Various notations for the derivative

There are several notations for the derivative. Four of which are listed below, organized by their namesake (left) and the derivative notation (right). The Leibniz and Lagrange notations are most commonly used; Newton's notation is common in physics when studying temporal derivatives. Euler's notation is not commonly used for single-variable differentiation but is used in multivariable functions[2].

Leibniz	$\frac{dy}{dx}$
Lagrange	f'
Newton	$\dot{y}$
Euler	$\mathbf{D}f$

You may find it confusing or annoying that there are different notations for the same concept. To some extent, that is a fair reaction. But each

[2]According to Wikipedia, this notation was actually developed by Louis Arbogast although it is refered to as Euler's notation.

notation has both conceptual and typographical advantages, and there are reasons for learning and using each format. I will write brief descriptions below, and you'll learn more about these notations as you progress through the book.

Leibniz (*dy/dx*)

This notation is conceptually the most enlightening because it matches the equation for the slope of a line. As I wrote earlier, you'll also see variations like df/dx and d/dx. Leibniz notation is also illuminating when learning about the chain rule $\left(\frac{dy}{du} = \frac{dy}{dx} \cdot \frac{dx}{du}\right)$ and differential equations $\left(\frac{dy}{dx} = k \Rightarrow dy = dx\,k\right)$.

A disadvantage of Leibniz notation is that it is cumbersome to write out. That may seem like an excuse for laziness, but when equations become dense and copious, compact notations can be the difference between comprehension and a massive headache.

Lagrange

This is a popular notation because it is compact and elegant. f' (read out loud as "eff prime") is the derivative of f. You can also write $f'(x)$ or f'_x to indicate the derivative with respect to x (c.f. f'_θ or f'_t). This notation is also convenient because it can be written without any LaTeX or other fancy formatting.

A disadvantage of the Lagrange notation is that it is not as conceptually illuminative as the Leibniz notation.

Newton

Newton's notation ($\dot{y}$) is compact, but also requires special formatting tools on a keyboard. The dot on top can also easily be missed if the font size is small, the pen is leaking, the printer has ink smudges, or the laptop screen is dirty. If you study changes over time in a physics course, you may use Newton's notation. I won't use it in this book, and mention it here only so that you can recognize it.

Euler's notation

Euler's notation for single-variable derivatives is not commonly used. I suspect that's because it has no real advantages over the Lagrange notation, and is easily confused with the notation to indicate the domain of a function. Euler's notation is more commonly used for partial derivatives, and I will introduce that flavor of the notation later in the book.

One more bit of notation: Instead of explicitly defining a function $f(x) =$ ... and then notating its derivative using $f'(x)$, you can indicate the derivative of an expression by applying one of the derivative operators directly to an expression, as in one of these examples:

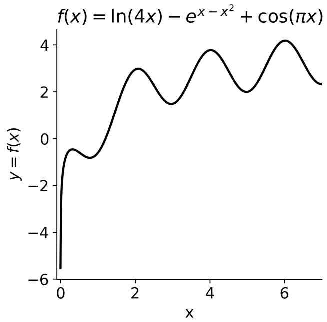

Figure 5.6: Equation 5.25 visualized.

$$\frac{d}{dx}\left[\ln(4x) - e^{x-x^2} + \cos(\pi x)\right] \tag{5.25}$$

$$\left(\ln(4x) - e^{x-x^2} + \cos(\pi x)\right)' \tag{5.26}$$

5.3.1 Geometric terms

There are several geometric terms that are relevant for differential calculus, and in this section you will learn two: *tangent* and *secant*.

I'm going to start by providing definitions that are over-simplified and therefore not entirely correct. But I believe that introducing the concepts this way — and then correcting them — will help build your critical thinking skills. While reading the definitions below, please inspect Figure 5.7 and think about why the definitions are imprecise, possible failure scenarios, and how the terms could be more precisely defined.

Tangent line (dotted line in Figure 5.7)
A tangent line passes through a function at one point.[3]

Secant line (dashed line in Figure 5.7)
A secant line passes through a function at two points.

Do you have any thoughts about how those definitions might go awry? In fact, the graphs in Figure 5.7 fit perfectly with the the definitions above.

The problem with those definitions is that they aren't specific enough. Consider Figure 5.8. Panel A fits the tangent definition above, but it is not a tangent line. Panel B shows a tangent line, but it passes through the function at more than one point.

Here are more precise definitions:

Tangent line
A tangent line is a straight line with a slope equal to the derivative of a

[3]You might be familiar with a "tangential conversation," like when you're talking to your mother about dinner plans and she suddenly reminds you of something inappropriate you said when you were 5 years old.

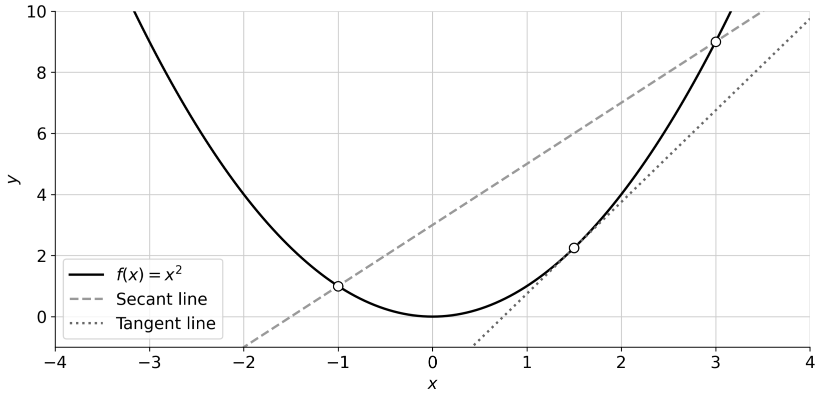

Figure 5.7: Tangent and secant lines on a function. You will write code to generate this figure in Exercise 6.

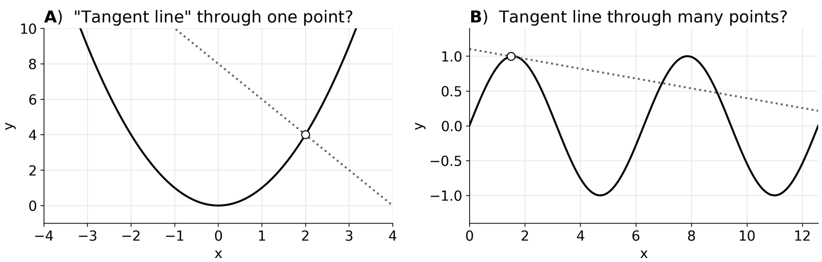

Figure 5.8: Panel A shows a line that touches the function once but is not a tangent line. Panel B shows a true tangent line that touches the function at multiple points.

function at the point at which the tangent line intersects the function. (The dotted line in Figure 5.8A is not a tangent line.)

Secant line

A secant line intersects a function at two or more points, and the slope of the secant line is defined from two intersection points. (The dotted line in Figure 5.8B is both a tangent and a secant.)

There are simple algorithms to find the equation for tangent and secant lines. You'll learn and implement those algorithms in Exercise 6.

5.4 Interpreting derivatives plots

The derivative is a *function.* It is a function of x just like the function it was computed from. It may seem like the derivative of the line is not a function, but it is: $f(x) = m$. There is no x term in the function definition, meaning this function does not vary — it's a flat line from $-\infty$ to $+\infty$.

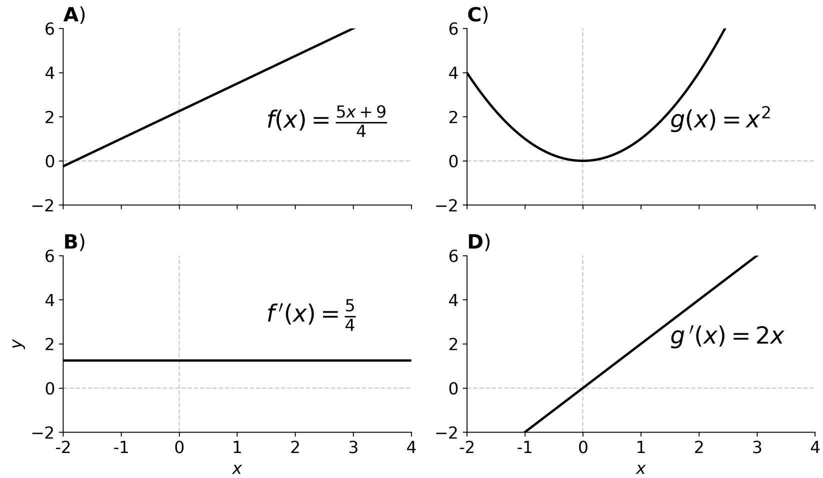

Figure 5.9: Two functions (top row) and their derivatives (bottom row).

Figure 5.9 shows two functions we worked with earlier (panels A and C on the top row) and their derivatives (panels B and D on the bottom row).

If you are confused at how the derivative relates to the function, then that is perfectly normal. Everyone gets confused when they are new to looking at plots of derivatives, because they expect them to be closely visually related to the original functions.

The linear function (panel A) has a constant slope for all values of x. The derivative of that function (panel B) is a flat line, because the slope doesn't change for different values of x. $f(x)$ goes up at exactly the same rate, regardless of whether x is negative, positive, large, small, etc.

On the other hand, the quadratic function (panel C) does not have a constant slope; the slope changes depending on the value of x. x^2 curves down for $x < 0$, is momentarily flat when $x = 0$, and then curves up for $x > 0$.

The graph of the derivative is a straight line with a positive slope that touches $x = y = 0$. First, just think about the sign of the derivative, ignoring its value. When x^2 curves downward, its derivative is negative. And when x^2 curves upwards, its derivative is positive. So, the sign of the derivative isn't telling us about the *value* of the function; it's telling us about the *direction* of the function as it moves to the right on the x-axis.

The first thing to look for in a derivative plot is the sign of the derivative (ignoring the actual values).

Now let's consider the numerical values of the derivative (not just the sign). The values of the derivative encode how fast the function changes. When the derivative is closer to zero on the y-axis, the function changes relatively slowly (gentle slope), and when the derivative is further from zero on the y-axis, the function changes faster. Now looking back to the original function in panel C, x^2 changes faster and faster (that is, greater slopes) the further away it gets from $x = 0$, while the closer it gets to $x = 0$, the more slowly it changes on the y-axis.

I hope that makes sense. Please take a while to stare at Figure 5.9 until it does make sense. The more time you invest in understanding derivatives plots now, the easier you'll find the rest of calculus.

5.4.1 More examples

Please take a few moments to study Figure 5.10. Try to see how each of the three functions in the top row relates to their derivatives in the bottom row. Don't worry about how I obtained the derivatives; you'll learn that over the next several sections. For now, just focus on trying to match the *changes* in the function to the *values* of the derivative.

In particular, answer the following two questions for each column: (**1**) What is the *sign* of the derivative when the function goes up vs. down; (**2**) What is the *value* of the derivative when the function *changes direction*?

Panels A and B

$f(x)$ mostly goes upwards as x increases, but there is a piece in the middle where $f(x)$ decreases. Where $f(x)$ increases, $f'(x)$ is positive, and where $f(x)$ decreases, $f'(x)$ is negative. Also notice that the x-axis values at which $f(x)$ transitions from going up to going down correspond to zero-crossings in $f'(x)$. In other words, $f' = 0$ when f changes direction. These are called "critical points" and I'll discuss them in more detail in the next chapter.

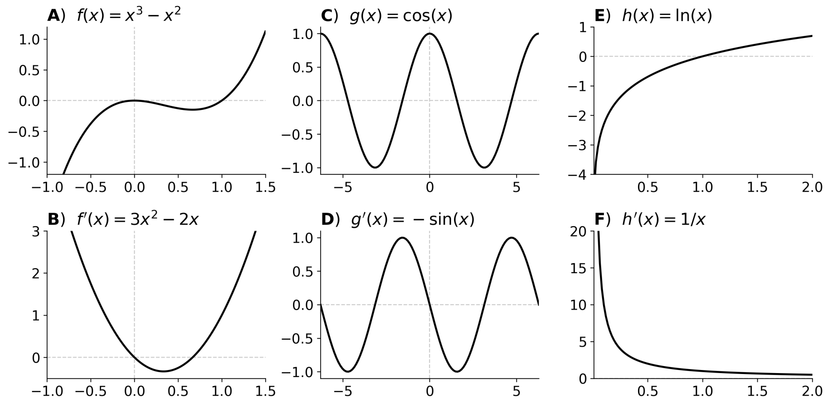

Figure 5.10: Three more functions (top row) and their derivatives (bottom row).

Panels C and D

Let's start at $x = 0$. $g(x)$ starts out momentarily flat; its slope at $x = 0$ is $g' = 0$. Then g goes down — first steeply then more gently — until it reaches $y = -1$, and which point we have another moment of a horizontal tangent, then it bends upwards until $y = 1$ and the cycle repeats. *Ignore your first instinct to think that g' is a shifted version of g.* (In fact, it is a shifted version, but that's not how you want to look at it.) Instead, focus on how the *values* of the derivative reflect the *changes* in the function. The steeper g decreases, the more negative g'; when g changes from going downwards to going upwards, $g' = 0$.

Panels E and F

The natural log function starts negative and goes upwards. It never stops going up. The *rate* at which $h(x)$ increases, decreases. That is, the larger the value of x, the slower $h(x)$ increases, but it's never flat and never bends downwards. Notice how this behavior manifests in the derivative: $h'(x)$ is always positive; is very large close to $x = 0$, and gets smaller and smaller as x increases but never touches zero or becomes negative.

Ignore your instinct to think that $1/x$ is the inverse of $\ln(x)$; e^x is the inverse of $\ln(x)$. Inverses are not derivatives.

It takes some time and practice to develop the skill of interpreting plots of functions and their derivatives. It's a skill that you will continue to improve throughout the rest of this book.

Three final messages for this section:

1. The most important thing to remember about the plot of a derivative is that it shows the slopes of the function at each point. Not the *values* of the function, but the *changes* in the function at each

x-axis value.

2. The plot of a derivative in general looks nothing like the plot of its function, with a few noteworthy examples that I'll introduce later, such as sine and e^x.

3. Look for different features in plots of functions vs. derivatives. When inspecting a function, look at its values, behavior, striking or unusual parts, etc. When inspecting a derivative, focus on the sign and zero-crossings, and look at the value of the derivative only thereafter.

5.5 Differentiation is linear

What does it mean for differentiation to be a linear operator? Mathematically, it is described by Equation 5.27; try to understand what this equation means before reading the text below (α and β are two constants).

$$\frac{d}{dx}\big[\,\alpha f(x) + \beta g(x)\,\big] \;=\; \alpha f'(x) + \beta g'(x) \tag{5.27}$$

There are two operations described by Equation 5.27, one about scalar multiplication and one about addition. Let's discuss the multiplication part first.

Scalar multiples can be factored out of derivatives, just like how they were factored out of limits.

$$\frac{d}{dx}\big[\,\alpha f(x)\,\big] \;=\; \alpha\frac{d}{dx}\big[\,f(x)\,\big] \tag{5.28}$$

In other words, the derivative of a constant times a function equals that constant times the derivative.

The next part of linearity is that the derivative of the sum of two functions equals the sum of the derivatives of those functions:

$$\frac{d}{dx}\big[\, f(x) + g(x) \,\big] \;=\; \frac{d}{dx}\big[\, f(x) \,\big] \;+\; \frac{d}{dx}\big[\, g(x) \,\big] \tag{5.29}$$

Equation 5.27 combines those two concepts into one line.

The linearity of differentiation is useful because it allows you to break down complicated multiterm functions into a collection of smaller, more manageable parts.

Figure 5.11 shows a demonstration of linearity. The plot shows the following two derivatives:

$$v_1(x) = \left[\; 3x^3 - \frac{2}{\pi}\sqrt{\cos(2\pi x) + 1} \;\right]' \tag{5.30}$$

$$v_2(x) = 3\left[\; x^3 \;\right]' + \frac{2}{\pi}\left[\; -\sqrt{\cos(2\pi x) + 1} \;\right]' \tag{5.31}$$

If differentiation is linear according to Equation 5.27, then the two lines should overlap perfectly in the plot. And they do! On the other hand, this is merely a demonstration with two specific functions and two specific constants. It is necessary to prove that the linearity claim is valid for any function and any constant.

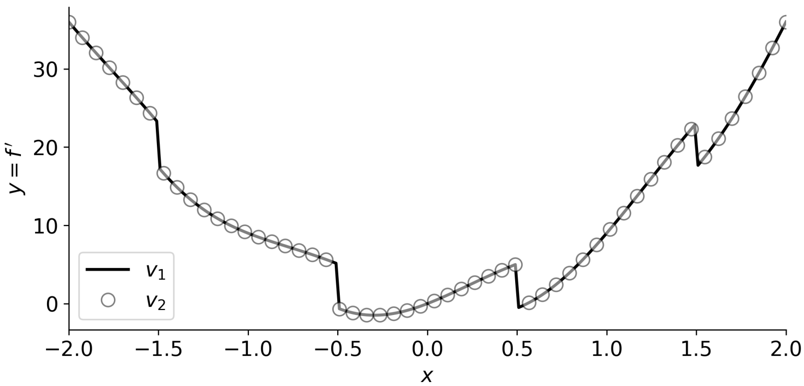

Figure 5.11: Demonstration of the linearity of differentiation. v_2 was down-sampled by a factor of four to facilitate visualization.

5.5.1 Proof of linearity

Intuitively, you can understand that derivatives are linear because they are merely an application of limits, and limits are linear.

That said, the linearity of differentiation is straightforward to prove, and involves expanding Equation 5.27 using the limits definition of differentiation. The sequence of equations is below. As with other proofs in this chapter, this one may look intimidating at first, but I think you'll find it manageable when comparing each equation to the previous.

$$
\begin{aligned}
k(x) &= \alpha f(x) + \beta g(x) && (5.32)\\
k'(x) &= \lim_{h\to 0} \frac{k(x+h) - k(x)}{h} && (5.33)\\
&= \lim_{h\to 0} \frac{[\alpha f(x+h) + \beta g(x+h)] - [\alpha f(x) + \beta g(x)]}{h} && (5.34)\\
&= \lim_{h\to 0} \frac{\alpha f(x+h) - \alpha f(x) + \beta g(x+h) - \beta g(x)}{h} && (5.35)\\
&= \lim_{h\to 0} \frac{\alpha\,[f(x+h) - f(x)] + \beta\,[g(x+h) - g(x)]}{h} && (5.36)\\
&= \lim_{h\to 0} \alpha \left[\frac{f(x+h) - f(x)}{h}\right] \;+\; \lim_{h\to 0} \beta \left[\frac{g(x+h) - g(x)}{h}\right] && (5.37)\\
&= \alpha \lim_{h\to 0} \left[\frac{f(x+h) - f(x)}{h}\right] \;+\; \beta \lim_{h\to 0} \left[\frac{g(x+h) - g(x)}{h}\right] && (5.38)\\
&= \alpha f'(x) + \beta g'(x) && (5.39)
\end{aligned}
$$

The only "trick" to this proof is re-ordering the terms to group together the $f(x)$ terms separately from the $g(x)$ terms, and applying the properties of limits you learned in the previous chapter.

5.6 Continuity is necessary but not sufficient

A derivative is defined at some point $x = a$ only if the function is defined at that point. Consider the following function:

$$f(x) = \frac{|x-1|}{x-1} \tag{5.40}$$

What is the derivative of that function at $x = 1$? Well, the function doesn't exist at that point ($f(1) = 0/0$), so it doesn't make sense to think about its derivative at that point. No function value, no slope, no two-sided limit, no derivative.

Another example is the logarithm: what is the derivative of the logarithm function at $x = -4$? The function is not defined for negative numbers, so it doesn't make sense to talk about a derivative for $x \leq 0$.

Is continuity sufficient for differentiation? You might have guessed that the answer is *no*, considering the bolded question. But the relevant question is *why* continuity is not sufficient for a derivative to exist.

Consider the function in Figure 5.12A. It is continuous everywhere, but how can you draw the tangent line at $x = 0$? The answer is you cannot: there is no unique slope at that cusp point. Instead, an infinite number of straight lines can pass through that point.

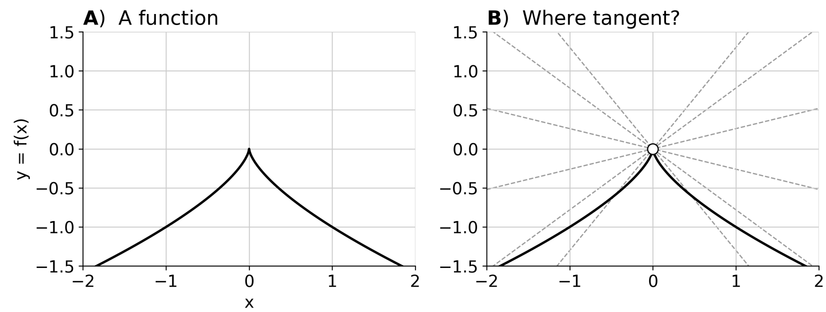

Figure 5.12: A function that is continuous at all points but has no derivative at $x = 0$. Panel B illustrates six of the unlimited number of possible straight lines that pass through the function at that point.

A function that is differentiable for all values of x is called a "differentiable function" (alternatively, that the function has "total differentiability"). If there are some points in the function where the derivative exists and other points where the derivative does not exist, then it's nondifferentiable.

Many nondifferentiable functions can be differentiable within restricted domains. For example, you could describe a function as being "differentiable everywhere except $x = 0$" or you could write that the function is "differentiable only in the intervals $(-\infty, 0) \cup (0, \infty)$".

5.7 Derivatives of polynomials

Differentiate polynomials with this one simple trick!

I will start this section by showing the rule and then a few examples.

$$f(x) = cx^r \tag{5.41}$$

$$f'(x) = rcx^{(r-1)} \tag{5.42}$$

This is called the "power rule" of differentiation.

where c is any constant and r is any constant in the exponent. In words: to differentiate a polynomial, bring the power down as a multiplying coefficient and reduce the power by one. Please apply this rule to differentiate the following functions, and then check your work in the footnote.[4]

a) x **b)** x^2 **c)** $5x^5$

d) $4 + 5x + 6x^{20}$ **e)** $\frac{x^2}{2}$ **f)** $x^5 - 3x^2 + .25x^4$

The exception to this rule is x^0: Any number raised to the power of zero is 1, and thus the derivative of x^0 is 0.

Can you differentiate the following function? Try it yourself before reading the next paragraph.

$$f(x) = e^{\pi}x^2 - \pi^e x^4 \tag{5.43}$$

The answer is *yes*, of course you can differentiate that function! The coefficients may look strange, but those are not functions of x; they are

[4] **a)** $1x^0 = 1$ **b)** $2x$ **c)** $25x^4$ **d)** $5 + 120x^{19}$ **e)** x **f)** $5x^4 - 6x + x^3$

just constants. Thus, the derivative of that function is $f' = 2e^{\pi}x - 4\pi^{e}x^{3}$

How about the following two functions; can you differentiate them?

$$f(x) = \sin^3(x^2) \tag{5.44}$$

$$g(x) = x^3 e^x \tag{5.45}$$

Here the answer is *no*, you cannot differentiate them. Well, to clarify: Both of these functions are perfectly differentiable, and can be differentiated with little time and effort. But at this point in the book, you haven't learned the rules to differentiate these two functions. Equation 5.44 requires the *chain rule* because of the composite function, and Equation 5.45 requires the *product rule* because two functions of x are multiplied. You will learn both of these rules in the next chapter. Until then, you can only differentiate individual polynomial terms.

Alternating power parity An interesting result of the power rule is that the derivative of an even polynomial is an odd polynomial, and the derivative of an odd polynomial is an even polynomial. Observe:

$$f = x^2, \quad f' = 2x^1$$
$$f = x^3, \quad f' = 3x^2$$
$$f = x^4, \quad f' = 4x^3$$
$$f = x^5, \quad f' = 5x^4$$

The power rule describes why this happens algebraically, and Exercise 8 will guide you to exploring why this happens geometrically.

Proving the power rule There are several proofs of the power rule. A few of the proofs work only for positive integer powers, although the power rule works for any number in the exponent (except zero). A general proof that works for all powers — and is compact and elegant — relies on a procedure called *implicit differentiation*. That's a topic for Chapter 7. So for now, please just trust me that the power rule really is correct, and rigorously proving that it is correct is a good motivation to continue working through this book ;)

That said, I would like you to demonstrate the power rule for one specific function, using the limits-based definition of the derivative (Equation 5.4 — that's the formula I told you write down many times in order to memorize).

Use the function $f(x) = 2x^3$, apply Equation 5.4, and work with the function until you can solve the limit problem. My answers are below, but I hope you take this opportunity to work through the problem yourself before checking my work.

$$\frac{dy}{dx} = \lim_{h\to 0}\left[\frac{2(x+h)^3 - 2x^3}{h}\right] \tag{5.46}$$

$$= \lim_{h\to 0}\left[\frac{2(x+h)(x+h)(x+h) - 2x^3}{h}\right] \tag{5.47}$$

$$= \lim_{h\to 0}\left[\frac{2(x^2+h^2+2xh)(x+h) - 2x^3}{h}\right] \tag{5.48}$$

$$= \lim_{h\to 0}\left[\frac{2(x^3+x^2h+h^2x+h^3+2x^2h+2xh^2) - 2x^3}{h}\right] \tag{5.49}$$

$$= \lim_{h\to 0}\left[\frac{2x^3+2x^2h+2xh^2+2h^3+4x^2h+4xh^2 - 2x^3}{h}\right] \tag{5.50}$$

$$= \lim_{h\to 0}\left[\frac{h(2x^2+2xh+2h^2+4x^2+4xh)}{h}\right] \tag{5.51}$$

$$= \lim_{h\to 0} 6x^2 + h(6x+2h) \tag{5.52}$$

$$= 6x^2 \tag{5.53}$$

As I wrote earlier in this chapter, differentiating via Equation 5.4 can be time-consuming and sometimes tricky. But I hope you agree that the insights gained are worth the effort.

5.7.1 Derivative of square root

The square root function $\sqrt{x}$ is not a polynomial, because polynomials are limited to positive-integer powers.

However, the square root function obeys the power rule of differentiation:

$$\sqrt{x} \;=\; x^{1/2} \tag{5.54}$$

$$\left[\sqrt{x}\,\right]' \;=\; \frac{1}{2}x^{-1/2} \;=\; \frac{1}{2\sqrt{x}} \tag{5.55}$$

Thus, you can differentiate the square root by rewriting it using fractional powers as in Equation 5.54.

5.8 Derivatives of cosine and sine

Cosine and sine have a neat cyclic pattern of derivatives. Behold:

$$\begin{aligned}
[\,\cos(x)\,]' &= -\sin(x)\\
[\,-\sin(x)\,]' &= -\cos(x)\\
[\,-\cos(x)\,]' &= \sin(x)\\
[\,\sin(x)\,]' &= \cos(x)
\end{aligned}$$

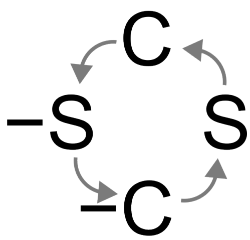

Figure 5.13: The cyclic nature of sine and cosine derivatives.

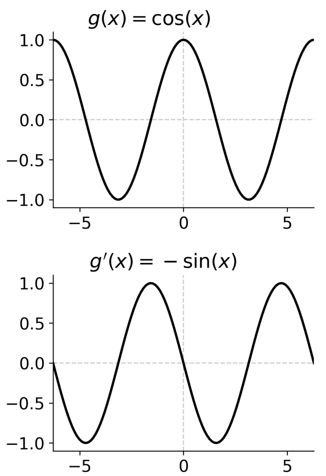

Figure 5.14: Minus the sine is the derivative of cosine. And on and on the cycle goes...

I previously discussed interpreting a plot of cosine and its derivative in Figure 5.10C-D; those panels are redrawn in Figure 5.14 so you don't have to flip back a few pages. Please take a moment to review why the negative sine is the derivative of cosine: Trace the cosine function with your finger and notice how its behaviors (upwards, downwards, transitions between them) are exactly captured by the minus sine function.

5.8.1 Proving the cosine-sin cycle

In this subsection, I will show the proofs that the derivative of cosine is minus sine, and that the derivative of sine is cosine. If you are new to calculus and mathematical proofs, you may find all the equations intimidating. But it's really not so bad. Just take a deep breath and pay attention to the differences between each equation and the previous.

These proofs rely on two trig identities:

$$\cos(\theta + \phi) = \cos(\theta)\cos(\phi) - \sin(\theta)\sin(\phi) \tag{5.56}$$

$$\sin(\theta + \phi) = \sin(\theta)\cos(\phi) + \cos(\theta)\sin(\phi) \tag{5.57}$$

These are called the "angle addition identities."

[cos(x)]' = $-$sin(x) I suggest to write down each of these steps and try to understand how each line follows from the previous.

$$\begin{aligned}
\frac{dy}{dx} &= \lim_{h\to 0}\left[\frac{\cos(x+h)-\cos(x)}{h}\right] && (5.58)\\
&= \lim_{h\to 0}\left[\frac{\cos(x)\cos(h)-\sin(x)\sin(h)-\cos(x)}{h}\right] && (5.59)\\
&= \lim_{h\to 0}\left[\frac{\cos(x)\cos(h)-\cos(x)-\sin(x)\sin(h)}{h}\right] && (5.60)\\
&= \lim_{h\to 0}\left[\frac{\cos(x)(\cos(h)-1)-\sin(x)\sin(h)}{h}\right] && (5.61)\\
&= \lim_{h\to 0}\left[\frac{\cos(x)(\cos(h)-1)}{h}\right] - \lim_{h\to 0}\left[\frac{\sin(x)\sin(h)}{h}\right] && (5.62)\\
&= \cos(x)\lim_{h\to 0}\left[\frac{\cos(h)-1}{h}\right] - \sin(x)\lim_{h\to 0}\left[\frac{\sin(h)}{h}\right] && (5.63)\\
&= \cos(x)\times 0 - \sin(x)\times 1 && (5.64)\\
&= -\sin(x) && (5.65)
\end{aligned}$$

Once you apply the angle addition identity to separate $\cos(x+h)$ into separate terms, the rest of the proof just involves grouping matching terms together and applying the properties of limits, and the trig limits, that you learned in the previous chapter.

Perhaps you were confused with pulling out the $\cos(x)$ and $\sin(x)$ terms in Equation 5.63. These are both functions of x and I explained in the previous chapter that only *constants* can be factored out of a limit. But, here we are taking the limit of h, not of x. h is related to x in that it is a short-hand for Δx, but from the perspective of h, a function purely of x can be considered a constant. That's why any function of x that does not also include h can be factored out. You will encounter this concept again in multivariable differentiation.

[sin(x)]' = cos(x) The proof that the derivative of sine is cosine follows the same procedure as above: apply the angle addition identity, group terms together, and apply properties of limits and the trig limits from the previous chapter. After working through the previous proof, I'm sure you will agree that this one is easier to follow.

$$\begin{aligned}
\frac{dy}{dx} &= \lim_{h\to 0}\left[\frac{\sin(x+h)-\sin(x)}{h}\right] &\quad (5.66)\\
&= \lim_{h\to 0}\left[\frac{\sin(x)\cos(h)+\cos(x)\sin(h)-\sin(x)}{h}\right] &\quad (5.67)\\
&= \lim_{h\to 0}\left[\frac{\sin(x)\cos(h)-\sin(x)+\cos(x)\sin(h)}{h}\right] &\quad (5.68)\\
&= \lim_{h\to 0}\left[\frac{\sin(x)(\cos(h)-1)+\cos(x)\sin(h)}{h}\right] &\quad (5.69)\\
&= \sin(x)\lim_{h\to 0}\left[\frac{\cos(h)-1}{h}\right]+\cos(x)\lim_{h\to 0}\left[\frac{\sin(h)}{h}\right] &\quad (5.70)\\
&= \sin(x)\times 0+\cos(x)\times 1 &\quad (5.71)\\
&= \cos(x) &\quad (5.72)
\end{aligned}$$

Other trig functions? Cosine and sine are arguably the two most important trig functions, but they're not the only ones. The derivative of the tangent function requires the quotient rule, and so I'll discuss that in the next chapter. Derivatives of other trig functions are used less often, and I won't go through all of them. But I do recommend committing the information presented in Figure 5.13 to memory.

5.9 Derivatives of absolute value and signum

Absolute value ($f(x) = |x|$) and signum ($f(x) = \text{sgn}(x)$; "signum" is the sign function, i.e., $y = -1$ for negative x and $y = 1$ for positive x) are nonlinear functions, and their derivatives are also nonlinear functions.

Before looking at the equations below, I would like you to inspect Figure 5.15 and try to infer the derivatives of these two functions.

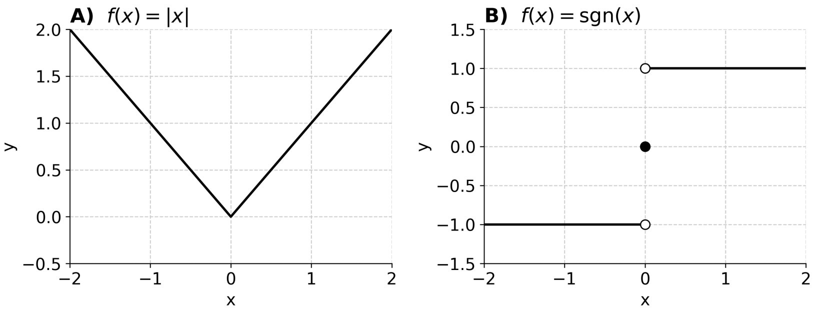

Figure 5.15: Please infer the derivatives of these two functions before reading further.

Let's start with the absolute value function. This function has two pieces: a straight line with a slope of $m = -1$ for any negative x, and a straight line with a slope of $m = +1$ for any positive x. What about when $x = 0$? Zero doesn't have a negative (there is no -0), so $x = 0$ can be included in the positive piece.

The derivative of $f(x) = |x|$ simply involves writing out the piecewise function corresponding to what I wrote above. It turns out that the piecewise definition can be elegantly described by one fraction.

$$|x|' = \left\{ \begin{array}{lll} -1 & \text{if} & x < 0 \\ 1 & \text{if} & x > 0 \\ \text{d.n.e.} & \text{if} & x = 0 \end{array} \right\} = \frac{|x|}{x}, \; x \neq 0 \tag{5.73}$$

Now for the signum function (also called the *sign function*). Before discussing its derivative, I would like to write out the formula for this function. This function is -1 for any negative x and $+1$ for any positive x. What is its value when $x = 0$? Zero doesn't have a sign, so the signum function is either undefined (d.n.e.) or given a value of zero for $x = 0$.

Taken together, signum can be defined using a three-segment piecewise function.

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x = 0 \end{cases} \tag{5.74}$$

You might see the signum function defined more compactly as:

$$\operatorname{sgn}(x) = \frac{|x|}{x}, \ x \neq 0 \tag{5.75}$$

That's pretty neat, because it means that the derivative of the absolute value function is the signum function. Even better: In many applications, a function value *exactly* equal to 0 can be rare due to computer rounding and precision errors. Therefore, in practice, `np.sign(x)` can be used as the derivative of `abs(x)`.

What do you think is the derivative of the signum function? You already know that the derivative of any constant is zero, but this function has three distinct constants in different restricted domains. Don't be confused by that — the derivative is still zero for all values of x except for the one point of transition.

$$[\ \operatorname{sgn}(x)\ ]' = \begin{cases} 0 & \text{for} \quad x \neq 0 \\ \text{d.n.e.} & \text{for} \quad x = 0 \end{cases} \tag{5.76}$$

5.10 Derivatives of log and exp

In this section, you will learn the derivatives of the natural log and natural exponential functions. As with a few other derivatives in this chapter, the formal proof of the claims I'll make below rely on a technique called implicit differentiation. So, for now please just trust that these claims are true, and then you'll be able to prove them yourself in the next chapter.

5.10.1 Logs of various bases, and their derivatives

Here is the main take-home message:

$$f(x) = \log_B(x) \tag{5.77}$$

$$f'(x) = \frac{1}{x\ln(B)} \tag{5.78}$$

Recall from Chapter 3 that the natural log and natural exponential are inverses of each other. Therefore, the natural log, defined as $B = e$, simplifies to $f' = 1/x$ because $\ln(e) = 1$.

I showed the graph of the natural log and its derivative in Figure 5.10 (page 166), and Figure 5.16 below shows the graph again with more log bases. Please take a moment to notice the following features: (**1**) All log bases, and therefore all of their derivatives, have the same basic shape, differing only in the steepness of the increases. (**2**) All log functions increase to ∞, and the bases define how quickly they increase; therefore, all derivatives of log functions are non-negative. (**3**) The log functions are defined only for $x > 0$, and therefore, so are their derivatives. Thus, the derivative of logarithms excludes $x \leq 0$, even though the function $1/x$ is defined for negative x.

Logs can be taken to any non-negative base except $B = 1$, but 2, e, and 10 are the most common bases.

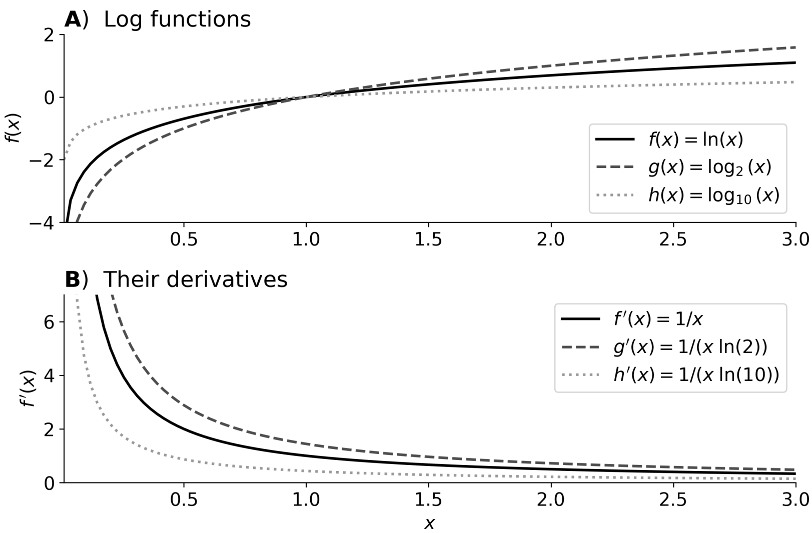

Figure 5.16: Plots of log functions and their derivatives.

If the log function is scaled, that scalar scales the derivative. In particular, for numbers a and b:

$$\Big[a \ln(x+b)\Big]' = \frac{a}{x+b} \tag{5.79}$$

Assuming $x + b > 0$.

5.10.2 Derivative of e^x

The irrational number e is one of the most remarkable and important numbers in all of human-invented mathematics[5], and therefore, the natural exponential function e^x (Figure 5.17) is also one of the most remarkable and important functions.

The remarkable thing about e^x in the context of calculus is that its derivative is also e^x.

$$f(x) = e^x \tag{5.80}$$

$$f'(x) = e^x \tag{5.81}$$

This is the only function that has this property. (The derivative of the constant function $f(x) = 0$ is also zero, but that function can be expressed as $f(x) = 0e^x$.) It means that as x gets larger, e^x increases to ∞. And how fast does that function increase? Well, it increases at the rate of e^x. Not only does it blow up to ∞, the rate at which it increases, increases. At each value of x, the slope of the tangent line going through the function is $m = e^x$.

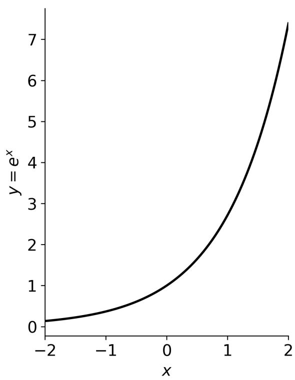

Figure 5.17: e^x

Proving that $[\ e^x\ ]' = e^x$, and that that is the only function that equals its own derivative, requires knowledge of integration. We'll get back to this point later in the book!

Constants in the exponents To differentiate $e^{\alpha x}$ for any constant α, bring down that constant as a multiplicative coefficient — but leave it in the exponent. Additive terms in the exponent are ignored. Some examples:

$$\left(\ e^{2x}\ \right)' = 2e^{2x} \tag{5.82}$$

$$\left(\ e^{\pi x+2}\ \right)' = \pi e^{\pi x+2} \tag{5.83}$$

[5]Whether this number is also important in alien mathematics is unknown and can be confirmed only if we meet our mathematically savvy galactic neighbors.

$$\left(e^{x/2} \right)' = \frac{e^{x/2}}{2} \tag{5.84}$$

For now, commit this rule to memory. When you learn about implicit differentiation in Chapter 7, you will know how to prove this rule.

Why are additive constants ignored? That's because of the law of exponents:

$$e^{2+\pi x} = e^2 \times e^{\pi x} \tag{5.85}$$

$$\left(e^{2+\pi x} \right)' = \pi e^2 e^{\pi x} \tag{5.86}$$

$$= \pi e^{2+\pi x} \tag{5.87}$$

In other words, the additive term in the exponent (in this example, e^2) is actually a multiplicative constant that is placed in the exponent simply because it makes the equation more compact. It's not a function of x.

5.11 Higher-order derivatives

The derivative of a function is itself a function. And that means that you can take the derivative of a derivative. And the derivative of the derivative is also a function, so you can take the derivative of the derivative of the derivative... and so on.

Some functions will eventually yield a constant derivative (e.g., x^n for any non-negative integer n), while other functions are infinitely differentiable (e.g., e^x and $\cos(x)$).

There are no new tricks or techniques to obtain higher-order derivatives. You simply differentiate again. For example:

$$f(x) = \cos(x) + 4x^2 + e^x \tag{5.88}$$

$$f'(x) = -\sin(x) + 8x + e^x \tag{5.89}$$

$$f''(x) = -\cos(x) + 8 + e^x \tag{5.90}$$

$$f^{(3)}(x) = \sin(x) + e^x \tag{5.91}$$

Notation The table below extends the notation table I showed earlier in this chapter; the new column shows how to indicate the second derivative.

Leibniz	$\frac{dy}{dx}$	$\frac{d^2y}{dx^2}$
Lagrange	f'	f''
Newton	$\dot{y}$	$\ddot{y}$
Euler	$\mathbf{D}f$	$\mathbf{D}^2f$

A few notes: (**1**) It may seem strange that the Leibniz notation squares the d and the x but not the y. However, this is short-hand notation. Think about it like this:

$$\left(\frac{d}{dx}\right)^2 y \quad \Rightarrow \quad \frac{d^2}{(dx)^2}y \quad \Rightarrow \quad \frac{d^2y}{dx^2}$$

In other words, (d/dx) is the differentiation operator being applied to y (or f when written df/dx), and that operator is squared.
(**2**) The Lagrange notation has one apostrophe for each derivative, but parenthesized numbers are used after the second derivative. That is, use $f^{(3)}$ instead of f''' and $f^{(10)}$ instead of f''''''''''. Parentheses are used to avoid confusion with powers. Thus: $f^3 = f \times f \times f$ while $f^{(3)}$ is the third derivative of f.

Applications of higher derivatives The first derivative signifies major changes in the function's behavior. And the second derivative signifies changes in the derivative, which means the change in the way that the function is changing.

Higher-order derivatives are used for identifying and interpreting critical and inflection points (Chapter 6), and for differentiation tricks such as L'Hopital's rule (Chapter 7).

Higher-order derivatives have many real-world applications in finance, physics, and engineering. For example, imagine being in a moving car and defining a function of the position of the car. The first derivative is velocity, the second derivative is acceleration, and the third derivative is jerk (in this context, "jerk" is not a rude driver; it is a sudden change in acceleration that causes your head to whiplash). Higher-order derivatives are also used in engineering and materials design, for example computing beam stresses as the fourth derivative, heat transfer, and so on. These calculations help ensure the building you are sitting in doesn't collapse on top of you.

5.12 Exercises

5.1. The purpose of this exercise is to help you understand the link between "local slopes," x-axis grid resolution, and the derivative.

Write `numpy` code to calculate the function $f(x) = x^2$ using $N = 5$ x-axis points linearly spaced between -1 and 5. Connecting these points creates a series of four straight line segments, which you can visualize as in Figure 5.18A. Make sure the segment-calculating code is implemented in a for-loop so that you can change the N parameter.

Next, write code to calculate the slope of each line segment. Visualize the four line segments using horizontal lines, where the x-axis points for each "mini-slope" are defined by the boundary points for each segment, and the y-axis value is the slope of that segment. Visualize this as in Figure 5.18B.

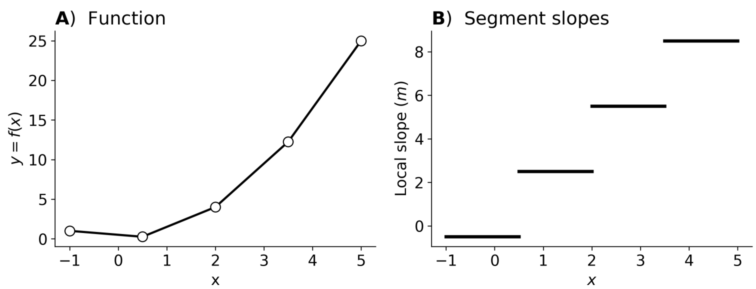

Figure 5.18: Visualization for Exercise 1.

Now that you have the code to create the figure, experiment with the N parameter. **First**, try changing N to larger numbers, e.g., 50 or 500. Pay close attention to the y-axis values in panel B: Does the range of the axis change when you change the x-axis resolution? **Second**, modify the code to remove the denominator from m calculation (that is, calculate the slope of each line segment as Δy without dividing by Δx). Again, vary the N parameter and observe what happens to the y-axis values in panel B. What conclusions can you draw from these experiments about the importance of scaling the slope by Δx?

Conceptual note: The derivative is what happens when you set the resolution of the x-axis grid to be infinitely fine. That is, as $N \to \infty$,

$\Delta x \to dx$ and the y-axis in panel B becomes the derivative.

5.2. `numpy` cannot calculate the analytic derivative of a function, because `numpy` is designed to process numbers, not symbols. However, you can use `numpy` to compute an "empirical derivative," which is a numerical approximation of a derivative. This is also called "numerical differentiation" or "numerical differences." In fact, the previous exercise involved numerically differentiating x^2.

The goal of this exercise is to explore the relation between numerical and analytic differentiation, and to understand how to plot empirical derivatives.

Implement the function $f(x) = x^3 - x^2$ in `numpy` using $N = 10$ x-axis points linearly spaced in the domain $[-1, 1]$. Then, compute two versions of the derivative of this function:

Analytic Write code that directly implements $f'(x)$ (that is, you differentiate the function on paper and then write Python code that implements the derivative function) and evaluates that analytic derivative at the same x-axis points used for the function.

Empirical Calculate $\Delta y/\Delta x$ empirically using the `np.diff` function.

Plot all three functions as in Figure 5.19. You might get an error message when plotting the empirical derivative; you'll need to figure out why that error happens and what to do about it!

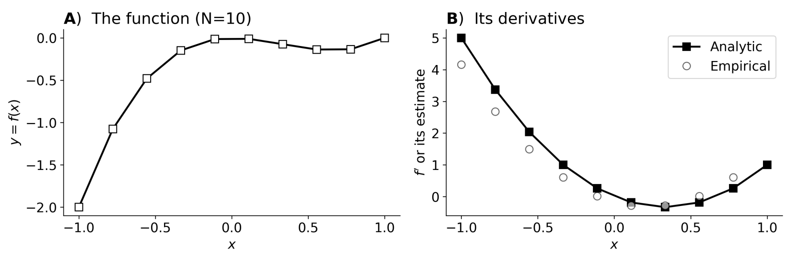

Figure 5.19: Visualization for Exercise 2.

What do you think of the two derivatives? The shapes appear largely similar, but the values don't seem to line up. Before reading the next paragraph, come up with a reason why this might be, and a possible solution.

The reason why the two derivative functions don't line up is the same underlying cause as the error message when plotting: The number of x-axis coordinate values is N whereas the number of points in the empirical derivative is $N-1$. I drew the figure using `x[:-1]`, that is, the first $N-1$ x-axis grid points. Is that the correct way to do it? You could also try `x[1:]`, which skips the first point.

The heart of the issue is this: Given a difference of $y_2 - y_1$, is the corresponding x value x_2 or x_1? The answer is that both are misaligned; they are equally correct and incorrect.

Perhaps you think that the correct solution is to use their average: $(x_1 + x_2)/2$. That average actually does *look* the best on the plot, in that it aligns well with the analytic derivative. But it is also incorrect because it means you are *linearly interpolating* the function; in other words, you are making assumptions about how the function behaves between two evaluated points, although you did not actually evaluate the function at those interpolated points. On the one hand, if the function is smooth and the x-axis resolution is high, then a linear interpolation is likely to be reasonably accurate. Still, the fact is that plotting the derivative at the boundary averages is based on assumptions. (The reason why averaging the boundaries looks most accurate on this plot is that the black lines for the analytic function are also linear interpolations; thus, the apparent close match simply reflects the fact that these are two linear interpolations of the same function.)

There are alternative formulations for numerical differentiation. Here's one, for example:

$$f' = \frac{f(x+h) - f(x-h)}{2h} \tag{5.92}$$

The idea here is to take the symmetric difference left and right of each center point x_0.

The important point is this: As h gets smaller (finer resolution of the x-axis grid), the various numerical approximation formulations converge to the analytic derivative.

All this discussion aside, you can now confirm that increasing the N value increases the match between the empirical and analytic functions. Spend a minute exploring different values of N on the resulting figure.

5.3. Did you take my advice to explore the qualitative impact of N on the visual match between the empirical and analytic functions? In this exercise, you will explore this relationship quantitatively.

RMS is a common way to measure divergences in data science and engineering.

Calculate the difference between the numerical approximation and the analytic derivatives using the root-mean-square error (RMS):

$$\mathrm{RMS}_N = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f'_{e_i} - f'_{a_i})^2} \tag{5.93}$$

N is the number of x-axis grid points, f'_{e_i} is the i^{th} value of the numerical difference (the empirical derivative), and f'_{a_i} is the analytic derivative.

The procedure described above provides an RMS for one value of N. Put the code for this procedure inside a for-loop and vary the N parameter from 10 to 300. Visualize the RMS series as a function of the x-axis points as in Figure 5.20.

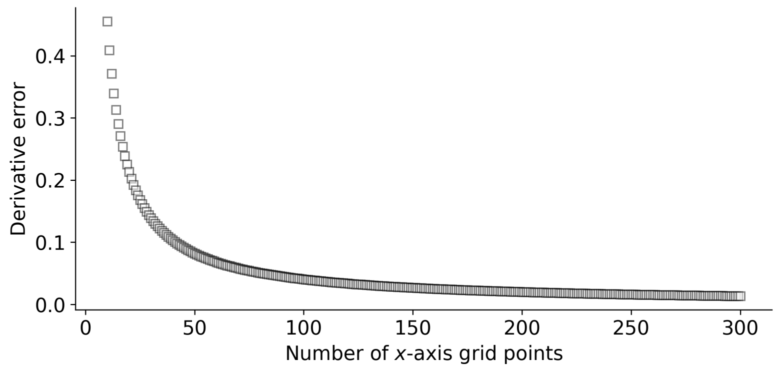

Figure 5.20: Visualization for Exercise 3.

Overall, this plot shows that the numerical approximation of the true derivative is relatively inaccurate when there are few x-axis grid points (which you saw qualitatively in the previous exercise), and becomes increasingly accurate as the number of grid points increases.

The error gets even smaller for larger N. I don't recommend running through this for-loop for really large N simply because of computation time, but you can either modify the code to skip values (e.g., N=10 to 1,000,000 in steps of 10000), or you can copy/paste the inside of the for-loop to evaluate the error for one large value of N. I implemented this latter suggestion in the online code, and found that when $N = 100000$, the RMS error is .00004 (*cf* error of .01 when $N = 300$).

The exact numerical values of this error depend on the behavior of the function within the simulated subdomain: The less the function varies, the smaller the error. You can try this by using different functions that don't change as much as the one in this exercise. Nonetheless, the overall shape of the error function — decreasing with increasing resolution with the errors tapering down quickly for large resolution — is a general expectation about the match between empirical and analytic functions.

5.4. `numpy` cannot calculate analytic (symbolic) derivatives, but `sympy` is specifically designed for these kinds of mathematical operations. The `sympy` function for differentiation is `sym.diff`. Your goal in this exercise is to figure out the correct way to use this function, and then obtain and print the derivatives of the functions shown in Figure 5.21. Make sure that the fractions 5/4 and 9/4 are represented as symbolic rational numbers, not as the floating-point numbers 1.25 and 2.25.

$$f(x) = \frac{5x}{4} + \frac{9}{4}$$

$$f'(x) = \frac{5}{4}$$

$$f(x) = x^3 - x^2$$

$$f'(x) = 3x^2 - 2x$$

$$f(x) = \sqrt{x} + \cos(x)$$

$$f'(x) = -\sin(x) + \frac{1}{2\sqrt{x}}$$

$$f(x) = e^x$$

$$f'(x) = e^x$$

Figure 5.21: Python output from Exercise 4.

Finally, use the `.subs()` method to determine both the symbolic and numerical approximations to the derivative of e^x for three values of x, and report the results as below.

$$f'(-1) = e^{-1} \approx 0.367879$$
$$f'(0) = 1 \approx 1$$
$$f'(2) = e^2 \approx 7.38906$$

Given that `sympy` can return the *true* derivative of a function, why would you ever use `numpy` to *estimate* a derivative? There are two situations in which numerical differentiation is necessary: (**1**) Not all functions can be differentiated analytically, and some theoretically differentiable functions cannot be analytically solved by `sympy` due to their complexity. (**2**) Data measured from the real world may not conform to a specific mathematical function, and thus empirical differentiation is the only option. Examples include EKG (heart rate), heat distributions, and Earthquake spread across a landscape.

5.5. The purpose of this exercise is to confirm the linearity of differentiation in `sympy` and to improve your general `sympy` skills. Here's what you should do: Create symbolic constants α and β, and differentiate the terms combined and separately as shown in the `sympy` output below.

$$\frac{d}{dx}\left[\alpha x^2 + \beta\cos(x)\right] = 2\alpha x - \beta\sin(x)$$

$$\alpha\frac{d}{dx}(x^2) + \beta\frac{d}{dx}(\cos(x)) = 2\alpha x - \beta\sin(x)$$

Figure 5.22: Screenshot of `sympy` output for Exercise 5.

The point is to see that the result is the same regardless of how the differentiation was set up.

The next part of this exercise is more about improving your `sympy` coding skills than reinforcing calculus concepts. Create a symbolic expression corresponding to the function:

$$f(x) = \alpha x^2 + \beta x^3 + \gamma e^{2x} \tag{5.94}$$

Use `sym.diff()` to calculate the derivative of the entire function and print that out. Then, figure out how to separate that expression into separate `sympy` variables corresponding to each term of that function (e.g., one variable for αx^2).

In a for-loop over each term, differentiate each term separately and print the result. Inside the for-loop, sum each piece into one expression and print out the final summed result. Your Python output should match my results in the screenshot below. If you need a hint for separating the parts of the function, see the footnote[6].

[6] `sympy` calls the terms of an expression "arguments," abbreviated as `args`.

Derivative of the entire function: $2\alpha x + 3\beta x^2 + 2\gamma e^{2x}$

Derivative of αx^2 is $2\alpha x$
Derivative of βx^3 is $3\beta x^2$
Derivative of γe^{2x} is $2\gamma e^{2x}$

Combination of individual components: $2\alpha x + 3\beta x^2 + 2\gamma e^{2x}$

Figure 5.23: Screenshot of `sympy` output for the rest of Exercise 5.

5.6. In this exercise, I will describe an algorithm to compute tangent and secant lines. Your goal will be to translate that description into Python code. You can apply this algorithm to any function and x-axis values you want, but I recommend choosing values that will match Figure 5.7.

1. **Define the function and its derivative**. These need to be regular Python functions or `lambda` expressions.

2. **Define x-axis intersection points.** Pick three values: one for the tangent line (x_t) and two for the secant line (x_{s_1} and x_{s_2}).

3. **Calculate the corresponding function values.** The result of the previous and this steps will be three (x, y) pairs (e.g., $(x_t, f(x_t))$).

4. **Calculate two slopes.** The slope of the secant line is $m_s = \Delta y / \Delta x$ for the two function intersection points, and the slope of the tangent line is $m_t = f'(x_t)$.

5. **Write a function to calculate the line values.** These should be Python functions or `lambda` expressions. The functions should return $y_s = m_s(x - x_{s_1}) + y_{s_1}$ for the secant line, and $y_t = m_t(x - x_t) + y_t$ for the tangent line.

6. **Plot the results!** Then print out the figure and send it to your Grandmother so she can be proud of you :D

5.7. Start with $f(x) = x^{10}$ and successively differentiate the function and set the function to be its derivative. Use `sympy`. Print the results as in Figure 5.24.

$f(x) = x^{10},\ f'(x) = 10x^9$

$f(x) = 10x^9,\ f'(x) = 90x^8$

$f(x) = 90x^8,\ f'(x) = 720x^7$

$f(x) = 720x^7,\ f'(x) = 5040x^6$

$f(x) = 5040x^6,\ f'(x) = 30240x^5$

$f(x) = 30240x^5,\ f'(x) = 151200x^4$

$f(x) = 151200x^4,\ f'(x) = 604800x^3$

$f(x) = 604800x^3,\ f'(x) = 1814400x^2$

$f(x) = 1814400x^2,\ f'(x) = 3628800x$

$f(x) = 3628800x,\ f'(x) = 3628800$

Figure 5.24: Screenshot of Python output from Exercise 7.

5.8. This exercise is focused on approximating derivatives using numerical differences in `numpy`. Start from the function $f(x) = x^4$, calculate the empirical derivative using `np.diff`, then the derivative of that (thus using `np.diff` again) and again until you reach the third derivative. Show the original function and three numerical derivatives as in Figure 5.25. If your result doesn't match mine, then make sure you're normalizing by Δx correctly.

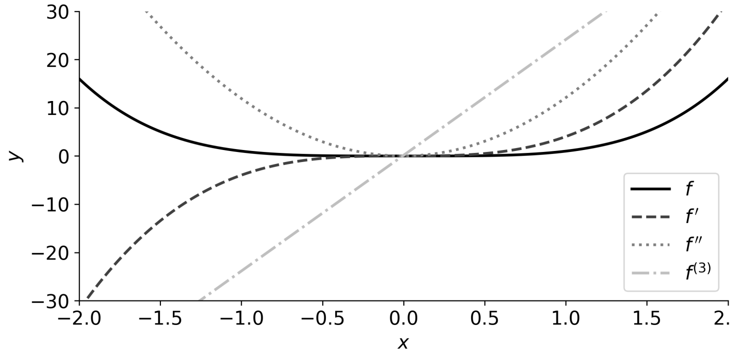

Figure 5.25: Visualization for Exercise 8.

The `np.diff` function takes an optional second input, which is the number of times to differentiate. So, `np.diff(fx,1)` is the first-order derivative (same as the default setting) and `np.diff(fx,2)` is the second-order derivative. Copy the code you wrote above into a new code block, and adjust the differentiation lines so that you compute all derivatives from the original function, not from the successive derivatives. Generate the plot again and make sure it matches perfectly. If it doesn't, think about the Δx scaling!

5.9. The goal of this exercise is to reproduce Figure 5.26. This will reinforce your understanding of trig derivatives, and will help improve your `matplotlib` skills.

The ϕ-axis grid spacing is defined as $\Delta\phi = .01$ in the interval $[-1.5\pi, 1.5\pi]$. The gray line is the derivative of cosine (approximated numerically in `numpy`) whereas the light gray circles are the minus sine function (down-sampled for visibility).

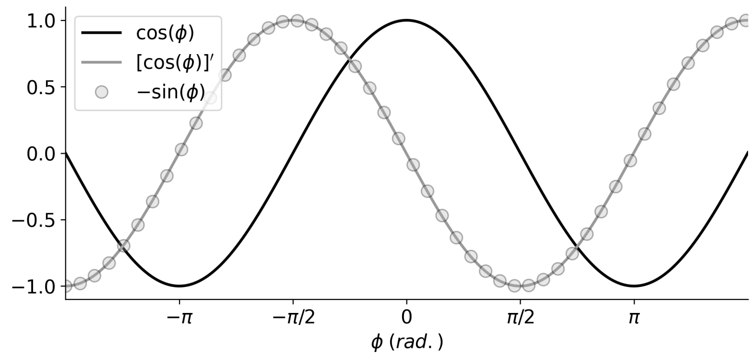

Figure 5.26: Visualization for Exercise 9.

5.10. This one's easy: reproduce the left-hand side of Figure 5.13 (page 174) in `sympy`: Differentiate $\cos(x)$ to get $-\sin(x)$, then differentiate that, and again and again until you get back to $\cos(x)$.

You can copy and adjust the code from Exercise 7, although I recommend challenging yourself and writing the code from scratch.

5.11. This exercise is just a bit of fun with plotting in `matplotlib`. Implement the following two equations in `numpy`:

$$f(x) = a \ln |x| \tag{5.95}$$

$$g(x) = \ln |x + b| \tag{5.96}$$

Then in a for-loop, vary a and b in 31 steps between -1 and 3, and draw each line as you can see in panels A and B of Figure 5.27. Finally, draw $f(x)$ on the x-axis and $g(x)$ on the y-axis as in panel C.

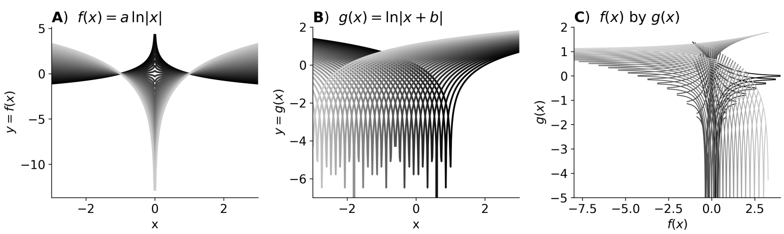

Figure 5.27: Visualization for Exercise 11.

I hope you enjoyed this exercise! Feel free to make other changes to the function definition, parameter ranges, colors, etc.