Notes working through Spivak's Calculus

Slava Akhmechet

September 26, 2024

Contents

0 Preface

I'm working through Spivak Calculus. Around the chapter on epsilon-delta limits the details get pretty confusing. I started supplementing with David Galvin's notes, which are often more clear but are still confusing. This is surprising because the topic of limits doesn't use anything beyond basic middle school math. Feels like it should be simple! And so I started writing these notes to properly understand the damned thing.

Some departures from the structure of Spivak's text:

- The very first chapter in these notes covers prerequisites necessary to study limits– some really basic limits intuitions, and material on bounding values with inequalities.
- In general each chapter in Spivak weaves between introducing concepts, exploring degenerate cases, showing examples of practice problems, and proving theorems. In my view this is delightful if you already understand the material, but distracting if you're trying to understand it for the first time. So instead I separate these categories into clear sections. I introduce concepts and proofs as quickly as possible (i.e. "the blessed path"), then have a separate section on edge cases, etc. I tend to skip and backtrack a lot through Spivak's material. The order of these notes reflects the order in which I internalized Spivak's text.
- This sometimes happens not only within a chapter, but also across chapters. Chapters 7 (Three Hard Theorems) and 8 (Least Upper Bounds) are swapped in these notes. Spivak first introduces the Intermediate Value theorem and the Extreme Value theorem as facts, then proves their consequences, then introduces completeness and its consequences, and finally proves IVT and EVT. I find it distracting and confusing. I introduce completeness and its consequences first. I then introduce and prove IVT and EVT, and finally cover their consequences. IMO this approach is much less confusing than Spivak's.
- Spivak covers variations of trigonometric functions as he goes through the book. In the early chapters I found it distracting as I didn't know any trig. I eventually buckled down and learned enough, and then revisited everything I skipped. I go through this exercise in chapter 8 of these notes.
- I break up derivatives into four chapters instead of three. The additional chapter is on the Leibniz notation. The issues of notation are sufficiently confusing that I found it difficult to study the concept of derivatives and two notational systems at the same time. Also, Leibniz notation requires considerable practice to internalize. So it gets its own chapter.

1 Limits, Part I (Prereqs)

Before we formally define limits, it helps to have a handwavy intuition for limits mechanics, and to understand inequalities. In this chapter we learn these prereqs as quickly as possible.

1.1 Handwavy limits definition

Here I only present a hand-wavy definition of limits and use it to explain the mechanics of computing limits of functions in practice. A proper definition and proofs of the theorems that make the mechanics work come in a later chapter.

A hand-wavy definition: a limit of $f(x)$ at a is the value $f(x)$ approaches close to (but not necessarily at) a.

A slightly less hand-wavy definition: let $f : \mathbf{R} \to \mathbf{R}$, let $a \in \mathbf{R}$ be some number on the x-axis, and let $l \in \mathbf{R}$ be some number on the y-axis. Then as x gets closer to a, $f(x)$ gets closer to l.

The notation for this whole thing is

$$\lim_{x \to a} f(x) = l$$

So for example $\lim_{x\to 5} x^2 = 25$ because the closer x gets to 5, the closer x^2 gets to 25 (we'll prove all this properly soon). Now suppose you have some fancy pants function like this one:

$$\lim_{x \to 0} \frac{1 - \sqrt{x}}{1 - x} \tag{1}$$

If you plot it, it's easy to see that as x approaches 0, the whole shebang approaches 1. But how do you algebraically evaluate the limit of this thing? Can you just plug 0 into the equation? It seems to work, but once we formally define limits, we'll have to prove somehow that plugging $a = 0$ into x gives us the correct result.

1.2 Limits evaluation mechanics

It turns out that it does in fact work because of a few theorems that make practical evaluation of many limits easy. Here I'll state these theorems as facts. Once I introduce the formal definition of limits in a later chapter I'll properly prove them.

1. **Constants**. $\lim_{x\to a} c = c$, where $c \in \mathbf{R}$. In other words if the function is a constant, e.g. $f(x) = 5$, then $\lim_{x\to a} f(x) = 5$ for any a.

2. **Identity**. $\lim_{x\to a} x = a$. In other words if the function is an identity function $f(x) = x$, then $\lim_{x\to 6} f(x) = 6$. Meaning we simply plug a into x.

3. **Addition**[1]. $\lim_{x\to a}(f+g)(x) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$. For example $\lim_{x\to a}(x+2) = \lim_{x\to a} x + \lim_{x\to a} 2 = a+2$.

4. **Multiplication**. $\lim_{x\to a}(f\cdot g)(x) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x)$. For example $\lim_{x\to a} 2x = \lim_{x\to a} 2 \cdot \lim_{x\to a} x = 2a$.

5. **Reciprocal**. $\lim_{x\to a}\left(\frac{1}{f}\right)(x) = \frac{1}{\lim_{x\to a} f(x)}$ when the denominator isn't zero. For example $\lim_{x\to a}\frac{1}{x} = \frac{1}{\lim_{x\to a} x} = \frac{1}{a}$ for $a \neq 0$.

To come back to 1, these theorems tells us that

$$\lim_{x\to 0}\frac{1-\sqrt{x}}{1-x} = \frac{\lim_{x\to 0} 1 - (\lim_{x\to 0} x)^{\frac{1}{2}}}{\lim_{x\to 0} 1 - \lim_{x\to 0} x} = \frac{1-0^{\frac{1}{2}}}{1-0} = 1$$

Holes

What happens if we try to take a limit as $x \to 1$ rather than $x \to 0$?

$$\lim_{x\to 1}\frac{1-\sqrt{x}}{1-x}$$

We can't use the same trick and plug in 1 because we get a nonsensical result 0/0 as the function isn't defined at 0. If we plot it, we clearly see the limit approaches 1/2 at 0, but how do we prove this algebraically? The answer is to do some trickery to find a way to cancel out the inconvenient term (in this case $1-\sqrt{x}$)

$$\lim_{x\to 1}\frac{1-\sqrt{x}}{1-x} = \lim_{x\to 1}\frac{1-\sqrt{x}}{(1-\sqrt{x})(1+\sqrt{x})} = \lim_{x\to 1}\frac{1}{1+\sqrt{x}} = \frac{1}{2}$$

Why is it ok here to divide by $1-\sqrt{x}$? Good question! Recall that the limit is defined *close to a* (or *around a*, or as *x approaches a*), but not **at** a. In other words $f(a)$ need not even be defined (as is the case here). This means that as we consider $1-\sqrt{x}$ at different values of x as it approaches a, the limit never requires us to evaluate the function at $x=a$. So we never have to consider $1-\sqrt{x}$ as $x=1$, $1-\sqrt{x}$ never takes on the value of 0, and it is safe to divide it out.

1.3 Absolute value inequalities

Consider an inequality $0 < |x-a| < \delta$. This will come up a lot soon. What does this inequality mean? The intuitive reading is that the difference between x and a is between 0 and δ. But it's a little subtle, so let's look at it carefully. There are actually two inequalities here: $0 < |x-a|$ and $|x-a| < \delta$. We should consider each separately.

[1] Spivak's book uses a slightly more verbose definition that assumes the limits of f and g exist near a, see p. 103

The left side, $0 < |x - a|$ is equivalent to $|x - a| > 0$. But $|x - a|$ is an absolute value, it's **always** true that $|x - a| \geq 0$. So this part of the inequality says $x - a \neq 0$, or $x \neq a$. I don't know why mathematicians say $0 < |x - a|$ instead of $x \neq a$, probably because confusing you brings them pleasure.

The right side is $|x - a| < \delta$. Intuitively this says that the difference between x and a should be less than δ. Put differently, x should be within δ of a. Algebraically we can write it as two cases:

1. $x - a < \delta$
2. $-(x - a) < \delta$

A little basic manipulation, and we can rewrite this as $a - \delta < x < a + \delta$.

1.4 Bounding with inequalities

We will often need to make an inequality of the following form work out:

$$|n||m| < \epsilon$$

Here ϵ is given to us, we have complete control over the upper bound of $|n|$, and $|m|$ can take on values outside our direct control. Obviously we can't make the inequality work without knowing *something* about $|m|$, so we'll try to find a bound for it in terms of other fixed values, or values we control.

For example, suppose we've discovered there is a fixed value a, and that $|m| < 3|a| + 4$. Given that we control $|n|$, how do we bound it in terms of ϵ and $|a|$ in such a way that the inequality $|n||m| < \epsilon$ holds?

Since we control $|n|$ and $(3|a| + 4)$ is fixed, we can find $|n|$ small enough so that $|n|(3|a| + 4) < \epsilon$ holds. Then certainly any inequality whose left side is smaller, e.g. $|n|(3|a| + 3) < \epsilon$, will also hold. And since $|m|$ is always smaller than $3|a| + 4$, it follows $|n||m| < \epsilon$ will hold as well.

All we have left to do is find a bound for $|n|$ such that $|n|(3|a| + 4) < \epsilon$ holds, which is of course easy:

$$|n| < \frac{\epsilon}{3|a| + 4}$$

Having bound $|n|$ in this way, we can verify that $|n|(3|a| + 4) < \epsilon$ holds by multiplying both sides of the above inequality by $3|a| + 4$.

2 Limits, Part II (Blessed Path)

2.1 Formal limits definition

Definition: $\lim_{x\to a} f(x) = L$ when for any $\epsilon \in \mathbf{R}$ there exists $\delta \in \mathbf{R}$ such that for all x, $0 < |x - a| < \delta$ implies $|f(x) - L| < \epsilon$. (Also $\epsilon > 0, \delta > 0$.)

Here is what this says. Suppose $\lim_{x\to a} f(x) = L$. You pick any interval on the y-axis around L. Make it as small (or as large) as you want. I'll produce an interval on the x-axis around a. You can take any number from my interval, plug it into f, and the output will stay within the bounds you specified.

So ϵ specifies the distance away from L along the y-axis, and δ specifies the distance away from a along the x-axis. Take any x within δ of a, plug it into f, and the result is guaranteed to be within ϵ of L. $\lim_{x\to a} f(x) = L$ just means there exists such δ for any ϵ.

Limit uniqueness

Suppose $\lim_{x\to a} f(x) = L$. It's easy to assume L is the only limit around a, but such a thing needs to be proved. We prove this here. More formally, suppose $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} f(x) = M$. We prove that $L = M$.

Suppose for contradiction $L \neq M$. Assume without loss of generality $L > M$. By limit definition, for all $\epsilon > 0$ there exists a positive $\delta \in \mathbf{R}$ such that $0 < |x - a| < \delta$ implies

- $|f(x) - L| < \epsilon \implies L - \epsilon < f(x)$
- $|f(x) - M| < \epsilon \implies f(x) < M + \epsilon$

for all x. Thus

$$\begin{aligned} &L - \epsilon < f(x) < M + \epsilon \\ &\implies L - \epsilon < M + \epsilon \\ &\implies L - M < 2\epsilon \end{aligned}$$

The above is true for all ϵ. Now let's narrow our attention and consider a concrete $\epsilon = (L - M)/4$, which we easily find leads to a contradiction[2]:

$$\begin{aligned} &L - M < 2\epsilon \\ &\implies (L - M)/4 < \epsilon/2 && \text{dividing both sides by 4} \\ &\implies \epsilon < \epsilon/2 && \text{recall we set } \epsilon = (L - M)/4 \end{aligned}$$

We have a contradiction, and so $L = M$ as desired.

[2] note we assumed $L > M$, thus $\epsilon = (L - M)/4 > 0$

Half-Value Neighborhood Lemma

This lemma will come in handy later, so we may as well prove it now. Suppose $M \neq 0$ and $\lim_{x\to a} g(x) = M$. We show that there exists some δ such that $0 < |x - a| < \delta$ implies $|g(x)| \geq |M|/2$ for all x.

Intuitively, the lemma states the following: when a function g approaches a nonzero limit M near a point, there exists an interval in which the values of g are closer to M than to zero.

Proof. The claim that $|g(x)| \geq |M|/2$ is equivalent to

$$g(x) \leq -|M|/2 \quad \text{or} \quad g(x) \geq |M|/2$$

There are two possibilities: either $M > 0$ or $M < 0$. Let's consider each possibility separately.

Case 1. Suppose $M > 0$. Then to show $|g(x)| \geq |M|/2$ it is sufficient to show *either* $g(x) \leq -M/2$ or $g(x) \geq M/2$. We will show $g(x) \geq M/2$. Fix $\epsilon = M/2$. By limit definition there is some δ such that $0 < |x - a| < \delta$ implies for all x

$$\begin{aligned}
&|g(x) - M| < M/2 && \\
&\implies -M/2 < g(x) - M && \\
&\implies M/2 < g(x) && \text{add } M \text{ to both sides} \\
&\implies g(x) > M/2 && \text{note} \geq \text{is correct but not tight}
\end{aligned}$$

Case 2. Suppose $M < 0$. We must show either $g(x) \leq M/2$ or $g(x) \geq -M/2$. We will show $g(x) \leq M/2$. Fix $\epsilon = -M/2$. Then

$$\begin{aligned}
&|g(x) - M| < -M/2 && \\
&\implies g(x) - M < -M/2 && \\
&\implies g(x) < M/2 && \text{add } M \text{ to both sides;} \\
& && \text{note} \leq \text{is correct but not tight}
\end{aligned}$$

QED.

2.2 Evaluation mechanics proofs

Armed with the formal definition, we can use it to rigorously prove the five theorems useful for evaluating limits (constants, identity, addition, multiplication, reciprocal). Let's do that now.

Constants

Let $f(x) = c$. We prove that $\lim_{x\to a} f(x) = c$ for all a.

Let $\epsilon > 0$ be given. Pick any positive δ. Then for all x such that $0 < |x - a| < \delta$, $|f(x) - c| = |c - c| = 0 < \epsilon$. QED.

(Note that we can pick any positive $\delta > 0$, e.g. $1, 10, \frac{1}{10}$.)

Identity

Let $f(x) = x$. We prove that $\lim_{x\to a} f(x) = a$ for all a.

Let $\epsilon > 0$ be given. We need to find $\delta > 0$ such that for all x in $0 < |x - a| < \delta$, $|f(x) - a| = |x - a| < \epsilon$. I.e. we need to find a δ such that $|x - a| < \delta$ implies $|x - a| < \epsilon$. This obviously works for any $\delta \leq \epsilon$. QED.

(Note the many options for δ, e.g. $\delta = \epsilon$, $\delta = \frac{\epsilon}{2}$, etc.)

Addition

Let $f, g \in \mathbf{R} \to \mathbf{R}$. We prove that

$$\lim_{x\to a}(f + g)(x) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$$

Let $L_f = \lim_{x\to a} f(x)$ and let $L_g = \lim_{x\to a} g(x)$. Let $\epsilon > 0$ be given. We must show there exists $\delta > 0$ such that for all x bounded by $0 < |x - a| < \delta$ the following inequality holds:

$$|(f + g)(x) - (L_f + L_g)| < \epsilon$$

I.e. we're trying to show $\lim_{x\to a}(f+g)(x)$ equals to L_f+L_g, the sum of the other two limits. Let's convert the left side of this inequality into a more convenient form:

$$\begin{aligned}
|(f + g)(x) - (L_f + L_g)| &= |f(x) + g(x) - (L_f + L_g)| \\
&= |(f(x) - L_f) + (g(x) - L_g)| \\
&\leq |(f(x) - L_f)| + |(g(x) - L_g)| \quad \text{by triangle inequality}
\end{aligned}$$

By limit definition there exist positive δ_f, δ_g such that for all x

- $0 < |x - a| < \delta_f$ implies $|f(x) - L_f| < \epsilon/2$
- $0 < |x - a| < \delta_g$ implies $|g(x) - L_g| < \epsilon/2$

Recall that we can make ϵ as small as we like. Here we pick deltas for $\epsilon/2$ because it's convenient to make the equations work, as you will see in a second. For all x bounded by $0 < |x - a| < \min(\delta_f, \delta_g)$ we have

$$|(f(x) - L_f)| < \epsilon/2 \quad \text{and} \quad |(g(x) - L_g)| < \epsilon/2$$

Fix $\delta = \min(\delta_f, \delta_g)$. Then for all x bounded by $0 < |x - a| < \delta$ we have

$$
\begin{aligned}
|(f+g)(x)-(L_f+L_g)| &\le |(f(x)-L_f)|+|(g(x)-L_g)| \\
&< \epsilon/2+\epsilon/2=\epsilon
\end{aligned}
$$

as desired.

Multiplication

Let $f, g \in \mathbf{R} \to \mathbf{R}$. We prove that

$$
\lim_{x\to a}(fg)(x) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x)
$$

Let $L_f = \lim_{x\to a} f(x)$ and let $L_g = \lim_{x\to a} g(x)$. Let $\epsilon > 0$ be given. We must show there exists $\delta > 0$ such that for all x bounded by $0 < |x-a| < \delta$ the following inequality holds:

$$
|(fg)(x) - (L_f L_g)| < \epsilon
$$

(i.e. we're trying to show $\lim_{x\to a}(fg)(x)$ equals to $L_f L_g$, the product of the other two limits.) Let's convert the left side of this inequality into a more convenient form:

$$
\begin{aligned}
|(fg)(x)-(L_fL_g)| &= |f(x)g(x)-L_fL_g| \\
&= |f(x)g(x)-L_fg(x)+L_fg(x)-L_fL_g| \\
&= |g(x)(f(x)-L_f)+L_f(g(x)-L_g)| \\
&\le |g(x)(f(x)-L_f)|+|L_f(g(x)-L_g)| && \text{by triangle inequality} \\
&= |g(x)||f(x)-L_f|+|L_f||g(x)-L_g| && \text{in general } |ab|=|a||b|
\end{aligned}
$$

We now need to show there exists δ such that $0 < |x-a| < \delta$ implies

$$
|g(x)||f(x)-L_f|+|L_f||g(x)-L_g| < \epsilon
$$

We will do that by finding δ such that

1. $|g(x)||f(x)-L_f| < \epsilon/2$
2. $|L_f||g(x)-L_g| < \epsilon/2$

First, we show $|g(x)||f(x)-L_f| < \epsilon/2$.

By limit definition we can find δ_1 to make $|f(x)-L_f|$ as small as we like. But how small? To make $|g(x)||f(x)-L_f| < \epsilon/2$ we must find a delta such that $|f(x)-L_f| < \epsilon/2g(x)$. But to do that we need to get a bound on $g(x)$. Fortunately we know there exists δ_2 such that $|g(x)-L_g| < 1$ (we pick 1 because we must pick some bound, and 1 is as good as any). Thus $|g(x)| < |L_g|+1$. And so, we can pick δ_1 such that $|f(x)-L_f| < \epsilon/2(|L_g|+1)$.

Second, we show $|L_f||g(x) - L_g| < \epsilon/2$.

That is easy. By limit definition there exists a δ_3 such that $0 < |x - a| < \delta_3$ implies $|g(x) - L_g| < \epsilon/2|L_f|$ for all x. Actually, we need a δ_3 such that $0 < |x - a| < \delta_3$ implies $|g(x) - L_g| < \frac{\epsilon}{2(|L_f|+1)}$ for all x to avoid divide by zero, and of course that exists too.

Fix $\delta = \min(\delta_1, \delta_2, \delta_3)$. Now

$$\begin{aligned} |(fg)(x) - (L_f L_g)| &\leq |g(x)||f(x) - L_f| + |L_f||g(x) - L_g| \\ &< e/2 + e/2 = e \end{aligned}$$

as desired.

Reciprocal

Let $\lim_{x \to a} f(x) = L$. We prove $\lim_{x \to a} \left(\frac{1}{f} \right)(x) = 1/L$ when $L \neq 0$.

First we show $\frac{1}{f}$ is defined near a. By half-value neighborhood lemma (see 2.1) there exists δ_1 such that $0 < |x - a| < \delta_1$ implies $|f(x)| \geq |L|/2$ where $L \neq 0$. Therefore $f(x) \neq 0$ near a, and thus $\frac{1}{f}$ near a is defined.

Now all we must do is find a delta such that $\left| \frac{1}{f}(x) - \frac{1}{L} \right| < \epsilon$. Let's make the equation more convenient:

$$\begin{aligned} \left| \frac{1}{f}(x) - \frac{1}{L} \right| &= \left| \frac{1}{f(x)} - \frac{1}{L} \right| \\ &= \left| \frac{L - f(x)}{Lf(x)} \right| \\ &= \frac{|f(x) - L|}{|L||f(x)|} \\ &= \frac{|f(x) - L|}{|L|} \cdot \frac{1}{|f(x)|} \end{aligned}$$

Above we showed there exists δ_1 such that $0 < |x - a| < \delta_1$ implies $|f(x)| \geq |L|/2$. Raising both sides to -1 we get $|\frac{1}{f(x)}| \leq \frac{2}{|L|}$. Continuing the chain of reasoning above we get

$$\begin{aligned} \frac{|f(x) - L|}{|L|} \cdot \frac{1}{|f(x)|} &\leq \frac{|f(x) - L|}{|L|} \cdot \frac{2}{|L|} \\ &= \frac{2}{|L|^2} |f(x) - L| \end{aligned}$$

(if you're confused about why this inequality works, left-multiply both sides of $|\frac{1}{f(x)}| \leq \frac{2}{|L|}$ by $\frac{|f(x)-L|}{|L|}$.) Thus we must find δ_2 such that

$$\frac{2}{|L|^2}|f(x) - L| < \epsilon$$

That is easy. Since $\lim_{x \to a} f(x) = L$ we can make $|f(x) - L|$ as small as we like. Dividing both sides by $\frac{2}{|L|^2}$, we must make $|f(x) - L| < \frac{|L|^2\epsilon}{2}$. Thus we must fix $\delta = \min(\delta_1, \delta_2)$. QED.

2.3 Low-level proofs

While high level theorems allow us to easily compute complicated limits, it's instructive to compute a few limits for complicated functions straight from the definition. We do that here.

Limits of quadratic functions

We will prove directly from the limits definition that $\lim_{x \to a} x^2 = a^2$. Let $\epsilon > 0$ be given. We must show there exists δ such that $|x^2 - a^2| < \epsilon$ for all x in $0 < |x - a| < \delta$.

Observe that

$$|x^2 - a^2| = |(x - a)(x + a)| = |x - a||x + a|$$

Thus we must pick δ such that $|x - a||x + a| < \epsilon$. Since $0 < |x - a| < \delta$, picking δ conveniently happens to bound $|x - a|$, letting us make it as small as we want. But to know how small, we need to find an upper bound on $|x + a|$. We can do it as follows.

Pick an arbitrary $\delta = 1$ (we may pick any arbitrary delta, e.g. 1/10, 10, etc.) Then since $|x - a| < \delta$:

$$\begin{aligned}
&|x - a| < 1 \\
&\implies -1 < x - a < 1 \\
&\implies 2a - 1 < x + a < 2a + 1 && \text{add } 2a \text{ to both sides}
\end{aligned}$$

We now have a bound on $x + a$, but we need one on $|x + a|$. It's easy to see $|x + a| < \max(|2a - 1|, |2a + 1|)$. By triangle inequality ($|a + b| \leq |a| + |b|$):

$$\begin{aligned}
|2a - 1| \leq |2a| + |-1| = |2a| + 1 \\
|2a + 1| \leq |2a| + |1| = |2a| + 1
\end{aligned}$$

Thus $|x + a| < |2a| + 1$, provided $|x - a| < 1$. Coming back to our original goal, $|x - a||x + a| < \epsilon$ when

- $|x - a| < 1$ and
- $|x - a| < \frac{\epsilon}{|2a|+1}$

Putting these together, $\delta = \min(1, \frac{\epsilon}{|2a|+1})$.

Limits of fractions

We will prove directly from the limits definition that $\lim_{x\to 2}\frac{3}{x} = \frac{3}{2}$. Let $\epsilon > 0$ be given. We must show there exists $\delta > 0$ such that $|\frac{3}{x} - \frac{3}{2}| < \epsilon$ for all x in $0 < |x-2| < \delta$.

Let's manipulate $|\frac{3}{x} - \frac{3}{2}|$ to make it more convenient:

$$\left|\frac{3}{x} - \frac{3}{2}\right| = \left|\frac{6-3x}{2x}\right| = \frac{3}{2}\frac{|x-2|}{|x|}$$

Thus we need to find δ such that

$$\begin{aligned} &\frac{3}{2}\frac{|x-2|}{|x|} < \epsilon \\ &\implies \frac{|x-2|}{|x|} < \frac{2\epsilon}{3} \end{aligned}$$

Conveniently $0 < |x-2| < \delta$ bounds $|x-2|$. But now we need to find a bound for $|x|$. It would be extra convenient if we could show $|x| > 1$. Then we could set $\delta = \frac{2\epsilon}{3}$ (and thus bound $|x-2| < \frac{2\epsilon}{3}$). A denominator greater than 1 would only make the fraction smaller than $\frac{2\epsilon}{3}$, ensuring $\frac{|x-2|}{|x|} < \frac{2\epsilon}{3}$ holds.

We will do exactly that. Pick an arbitrary $\delta = 1$ (we may pick any arbitrary delta, e.g. 1/10, 10, etc.) Then since $|x-2| < \delta$

$$\begin{aligned} &|x-2| < 1 \\ &\implies -1 < x-2 < 1 \\ &\implies 1 < x < 3 \\ &\implies 1 < |x| < 3 \end{aligned}$$

Yes!! Luckily $\delta = 1$ implies $|x| > 1$! Thus, provided that $|x-2| < 1$ and $|x-2| < \frac{2\epsilon}{3}$, the inequality $|\frac{3}{x} - \frac{3}{2}| < \epsilon$ holds. Putting the two constraints together, we get $\delta = \min(1, \frac{2\epsilon}{3})$.

3 Limits, Part III (Edge Cases)

3.1 Absence of limits

What does it mean to say L is not a limit of $f(x)$ at a? It flows out of the definition– there exist some ϵ such that for any δ there exists an x in $0 < |x - a| < \delta$ such that $|f(x) - L| \geq \epsilon$.

A stronger version is to say there is no limit of $f(x)$ at a. To do that we must prove that *any* L is not a limit of $f(x)$ at a.

Example: Absolute value fraction

Consider $f(x) = \frac{x}{|x|}$. It's easy to see that

$$f(x) = \begin{cases} -1 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases}$$

We will show there is no limit of $f(x)$ near 0.

Weak version. First, let's prove a weak version– that $\lim_{x \to 0} f(x) \neq 0$. That is easy. Pick some reasonably small epsilon, say $\epsilon = \frac{1}{10}$. We must show that for any δ there exists an x in $0 < |x - a| < \delta$ such that $|f(x) - 0| \geq \frac{1}{10}$.

Let's pick some arbitrary x out of our permitted interval, say $x = \delta/2$. Then

$$|f(x) - 0| = |f(\delta/2)| = \left|\frac{\delta/2}{|\delta/2|}\right| = 1 \geq \frac{1}{10}$$

Strong version. Now we prove that $\lim_{x \to 0} f(x) \neq L$ for *any* L. Sticking with $\epsilon = \frac{1}{10}$ we proceed as follows.

If $L < 0$ take $x = \delta/2$. Then

$$|f(x) - L| = |f(\delta/2) - L| = \left|\frac{\delta/2}{|\delta/2|} - L\right| = |1 - L| > \frac{1}{10}$$

Similarly if $L \geq 0$ take $x = -\delta/2$. Then

$$|f(x) - L| = |f(-\delta/2) - L| = \left|\frac{-\delta/2}{|-\delta/2|} - L\right| = |-1 - L| > \frac{1}{10}$$

Example: Dirichlet function

The *dirichlet* function f is defined as follows:

$$f(x) = \begin{cases} 1 & \text{for rational } x, \\ 0 & \text{for irrational } x. \end{cases}$$

We prove $\lim_{x \to a} f(x)$ does not exist for any a.

Proof. Let $\epsilon = \frac{1}{10}$. Suppose for contradiction there exists L such that $\lim_{x\to a} f(x) = L$. There are two possibilities: either $L \le \frac{1}{2}$ or $L > \frac{1}{2}$.

First suppose $L \le \frac{1}{2}$. Pick any rational x from the interval $0 < |x-a| < \delta$. Then $|f(x) - L| = |1 - L| \ge \frac{1}{2}$. Thus $|f(x) - L| \ge \frac{1}{10}$.

Similarly, suppose $L > \frac{1}{2}$. Pick any irrational x from the interval $0 < |x-a| < \delta$. Then $|f(x) - L| = |0 - L| > \frac{1}{2}$. Thus $|f(x) - L| \ge \frac{1}{10}$.

Thus $\lim_{x\to a} f(x)$ does not exist for any a, as desired.

3.2 One-sided limits

We have seen that the following function has no limit approaching 0:

$$f(x) = \begin{cases} -1 & x < 0, \\ 1 & x > 0 \end{cases}$$

However, f has properties around 0 we may want to be able to formally describe. First, intuitively f approaches -1 as we approach zero from the left (from "below"). Not surprisingly, a notation for this exists:

$$\lim_{x\to 0^-} f(x) = -1$$

If we take $l = -1$, this notation compiles down to the following definition. For every $\epsilon > 0$ there exists $\delta > 0$ such that $0 < a - x < \delta$ implies $|f(x) - l| < \epsilon$ for all x. This is our usual limit definition, except instead of looking at both sides of a we say $x < a$ (i.e. we look from left of a).

Second, intuitively f approaches 11 as we approach zero from the right (from "above"). The notation for this is:

$$\lim_{x\to 0^+} f(x) = 1$$

If we take $l = 1$, the definition is as follows. For every $\epsilon > 0$ there exists $\delta > 0$ such that $0 < x - a < \delta$ implies $|f(x) - l| < \epsilon$ for all x. Again, this is our usual limit definition, except instead of looking at both sides of a we say $x > a$ (i.e. we look from right of a).

3.3 Limits at infinity

Consider the function $f(x) = \frac{1}{x}$. Clearly as x gets very large, $f(x)$ trends toward zero. Again, we have a notation that encodes this property of f:

$$\lim_{x\to\infty} \frac{1}{x} = 0$$

Take $l = 0$, and this compiles down to the following definition. For every $\epsilon > 0$ there is a number N such that $|f(x) - l| < \epsilon$ for all $x > N$.

Intuitively, for any ϵ, $f(x)$ will get within ϵ of the limit for x large enough. Here we simply produce a large enough N instead of δ.

3.4 Infinite limits

Consider the function $f(x) = \frac{1}{x^2}$. Near zero f shoots up, and again, we want to be able to encode that. The notation for this property is

$$\lim_{x \to 0} f(x) = \infty$$

This compiles down to the following definition. Given any $M > 0$ there exists $\delta > 0$ such that $0 < |x - a| < \delta$ implies $f(x) > M$ for all x. Intuitively, given an arbitrarily large $f(x) = M$ we can produce a bound on the x-axis, within which $f(x)$ is never smaller than M.

Example. Suppose we want to prove $\lim_{x \to 0} \frac{1}{x^2} = \infty$. Let $M > 0$ be given. We must produce $\delta > 0$ such that $0 < |x| < \delta$ implies $\frac{1}{x^2} > M$ for all x. Suppose we fix $|x| < \frac{1}{\sqrt{M}}$. Then:

$$\begin{aligned}
& |x| < \frac{1}{\sqrt{M}} && \text{note } M > 0 \\
\implies & x^2 < \frac{1}{M} \\
\implies & \frac{1}{x^2} > M
\end{aligned}$$

Thus $\delta \le \frac{1}{\sqrt{M}}$ implies $\frac{1}{x^2} > M$ as desired.

4 Continuity, Part I (On a Point)

4.1 Definition of continuity

A function f is **continuous** at a when

$$\lim_{x \to a} f(x) = f(a)$$

Inlining the limits definition, f is continuous at a if for all $\epsilon > 0$ there exists $\delta > 0$ such that $0 < |x - a| < \delta$ implies $|f(x) - f(a)| < \epsilon$.

We can simplify this definition slightly. Observe that in continuous functions $f(a)$ exists, and at $x = a$ we get $f(x) - f(a) = 0$. Thus we can relax the constraint $0 < |x - a| < \delta$ to $|x - a| < \delta$.

A function f is **continuous on an interval** (a, b) if it's continuous at all $c \in (a, b)$[3].

Nonzero Neighborhood Lemma

Armed with these definitions we can extend the half-value neighborhood lemma (see 2.1) in a useful way. The *nonzero neighborhood lemma* will come in handy when we prove the intermediate value theorem (see 6.1), so we may as well prove the lemma now.

Suppose f is continuous at a, and $f(a) \neq 0$. Then there exists $\delta > 0$ such that:

1. if $f(a) < 0$ then $f(x) < 0$ for all x in $|x - a| < \delta$.
2. if $f(a) > 0$ then $f(x) > 0$ for all x in $|x - a| < \delta$.

Intuitively the lemma states that there is some interval around a on which $f(x) \neq 0$ and has the same sign as $f(a)$.

Proof. The proof follows trivially from the half-value neighborhood lemma.

4.2 Recognizing continuous functions

The following theorems allow us to tell at a glance that large classes of functions are continuous (e.g. polynomials, rational functions, etc.)

Five easy proofs

Constants. Let $f(x) = c$. Then f is continuous at all a because

$$\lim_{x \to a} f(x) = c = f(a)$$

[3]Closed intervals are a tiny bit harder, and I'm keeping them out for brevity.

Identity. Let $f(x) = x$. Then f is continuous at all a because

$$\lim_{x\to a} f(x) = a = f(a)$$

Addition. Let $f, g \in \mathbf{R} \to \mathbf{R}$ be continuous at a. Then $f + g$ is continuous at a because

$$\lim_{x\to a}(f+g)(x) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x) = f(a) + g(a) = (f+g)(a)$$

Multiplication. Let $f, g \in \mathbf{R} \to \mathbf{R}$ be continuous at a. Then $f \cdot g$ is continuous at a because

$$\lim_{x\to a}(fg)(x) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x) = f(a) \cdot g(a) = (fg)(a)$$

Reciprocal. Let g be continuous at a. Then $\frac{1}{g}$ is continuous at a where $g(a) \neq 0$ because

$$\lim_{x\to a}\left(\frac{1}{g}\right)(x) = \frac{1}{\lim_{x\to a} g(x)} = \frac{1}{g(a)} = \left(\frac{1}{g}\right)(a)$$

Slightly harder proof: composition

Let $f, g \in \mathbf{R} \to \mathbf{R}$. Let g be continuous at a, and let f be continuous at $g(a)$. Then $f \circ g$ is continuous at a. Put differently, we want to show

$$\lim_{x\to a}(f \circ g)(x) = (f \circ g)(a)$$

Unpacking the definitions, let $\epsilon > 0$ be given. We want to show there exists $\delta > 0$ such that $|x - a| < \delta$ implies

$$\begin{aligned} &|(f \circ g)(x) - (f \circ g)(a)| \\ &\qquad = |f(g(x)) - f(g(a))| < \epsilon \end{aligned}$$

By problem statement we have two continuities.

First, f is continuous at $g(a)$, i.e. $\lim_{X\to g(a)} f(X) = f(g(a))$. Thus there exists $\delta' > 0$ such that $|X - g(a)| < \delta'$ implies $|f(X) - f(g(a))| < \epsilon$.

Second, g is continuous at a, i.e. $\lim_{x\to a} g(x) = g(a)$. Thus there exists $\delta > 0$ such that $|x - a| < \delta$ implies $|g(x) - g(a)| < \epsilon$. Since we can make ϵ be anything, we can set it to δ'.

I.e. there exists $\delta > 0$ such that $|x - a| < \delta$ implies $|g(x) - g(a)| < \delta'$. Intuitively, $g(x)$ is close to $g(a)$. But by the first continuity, any X close to $g(a)$ implies

$$|f(X) - f(g(a))| < \epsilon$$

Thus $|f(g(x)) - f(g(a))| < \epsilon$, as desired.

4.3 Example: Stars over Babylon

Stars over Babylon is a modification of the Dirichlet function (see 3.1), defined as follows:

$$f(x) = \begin{cases} 0, & x \text{ irrational}, 0 < x < 1 \\ 1/q, & x = p/q \text{ in lowest terms}, 0 < x < 1. \end{cases}$$

Claim: for $0 < a < 1$, $\lim_{x\to a} f(x) = 0$.

Proof. Let $\epsilon > 0$ be given. We must find $\delta > 0$ such that $0 < |x - a| < \delta$ implies $|f(x) - 0| < \epsilon$. For *any* $\delta > 0$, $0 < |x - a| < \delta$ implies one of two cases for all x: either x is irrational or it is rational.

If x is irrational, $|f(x) - 0| = 0 < \epsilon$.

Otherwise, if $x = p/q$ in the lowest terms is rational, $f(x) = 1/q$. Let $n \in \mathcal{N}$ such that $1/n < \epsilon$. We will look for δ such that:

$$f\left(\frac{p}{q}\right) = \frac{1}{q} < \frac{1}{n} < \epsilon$$

Observe that when $q > n$, $f(\frac{p}{q}) = \frac{1}{q} < \frac{1}{n}$. Thus the only rationals that *could* result in $f(\frac{p}{q}) \geq 1/n$ are ones where $q \leq n$:

$$A = \{\frac{1}{2};\quad \frac{1}{3}, \frac{2}{3};\quad \frac{1}{4}, \frac{3}{4};\quad \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5},\quad \ldots,\quad \frac{1}{n}, \ldots, \frac{n-1}{n}\}$$

This set has a finite length, and thus *one* $p/q \in A$ is closest to a. Fix $\delta = |a - p/q|$ (i.e. anything less than this distance). This guarantees $0 < |x - a| < \delta$ implies $x \notin A$ for all x, and thus $f(x) < 1/n < \epsilon$ for all x, as desired.

Claim: $f(x)$ is continuous at all irrationals, discontinuous at all rationals.

Proof: we've just proven for $0 < a < 1$, $\lim_{x\to a} f(x) = 0$. By definition $f(x)$ is zero for all irrationals, and nonzero for all rationals. Thus $\lim_{x\to a} f(x) = f(x)$ for all irrationals, and $\lim_{x\to a} f(x) \neq f(x)$ for all rationals.

5 Complete ordered fields

5.1 Motivation

The twelve ordered field axioms are sufficient to define limits, continuity, and prove all the theorems in the previous sections. Since the set $\mathcal{Q}$ of rational numbers is an ordered field[4], rationals have been sufficient for the work we've done so far. However, we are about to start proving slightly more sophisticated theorems about continous functions, and ordered fields will quickly start breaking our intuitions.

For example, consider the function $f(x) = x^2 - 2$ (a parabola shifted down two units). It's easy to see f is a continuous function, and thus our intuition is that we should be able to draw it without "lifting the tip of the pencil off the sheet of paper". Upon reflection however, it becomes obvious that in the universe limited to ordered fields this is impossible. f intersects the x-axis when $x^2 = 2$, but every high school student knows $\sqrt{2} \notin \mathcal{Q}$ (see 5.1 for proof). Thus there is no $x \in \mathcal{Q}$ such that $f(x) = 0$. And since $\mathcal{Q}$ is an ordered field, it follows ordered fields alone aren't sufficient to resolve this problem.

The *intermediate value theorem* (see 6.1) formalizes the claim that a continuous function segment that starts below the x-axis and ends above the x-axis intersects the x-axis. But as we can see from the example above, this is not possible to prove with ordered field axioms alone. So before we proceed with further study of continuity, we need one more axiom called *the completeness axiom*, which we introduce in this chapter.

Combined with the twelve ordered field axioms, the completeness axiom forms *complete ordered fields*. These objects are sufficient to proceed with our study of calculus. We will see that rational numbers $\mathcal{Q}$ are not a complete ordered field, whereas real numbers $\mathcal{R}$ are.[5] Thus from here $\mathcal{R}$-valued functions will become our primary object of study.

Aside: sqrt(2) is irrational

Suppose $\sqrt{2} \in \mathcal{Q}$. Then there exist $a, b \in \mathcal{N}$ such that $\left(\frac{a}{b}\right)^2 = 2$. Assume a, b have no common divisor (since we can obviously keep simplifying until this is the case). Observe that both a and b cannot be even, otherwise we could simplify further.

Now we have $a^2 = 2b^2$. Thus a^2 is even, a must be even[6], and there exists $k \in \mathcal{N}$ such that $a = 2k$. Then $a^2 = 4k^2 = 2b^2$ so $2k^2 = b^2$. Thus b^2 is even and

[4] The proof is straightforward, so I'm not including it here.

[5] Proof that $\mathcal{R}$ is a complete ordered field requires construction of $\mathcal{R}$, which doesn't happen in Spivak until the last chapters. Thus I will not be delving into that here and ask the reader (i.e., currently myself) to take this on faith.

[6] Even numbers have even squares because $(2k)^2 = 4k^2 = 2 \cdot (2k^2)$

so b is even. Since both a and b cannot be even, this is a contradiction. Thus $\sqrt{2} \notin \mathcal{Q}$ as desired.

5.2 Least Upper Bound

Definition: b is an **upper bound** for S if $s \leq b$ for all $s \in S$.

For example:

- Any $b \geq 1$ is an upper bound for $S = \{x : 0 \leq x < 1\}$. E.g. $1, 2, 10$ are all upper bounds of S.
- By convention, *every* number is an upper bound for $\emptyset$.
- The set $\mathcal{N}$ of natural numbers has no natural upper bound. The proof is easy. Suppose $b \in \mathcal{N}$ is an upper bound for $\mathcal{N}$. But $b+1 \in \mathcal{N}$, and $b+1 > b$, which is a contradiction. Thus b isn't an upper bound for $\mathcal{N}$.[7]

Definition: x is a **least upper bound** of A, if

1. x is an upper bound of A,
2. *and* if y is an upper bound of A, then $x \leq y$.

A set can have only one least upper bound. The proof is easy. Suppose x and x' are both least upper bounds of S. Then $x \leq x'$ and $x' \leq x$. Thus $x = x'$. Consequently, we can use a convenient notation $\sup A$ to denote the least upper bound of A.

Obligatory examples:

- Let $S = \{x : 0 \leq x < 1\}$. Then $\sup S = 1$.
- By convention, the empty set $\emptyset$ has no least upper bound.

5.3 Completeness axiom

We are now ready to state the completeness axiom.

Completeness [P13]: If A is a non-empty set of numbers that has an upper bound, then it has a least upper bound.

Claim: rational numbers are not complete.
Proof: Let $C = \{x : x^2 < 2 \text{ and } x \in \mathcal{Q}\}$. Suppose for contradiction rational numbers are complete. Then there exists $b \in \mathcal{Q}$ such that $b = \sup C$. Observe that

[7] We need to do a little more work to show $\mathcal{N}$ has no upper bound, natural or not. Be patient! We will prove this by the end of the section.

- $b^2 \neq 2$ as that would imply $b = \sqrt{2}$ and thus $b \notin \mathcal{Q}$.
- $b^2 \not< 2$ as there would exist some $x \in C$ such that $b^2 < x^2 < 2$. Thus $b < x$ and b is not the upper bound.

Therefore $b^2 > 2$. But this implies there exists some $x \in \mathcal{Q}$ such that $2 < x^2 < b^2$. Thus x is greater than every element in C, and $x < b$. So b is not the *least* upper bound. We have a contradiction, therefore rational numbers are not complete, as desired.

Claim: completeness cannot be derived from ordered fields.
Proof: $\mathcal{Q}$ is not complete and $\mathcal{Q}$ is an ordered field. Thus completeness is not a property of ordered fields.

Claim: real numbers are complete.
Proof [deferred]: The completeness property can be derived from the construction of real numbers $\mathcal{R}$, which makes reals a **complete ordered field**. The proof requires we study the actual construction of $\mathcal{R}$, which Spivak leaves until the last chapters. Thus for the moment the proof will be taken on faith. In any case, it is better to build calculus upon abstract complete ordered fields than upon concrete real numbers.

5.4 Consequences of completeness

$\mathcal{N}$ is not bounded above

We've shown $\mathcal{N}$ has no upper bound in $\mathcal{N}$. Now we show $\mathcal{N}$ has no upper bound in $\mathcal{R}$.

Suppose for contradiction $\mathcal{N}$ has an upper bound. Since $\mathcal{N} \neq \emptyset$ then by completeness $\mathcal{N}$ has a least upper bound. Let $\alpha = \sup \mathcal{N}$. Then:

$$\begin{aligned} & \alpha \geq n \text{ for all } n \in \mathcal{N} \\ \implies & \alpha \geq n+1 \text{ for all } n \in \mathcal{N} && \text{since } n+1 \in \mathcal{N} \text{ if } n \in \mathcal{N} \\ \implies & \alpha - 1 \geq n \text{ for all } n \in \mathcal{N} \end{aligned}$$

Thus $\alpha - 1$ is *also* an upper bound for $\mathcal{N}$. This contradicts that $\alpha = \sup \mathcal{N}$. Therefore $\mathcal{N}$ is not bounded above, as desired.

$\sqrt{2}$ exists

We show $\sqrt{2} \in \mathcal{R}$. Let $S = \{y \in \mathcal{R} : y^2 < 2\}$. Obviously S is non-empty and has an upper bound. Thus by completeness property it has a least upper bound. Let $x = \sup S$. Note that $1 \in S$ and 2 is an upper bound of S. Thus $1 \leq x \leq 2$. We show $x^2 = 2$ by showing $x^2 \not< 2$ and $x^2 \not> 2$.

Case 1. Suppose for contradiction $x^2 < 2$. Let $0 < \epsilon < 1$ be a small number. Then

$$\begin{aligned}
(x+\epsilon)^2 =& x^2 + 2\epsilon x + \epsilon^2 & \\
&\leq x^2 + 4\epsilon + \epsilon && \text{since } x < 2 \text{ and } \epsilon < 1 \\
&= x^2 + 5\epsilon < 2 && \text{since } x^2 < 2 \text{ (by supposition), we can pick} \\
& && \text{a small enough } \epsilon \text{ to make this true}
\end{aligned}$$

Thus there exists ϵ such that $(x+\epsilon)^2 < 2$. By definition of S it follows $x+\epsilon \in S$, which contradics that x is the least upper bound. Therefore $x^2 \nless 2$

Case 2. Suppose for contradiction $x^2 > 2$. Let $0 < \epsilon < 1$ be a small number. Then

$$\begin{aligned}
(x-\epsilon)^2 =& x^2 - 2\epsilon x + \epsilon^2 & \\
&\geq x^2 - 2\epsilon x && \text{since } \epsilon^2 > 0 \\
&\geq x^2 - 4\epsilon && \text{since } x \leq 2 \\
&> 2 && \text{since } x^2 > 2 \text{ (by supposition), we can pick} \\
& && \text{a small enough } \epsilon \text{ to make this true}
\end{aligned}$$

Thus $(x-\epsilon)^2 > 2$, which by definition of S implies $x - \epsilon > y$ for all $y \in S$. So $x - \epsilon$ is an upper bound of S. We have a contradiction– since $x - \epsilon < x$, it follows x is not a least upper bound. Therefore $x^2 \ngtr 2$ as desired.

Since $x^2 \nless 2$ and $x^2 \ngtr 2$, it follows $x^2 = 2$ as desired.

Archimedean property

Handwavy definition: the Archimedean property states that you can fill the universe with tiny grains of sand.

Formal defition: let $\epsilon > 0$ be small and let $r > 0$ be large. Then there exists $n \in \mathcal{N}$ such that $n\epsilon > r$.

Proof: suppose for contradiction the property is false. Then there exist ϵ, r such that for all $n \in \mathcal{N}$, $n\epsilon \leq r$. Therefore $n \leq \frac{r}{\epsilon}$. This implies $\mathcal{N}$ is bounded, which a contradiction.

A useful special case is when $r = 1$. In this case the Archimedean property can be restated as follows. Let $\epsilon > 0$ be small. Then there exists $n \in \mathcal{N}$ such that $n\epsilon > 1$. Put differently, there exists $n \in \mathcal{N}$ such that $\frac{1}{n} < \epsilon$.

A few more notes on the Archimedean property:

- Obviously the Archimedean property follows from completeness, as shown above.

- The Archimedean property is true in $\mathcal{Q}$ and can be proven without being assumed[8].
- Completeness does not follow from the Archimedean property. The proof is easy: the Archimedean property holds on $\mathcal{Q}$, and we know $\mathcal{Q}$ is not complete as shown above.

Density

Let $x, y \in \mathcal{R}$. Then S is a **dense subset** of $\mathcal{R}$ if there is an element of S in (x, y). Put differently, there is an element of S between any two points in $\mathcal{R}$.

- Obviously $\mathcal{R}$ is a dense subset of itself (if $x, y \in \mathcal{R}$ then $\frac{x+y}{2} \in (x, y)$).
- Integers are not a dense subset of $\mathcal{R}$. E.g. there is no integer between 1.1 and 1.9.
- The set of positive numbers $\{x : x \in \mathcal{R}, x > 0\}$ is not a dense subset of $\mathcal{R}$. E.g. there is no positive number between -2 and -1.

Claim: the set of rational numbers $\mathcal{Q}$ is dense.
Proof: let $x, y \in \mathcal{R}$ be given. Suppose we can show there exists a rational in (x, y) for $0 \leq x < y$. Then:

- Given $x < y \leq 0$, there is a rational r in $(-y, -x)$. So $-r$ is in (x, y).
- Given $x < 0 < y$, there is a rational r in $(0, y)$. So r is of course also in (x, y).

Thus all we must do is prove there exists a rational in (x, y) for $0 \leq x < y$.

Let $0 \leq x < y$ be given. By the Archimedean property there exists $n \in \mathcal{N}$ such that $\frac{1}{n} < y - x$. Because (a) $\mathcal{N}$ is unbounded and (b) $\mathcal{N}$ is well-ordered, there exists the least integer $m \in \mathcal{N}$ such that $m \geq ny$.

First, observe that

$$\begin{aligned} & m - 1 < ny && \text{or } m \text{ wouldn't be the } \textit{least} \text{ integer } m \geq ny \\ \implies & \frac{m-1}{n} < y \end{aligned}$$

[8] Excluding the proof here, but it's fairly simple

Second, suppose for contradiction $\frac{m-1}{n} \leq x$. Then

$$\begin{aligned}
&\frac{m-1}{n} \leq x \\
&\implies \frac{m}{n} - \frac{1}{n} \leq x \\
&\implies -\frac{1}{n} \leq x - \frac{m}{n} \\
&\implies \frac{1}{n} \geq \frac{m}{n} - x \\
&\implies \frac{1}{n} \geq y - x \qquad\qquad \text{recall } m \geq ny \text{, thus } \frac{m}{n} \geq y
\end{aligned}$$

This is a contradiction, thus $\frac{m-1}{n} > x$.

Therefore $\frac{m-1}{n} \in (x, y)$ as desired.

Claim: the set of irrational numbers $\mathcal{R} \setminus \mathcal{Q}$ is dense.
Proof: let $x, y \in \mathcal{R}$ be given. By density of the rationals there exists $r \in \mathcal{Q}$ such that $\frac{x}{\sqrt{2}} < r < \frac{y}{\sqrt{2}}$. Multiplying each side by $\sqrt{2}$, we get $x < \sqrt{2}r < y$. We know $\sqrt{2}r$ is irrational. Thus there exists an irrational number between any two numbers in $\mathcal{R}$, and the set of irrationals $\mathcal{R} \setminus \mathcal{Q}$ is dense as desired.

6 Continuity, Part II (On an Interval)

6.1 Intermediate Value Theorem

Theorem: if f is continuous on $[a, b]$ and $f(a) < 0 < f(b)$, then there exists $x \in [a, b]$ such that $f(x) = 0$.

Or intuitively, if $f(a)$ is below zero and $f(b)$ is above zero, f must cross the x-axis somewhere.

Proof: intuitively, we will locate the smallest number x on the x-axis where $f(x)$ first crosses from negative to positive, and show that $f(x)$ must be zero.

First, we define a set A that contains all inputs to f before f crosses from negative to positive for the first time:

$$A = \{x : a \leq x \leq b, \text{and } f \text{ is negative on the interval } [a, x]\}$$

We know $A \neq \emptyset$ since $a \in A$, and b is an upper bound of A. Thus A has a least upper bound α such that $a \leq \alpha \leq b$. By nonzero neighborhood lemma (see 4.1) we know there is some interval around a on which f is negative, and some interval around b on which f is positive. Thus we can further refine the bound on α to $a < \alpha < b$.

We now show $f(\alpha) = 0$ by eliminating the possibilities $f(\alpha) < 0$ and $f(\alpha) > 0$.

Case 1. Suppose for contradiction $f(\alpha) < 0$. By nonzero neighborhood lemma there exists $\delta > 0$ such $|x - \alpha| < \delta$ implies $f(x) < 0$ for all x. But that means numbers in $(\alpha - \delta, \alpha + \delta)$ are in A. E.g. $(\alpha + \delta/2) \in A$. Since $\alpha + \delta/2 > \alpha$, α is not an upper bound of A, and is thus not the least upper bound.

Case 2. Suppose for contradiction $f(\alpha) > 0$. By nonzero neighborhood lemma there exists $\delta > 0$ such $|x - \alpha| < \delta$ implies $f(x) > 0$ for all x. But that means numbers in $(\alpha - \delta, \alpha + \delta)$ are *not* in A, and there exist many upper bounds of A less than α. E.g. $\alpha - \delta/2$ is an upper bound of A, and since $\alpha - \delta/2 < \alpha$, α is not the *least* upper bound.

Both cases lead to contradiction, therefore $f(\alpha) = 0$. QED.

IVT generalization

The intermediate value theorem is usually presented in a more general way. If f is continuous on $[a, b]$ and $f(a) < c < f(b)$ or $f(a) > c > f(b)$ then there is some x in $[a, b]$ such that $f(x) = c$.

Intuitively, f takes on any value between $f(a)$ and $f(b)$ at some point in the interval $[a, b]$.

Proof. This trivially follows from the the theorem as initially stated. There are two cases:

Case 1: $f(a) < c < f(b)$. Let $g = f - c$. Then g is continuous and $g(a) < 0 < g(b)$. Thus there is some x in $[a, b]$ such that $g(x) = 0$. But that means $f(x) = c$.

Case 2: $f(a) > c > f(b)$. Observe that $-f$ is continuous on $[a, b]$ and $-f(a) < -c < -f(b)$. By case 1 there is some x in $[a, b]$ such that $-f(x) = -c$, which means $f(x) = c$.

QED.

6.2 Boundedness theorem

The boundedness theorem states that if f is continuous on $[a, b]$, then f is bounded above (i.e. f lies below some line). Before we prove this, we first prove a simple lemma.

Bounded neighborhood lemma: if f is continuous at a, then there is $\delta > 0$ such that f is bounded above on the interval $(a - \delta, a + \delta)$.

Intuitively, if f is continuous at a then there is some interval around a on which f is bounded above.

Proof: The proof is trivial. Inlining the definition of continuity, for any $\epsilon > 0$ there exists $\delta > 0$ such that $|x - a| < \delta$ implies $|f(x) - f(a)| < \epsilon$ for all x. Thus $f(a) + \epsilon$ is the upper bound on f within $(a - \delta, a + \delta)$, as desired.

(Note that we can pick any ϵ to concretize the proof, for example $\epsilon = 1$.)

Boundedness theorem: if f is continuous on $[a, b]$, then f is bounded above on $[a, b]$. I.e. there is some numbers N such that $f(x) \leq N$ for all x in $[a, b]$.

Proof: intuitively, we will try to find the smallest number x on the x-axis where $f(x)$ becomes unbounded above, and discover that there is no such number in $[a, b]$.

First, we define a set A that contains all inputs to f before f stops being bounded above:

$$A = \{x : a \leq x \leq b, \text{ and } f \text{ is bounded above on } [a, x]\}$$

By bounded neighborhood lemma f is bounded above in the neighborhood of a[9]. Thus we know $A \neq \emptyset$ because $a \in A$. Further, b is an upper bound of A. Thus A has a least upper bound.

[9] We are being sloppy here as we actually need a left-sided and right-sided version of the bounded neighborhood lemma. I am papering over this for now, but will need to fix at some point by giving proper one sided proofs

Let $\alpha = \sup A$. To prove the boundedness theorem we must prove two claims:

1. $\alpha = b$, i.e. f does not ever stop being bounded above before b.
2. $(\alpha = b) \in A$, as $\sup A$ is not necessarily a member of A.

First, we prove $\alpha = b$. Suppose for contradiction $\alpha < b$. By bounded neighborhood lemma there is some $\delta > 0$ such that f is bounded above in $(\alpha - \delta, \alpha + \delta)$. But that means there are many upper bounds greater than α, for example $\alpha + \delta/2$. Thus α is not the *least* upper bound. We have a contradiction, and so $\alpha = b$.

Second, we prove $(\alpha = b) \in A$. By bounded neighborhood lemma there is some $\delta > 0$ such that f is bounded above in $(b - \delta, b]$. Pick any x_0 such that $b - \delta < x_0 < b$. Then:

- $x_0 < b = \alpha$. Since α is the least upper bound it follows $x_0 \in A$. Thus f is bounded above on $[a, x_0]$.
- f is bounded above on $[x_0, b]$.

Since f is bounded above on $[a, x_0]$ and on $[x_0, b]$, it follows f is bounded above on $[a, b]$ as desired. QED.

Boundedness theorem generalization

The *boundedness theorem* is usually presented slightly more generally: it proves f is bounded above *and* below. We already proved the former. Put more formally, the latter states:

If f is continuous on $[a, b]$, then f is bounded *below* on $[a, b]$. I.e. there is some numbers N such that $f(x) \geq N$ for all x in $[a, b]$.

Proof: observe that $-f$ is continuous on $[a, b]$. By claim 2 there exists a number M such that $-f(x) \leq M$ for all x in $[a, b]$. But that means $f(x) \geq -M$ for all x in $[a, b]$. QED.

6.3 Extreme Value Theorem

The *extreme value theorem* states that is f is continuous on $[a, b]$, then f attains its maximum on $[a, b]$. To see why we need the extreme value theorem, consider $f = \frac{1}{x}$. f is discontinuous at 0 and approaches infinity. Thus f does not attain a maximum value on the interval $[0, 1]$.

Extreme value theorem: If f is continuous on $[a, b]$, then there is some number y in $[a, b]$ such that $f(y) \geq f(x)$ for all x in $[a, b]$.

Proof: Let A be the set of f's outputs on $[a, b]$:

$$A = \{f(x) : x \text{ in } [a, b]\}$$

Since $[a, b]$ isn't empty, $A \neq \emptyset$. By boundedness theorem, f is bounded on $[a, b]$, and so A has an upper bound. Thus A has a least upper bound. Let $\alpha = \sup A$. By definition $\alpha \geq f(x)$ for x in $[a, b]$. Thus it suffices to show $\alpha \in A$ (i.e. $\alpha = f(y)$ for some y in $[a, b]$).

Let's consider a function g[10]:

$$g = \frac{1}{\alpha - f(x)}, \quad x \text{ in } \ [a, b]$$

Suppose for contradiction $\alpha \notin A$. Then the denominator is never zero and g is continuous. Therefore:

$$\begin{aligned}
&\frac{1}{\alpha - f(x)} < M && \text{by boundedness theorem} \\
& && \text{for some bound } M \\
&\implies \alpha - f(x) > \frac{1}{M} && \text{take reciprocal} \\
&\implies -f(x) > \frac{1}{M} - \alpha && \\
&\implies f(x) < \alpha - \frac{1}{M} && \text{times } -1
\end{aligned}$$

But this contradicts that α is the *least* upper bound. Thus $\alpha \in A$ as desired. QED.

EVT generalization

The extreme value theorem is usually presented slightly more generally: a continuous f attains both its maximum and its minimum. We already proved the former. Put more formally, the latter states:

If f is continuous on $[a, b]$, then there is some number y in $[a, b]$ such that $f(y) \leq f(x)$ for all x in $[a, b]$.

Proof: Observe that $-f$ is continous on $[a, b]$. By claim 3 there is some y in $[a, b]$ such that $-f(y) \geq -f(x)$ for all x in $[a, b]$. But that means that $f(y) \leq f(x)$ for all x in $[a, b]$. QED.

6.4 IVT and EVT consequences

Claim 1a: Every positive number has a square root. I.e. if $\alpha > 0$, then there is some number x such that $x^2 = \alpha$.

Proof: Consider the function $f(x) = x^2$. If f takes on the value of α as its output, then $x = \sqrt{\alpha}$ is the input (i.e. $x^2 = \alpha$). Thus all we must show is that f takes on the value of α.

[10] g is a bit of a rabbit pulled out of a magic hat, but to quote a great British statesman, them's the breaks

We can do it as follows. Show there exist a, b such that $f(a) < \alpha < f(b)$. Since f is continuous, by intermediate value theorem there exists x such that $f(x) = \alpha$. So, let's find a and b:

- First, find a such that $f(a) < \alpha$. Observe that $f(0) = 0 < \alpha$, thus fix $a = 0$.
- Second, find b such that $\alpha < f(b)$.
  - If $\alpha < 1$ then $f(1) = 1 > \alpha$. Thus fix $b = 1$.
  - If $\alpha > 1$ then $f(\alpha) = \alpha^2 > \alpha$. Thus fix $b = \alpha$.

By intermediate value theorem, there is some x in $[0, b]$ such that $f(x) = \alpha$. QED.

Claim 1b: Every positive number has an nth root. I.e. if $\alpha > 0$, then there is some number x such that $x^n = \alpha$.

Proof: We can use the exact same argument as 1a, just consider $f(x) = x^n$.

Claim 1c: Let n be odd. Then every number has an nth root. I.e. there is some number x such that $x^n = \alpha$ for all α.

Proof: This is also easy:

- *Case* $\alpha > 0$. By claim 2b, there is an x such that $x^n = \alpha$.
- *Case* $\alpha < 0$. By claim 2b, there is an x such that $x^n = -\alpha$. Then $(-x)^n = \alpha$.

QED.

Claim 2: If n is odd, then any equation of the form

$$x^n + a_{n-1}x^{n-1} + \ldots + a_0 = 0$$

has a root.

Proof: Let $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0$. Here is an intuitive outline of the proof:

1. We will show that f must take on negative and positive values. Thus by the intermediate value theorem, there exists some x such that $f(x) = 0$.
2. To do that we will show that as $|x|$ gets large, x^n completely dominates other terms. (This is obvious if you consider Big-Oh of each term.)
3. Since n is odd, x^n takes on a negative value when x is negative, and a positive value when x is positive. And since x^n dominates other terms, when x is sufficiently large, f takes on both negative and positive values.

We must find a way to bound the magnitude of $a_{n-1}x^{n-1} + \ldots + a_0$ to show that for large enough x, it's smaller than the magnitude of x^n. This way we guarantee $f(x)$ has the same sign as x^n. This is trivial to do by adopting Big-Oh notation, but both math books I looked at do it the old-fashioned way, so we will too.

Let's start with some obvious transformations we can make:

$$\begin{aligned} |a_{n-1}x^{n-1} + \ldots + a_0| &\leq |a_{n-1}x^{n-1}| + \ldots + |a_0| && \text{by triangle inequality} \\ &= |a_{n-1}||x^{n-1}| + \ldots + |a_0| && \text{in general } |ab| = |a||b| \end{aligned}$$

If we only consider behavior of f on large x (i.e. when $|x| > 1$), we can further bound the expression. Observe that when $|x| > 1$ then $x^{n-1} > x^{n-2} > \ldots > x > 1$. Therefore:

$$\begin{aligned} |a_{n-1}x^{n-1} + \ldots + a_0| &\leq |a_{n-1}||x^{n-1}| + \ldots + |a_0| \\ &\leq |a_{n-1}||x^{n-1}| + \ldots + |a_0||x^{n-1}| \\ &= x^{n-1}(|a_{n-1}| + \ldots + |a_0|) \end{aligned}$$

Let $M = |a_{n-1}| + \ldots + |a_0| + 1$, i.e. a bound on the sum of the coefficients, plus a little extra to ensure $M > 1$. Then

$$\begin{aligned} |a_{n-1}x^{n-1} + \ldots + a_0| &\leq x^{n-1}(|a_{n-1}| + \ldots + |a_0|) \\ &< M|x^{n-1}| \end{aligned}$$

Given this bound it follows that for all $|x| > 1$:

$$x^n - M|x^{n-1}| < x^n + (a_{n-1}x^{n-1} + \ldots + a_0) < x^n + M|x^{n-1}|$$

or put differently:

$$x^n - M|x^{n-1}| < f(x) < x^n + M|x^{n-1}|$$

We will now find x_1 and x_2 such that $f(x_1) < 0$ and $f(x_2) > 0$. Let $x_1 = -2M$ (note that x_1 satisfies our condition $|x_1| > 1$ since $M > 1$). Then for all $x \leq x_1$:

$$\begin{aligned} f(x) &< x^n + M|x^{n-1}| \\ &= x^n + Mx^{n-1} && n \text{ is odd, thus } n-1 \text{ is even, thus } x^{n-1} > 0 \\ &= x^{n-1}(x + M) && \text{factor out } x^{n-1} \\ &\leq -2^{n-1}M^n && \text{substitute } -2M \text{ and simplify} \\ &< 0 \end{aligned}$$

Similarly, let $x_2 = 2M$. Then for all $x \geq x_2$:

$$\begin{aligned} f(x) &> x^n - M|x^{n-1}| \\ &= x^n - Mx^{n-1} \\ &= x^{n-1}(x - M) \\ &\geq 2^{n-1}M^n \\ &> 0 \end{aligned}$$

QED.

Claim 3: If n is even and $f(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0$, then there is a number y such that $f(y) \leq f(x)$ for all x.

Intuitively, even degree polynomials achieve their minimum on $\mathcal{R}$ because when you zoom out enough they are U-shaped (consider the graph $f(x) = x^2$ as a simple example).

Proof: It's easy to intuitively see why the claim makes sense. x^n dominates the rest of the terms when x is very large. Since n is even, $x^n > 0$. Thus on very large $|x|$ the graph shoots up (i.e. it has a U shape).

Here is the outline for our proof:

1. Observe that $f(0) = a_0$.
2. We will prove f is U-shaped by proving there exist two points:
   - $x_0 < 0$ such that $f(x) > a_0$ on $(-\infty, x_0]$.
   - $x_1 > 0$ such that $f(x) > a_0$ on $[x_1, \infty]$.
3. By extreme value theorem f achieves a minimum m on $[x_0, x_1]$. Note $m \leq a_0$ (otherwise it wouldn't be a minimum).
4. Thus f achieves a minimum m on $\mathcal{R}$, as we've shown that outside $[x_0, x_1]$, $f(x) > a_0$ (and thus $f(x) > m$) for all x.

All we must do now to complete the proof is find $x_0 < 0 < x_1$. Let $M = |a_{n-1}| + \ldots + |a_0| + 1$, i.e. a bound on the sum of the coefficients, plus a little extra to ensure $M > 1$. In Claim 2 we discovered that for $|x| > 1$

$$x^n - M|x^{n-1}| < f(x) < x^n + M|x^{n-1}|$$

Let $x_1 = -2M$. Note that x_1 satisfies our condition $|x_1| > 1$ since $M > 1$. Then for all $x < x_1$:

$$\begin{aligned} f(x) &> x^n - M|x^{n-1}| \\ &= x^n + Mx^{n-1} && x \text{ is negative, and } n-1 \text{ is odd} \\ &= x^{n-1}(x + M) \\ &\geq 2^{n-1}M^n && \text{substitute } -2M \text{ and simplify} \end{aligned}$$

Similarly let $x_2 = 2M$. Then for all $x > x_1$:

$$\begin{aligned}
f(x) &> x^n - M|x^{n-1}| \\
&= x^n + Mx^{n-1} && x \text{ is positive} \\
&= x^{n-1}(x+M) \\
&\geq 2^{n-1}M^n && \text{substitute } 2M \text{ and simplify}
\end{aligned}$$

Since $M > 1$ we have

$$2^{n-1}M^n \geq M \geq |a_n + 1| \geq a_n + 1 > a_n$$

Therefore for all $x < x_1$ and $x > x_2$, $f(x) > a_n$ as desired.

Claim 4: Consider the equation

$$x^n + a_{n-1}x^{n-1} + \ldots + a_0 = c$$

and suppose n is even. Then there is a number m such that the equation has a solution for $c \geq m$ and has no solution for $c < m$.

Proof: In claim 3 we saw that even degree polynomials achieve a minimum. Let that be m. There are three cases:

- If $c < m$ there is no solution, as the polynomial doesn't take on values less than m.
- If $c = m$ there is a solution, as the polynomial obviously takes on the value m (by claim 3).
- Suppose $c > m$. Let $y, z \in \mathcal{R}$ such that $f(y) = m$ and $z > y, f(z) > c$.[11] Then $f(y) = m < c < f(z)$. By intermediate value theorem there is a number k in $[y, z]$ such that $f(k) = c$.

QED.

6.5 Uniform continuity

TODO (skipping until it comes up in Spivak or I hit it in Galvin's notes)

[11] Technically we have to prove such a z exists, but somehow Spivak rolls right past this.

7 Derivatives, Part I (Fundamentals)

7.1 Formal definitions

Definition: the *derivative* at a of a function f, denoted $f'(a)$, is defined as:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

There are *three intuitions* to convey about the derivative:

- *Algebraic interpretation.* The derivative tells how $f(a+h)$, the value of f as small distance from a, changes relative to $f(a)$ as h becomes very small.
- *Geometric interpretation.* Draw a line through points $(a, f(a))$ and $(a + h, f(a+h))$ for some small h. Then make h "infinitely small". Our $f'(a)$ is the slope of that line. The tagent line is a *linear approximation* of f near a.
- *Physics interpretation.* Suppose $f(t)$ maps time to position of a car on a road (or of any object on a straight line). Suppose you want to know the average velocity between any two points in time t_1, t_2. If $h = t_2 - t_1$ then the average velocity is $\frac{f(t_1+h)-f(t_1)}{h}$. This is the quotient of the derivative! Only with h "reduced to an infinitesimal", producing *instantanious velocity.*

Definition: f is called *differentiable* at a if the limit $f'(a)$ exists.

The notation $f'(a)$ suggests f' is a function. Indeed, we define f' as follows. Its domain is the set of all numbers a where f is differentiable, and its value at such a point a is the limit above. Not surprisingly, we call f' the *derivative* of f. Note that the domain of f' could be much smaller than the domain of f.

We can apply the definition of the derivative to f' yielding the *second derivative* $(f')'$, denoted f'' or $f^{(2)}$. The domain of f'' is all points a such that f' is differentiable at a. If $f''(a)$ exists, we say f is *twice differentiable* at a.

7.2 Differentiability implies continuity

We are about to prove an important theorem– that differentiability implies continuity. To do that, we begin with a convenient (simple) lemma.

Lemma: $\lim_{x \to a} f(x)$ is equivalent to $\lim_{h \to 0} f(a+h)$.
Proof. Let $x = a + h$. Then

$$\begin{aligned} &\lim_{h \to 0} f(a+h) \\ &\iff \lim_{x-a \to 0} f(x) \\ &\iff \lim_{x \to a} f(x) \end{aligned}$$

Note that the last implication is true because

$$0 < |(x-a) - 0| < \delta \iff 0 < |x-a| < \delta$$

QED.

Theorem: if f is differentiable at a, then f is continuous at a.
Proof. We must show that:

$$\lim_{x \to a} f(x) = f(a)$$

We begin as follows:

$$\begin{aligned}
\lim_{h \to 0}[f(a+h) - f(a)] &= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \cdot h \\
&= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \cdot \lim_{h \to 0} h \\
&= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \cdot 0 \\
&= 0
\end{aligned}$$

It follows that

$$\begin{aligned}
&\lim_{h \to 0}[f(a+h) - f(a)] = 0 \\
&\implies \lim_{h \to 0} f(a+h) - \lim_{h \to 0} f(a) = 0 \\
&\implies \lim_{h \to 0} f(a+h) = \lim_{h \to 0} f(a) = f(a)
\end{aligned}$$

7.3 Low-level proofs

In the next chapter we prove theorems that make finding derivatives for many classes of functions easy. But for now we show four low-level derivations directly from the definition. Here we will be looking at constant functions, linear functions, quadratic, and cubic functions.

Constant functions

Let $f(x) = c$. Then:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = \lim_{h \to 0} \frac{c - c}{h} = 0$$

Thus f is differentiable at a for every number a, and $f'(a) = 0$.

Linear functions

Let $f(x) = cx + d$. Then:

$$\begin{aligned} f'(a) &= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \\ &= \lim_{h \to 0} \frac{c(a+h) + d - (ca+d)}{h} \\ &= \lim_{h \to 0} \frac{ch}{h} = c \end{aligned}$$

Thus f is differentiable at a for every number a, and $f'(a) = c$.

Quadratic functions

Let $f(x) = x^2$. Then:

$$\begin{aligned} f'(a) &= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \\ &= \lim_{h \to 0} \frac{(a+h)^2 - a^2}{h} \\ &= \lim_{h \to 0} \frac{a^2 + 2ah + h^2 - a^2}{h} \\ &= \lim_{h \to 0} \frac{2ah + h^2}{h} \\ &= \lim_{h \to 0} 2a + h \\ &= \lim_{h \to 0} 2a \end{aligned}$$

Thus f is differentiable at a for every number a, and $f'(a) = 2a$.

Cubic functions

Let $f(x) = x^3$. Then:

$$\begin{aligned} f'(a) &= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \\ &= \lim_{h \to 0} \frac{(a+h)^3 - a^3}{h} \\ &= \lim_{h \to 0} \frac{a^3 + 3a^2h + 3ah^2 + h^3 - a^3}{h} \\ &= \lim_{h \to 0} \frac{3a^2h + 3ah^2 + h^3}{h} \\ &= \lim_{h \to 0} 3a^2 + 3ah + h^2 \\ &= 3a^2 \end{aligned}$$

Thus f is differentiable at a for every number a, and $f'(a) = 3a^2$.

7.4 Non-differentiability

Continuous functions are "nice". Functions that are differentiable everywhere are "nicer". Functions that are differentiable everywhere and whose first derivative is differentiable everywhere are nicer still. Thus to fully understand the derivative we must understand examples where it does not exist.

We now turn our attention to functions that aren't differentiable at some points a. We first look at four simple examples where there isn't everywhere a first derivative. We then turn our attention to a more subtle example– a function that's differentiable in the first, but not everywhere in the second derivative.

First derivative

Example 1
Let $f(x) = |x|$. Consider $f'(0)$:

$$f'(0) = \lim_{h \to 0} \frac{f(0+h) - f(0)}{h} = \lim_{h \to 0} \frac{|h|}{h}$$

Observe that $\lim_{h \to 0^+} \frac{|h|}{h} = 1$ and $\lim_{h \to 0^-} \frac{|h|}{h} = -1$. This $\lim_{h \to 0} \frac{|h|}{h}$ does not exist, and f is not differentiable at 0. Note that f is differentiable at every other point: $f'(a) = -1$ for $a < 0$ and $f'(a) = -1$ for $a > 0$.

Example 2
Let f be defined as follows:

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

Now consider $f'(0)$:

$$f'(0) = \lim_{h \to 0} \frac{f(0+h) - f(0)}{h} = \lim_{h \to 0} \frac{f(h)}{h}$$

Observe that

$$\frac{f(h)}{h} = \begin{cases} \frac{h^2}{h} = h, & h \leq 0 \\ \frac{h}{h} = 1, & h \geq 0 \end{cases}$$

Therefore $\lim_{h \to 0^-} \frac{f(h)}{h} = 0$ and $\lim_{h \to 0^+} \frac{f(h)}{h} = 1$. Thus $\lim_{h \to 0} \frac{f(h)}{h}$ does not exist, and f is not differentiable at 0.

Example 3
Let $f(x) = \sqrt{|x|}$. Consider $f'(0)$:

$$f'(0) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = \lim_{h \to 0} \frac{\sqrt{|h|}}{h}$$

Observe that

$$\frac{\sqrt{|h|}}{h} = \begin{cases} \frac{\sqrt{-h}}{h} = -\frac{1}{\sqrt{-h}}, & h < 0 \\ \frac{\sqrt{h}}{h} = \frac{1}{\sqrt{h}}, & h > 0 \end{cases}$$

Therefore $\lim_{h\to 0^+} \frac{\sqrt{|h|}}{h} = \infty$ and $\lim_{h\to 0^-} \frac{\sqrt{|h|}}{h} = -\infty$. Thus $\lim_{h\to 0} \frac{\sqrt{|h|}}{h}$ does not exist, and f is not differentiable at 0.

Example 4
Let $f(x) = \sqrt[3]{x}$. Here $\frac{\sqrt[3]{h}}{h}$ plays out as follows:

$$\frac{\sqrt[3]{h}}{h} = \frac{h^{1/3}}{h} = \frac{1}{h^{2/3}} = \frac{1}{\left(\sqrt[3]{h}\right)^2}$$

This expression becomes arbitrarily large as h goes to 0, i.e. $\lim_{h\to 0} \frac{1}{\left(\sqrt[3]{h}\right)^2} = \infty$. Thus f is not differentiable at zero (or put differently, the tagent line to f at 0 is vertical).

Second derivative

We now come to our more subtle example– a function that's differentiable in the first but not everywhere in the second derivative:

Example 1
Let

$$f(x) = \begin{cases} x^2, & x \geq 0 \\ -x^2, & x \leq 0 \end{cases}$$

As we've seen in the quadratic functions example above, $\frac{dx^2}{dx} = 2x$. By very similar logic, $\frac{d(-x^2)}{dx} = -2x$. Thus $f'(a) = 2a$ if $a \geq 0$, and $f'(a) = -2a$ if $a \leq 0$. Or, put differently, $f'(x) = 2|x|$.

So, $f'(0) = 0$. But what about $f''(0)$? We've already seen that $g(x) = |x|$ is not differentiable, and by very similar logical, $f''(0)$ does not exist! So even a "smooth looking" function may not have a second derivative– a fact that implies existence of a second derivative is a strong critereon for a function to satisfy.

7.5 Tangent lines

Spivak now handles a question—- how many times does a tagent line to f at a intersect the graph of f? He doesn't yet motivate the question, though I suspect he will at a later time (at which point I will come back and make a note of it here). To answer this question we must first find the linear equation that describes the tagent line, which we will now do. We'll then address the question of intersection for quadratic and cubic functions.

Point-slope form

The slope m of a line is determined by

$$m = \frac{y - y_1}{x - x_1}$$

For a function f differentiable at a, the slope of a tagent line at a is $m = f'(a)$. We also know one of the points on the line– $(a, f(a))$ (the point where the tangent line intersects with f). Plugging that in we get

$$\begin{aligned} f'(a) &= \frac{y - f(a)}{x - a} \\ &\implies y - f(a) = f'(a)(x - a) \\ &\implies y = f'(a)(x - a) + f(a) \end{aligned}$$

This gives us a linear equation for the tangent line– a *linear approximation* of f near a.

Intersections

Quadratic functions. For $f(x) = x^2$ we've seen that $f'(a) = 2a$. Plugging that into $y = f'(a)(x - a) + f(a)$ we get:

$$\begin{aligned} y &= 2a(x - a) + a^2 \\ &= 2ax - a^2 \end{aligned}$$

Let $g(x) = 2ax - a^2$. We can now solve for all x such that $f(x) = g(x)$:

$$\begin{aligned} &f(x) = g(x) \\ &\implies x^2 = 2ax - a^2 \\ &\implies x^2 - 2ax + a^2 = 0 \\ &\implies (x - a)^2 = 0 \\ &\implies x = a \end{aligned}$$

So the only solution is $x = a$, therefore the only point of intersection is $(a, f(a) = g(a) = a^2)$.

Cubic functions. For $f(x) = x^3$ we've seen that $f'(a) = 3a^2$. Plugging that into $y = f'(a)(x - a) + f(a)$ we get:

$$\begin{aligned} y &= 3a^2(x - a) + a^3 \\ &= 3a^2x - 2a^3 \end{aligned}$$

Let $g(x) = 3a^2x - 2a^3$. We can now solve for all x such that $f(x) = g(x)$:

$$\begin{aligned} &f(x) = g(x) \\ &x^3 = 3a^2x - 2a^3 \\ &x^3 - 3a^2x + 2a^3 = 0 \end{aligned}$$

Factoring this in Mathematica, we get

$$(a - x)^2(2a + x) = 0$$

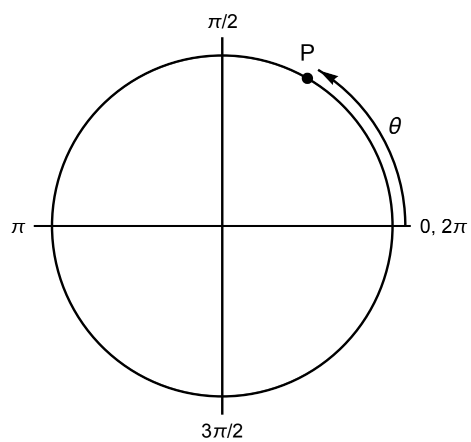
Thus one of the solutions is $x = a$ (we already knew (a, a^3) is a point of intersection– we explicitly constructed the tangent line at this point). The other solution is $x = -2a$, and thus the other point of intersection is $(-2a, -8a^3)$.

8 Trigonometric functions

Spivak covers variations of trigonometric functions as he goes through the book. I initially found it distracting– it's hard enough to understand $\epsilon - \delta$ limits; it's harder still if you're also trying to learn trig as you go. But trig is important, and sooner or later it's time to understand it. Because trig examples are crucial in understanding the chain rule in the next chapter, that time is now. Here I introduce the basics of trigonometric functions, and then cover everything trig-related from Spivak's early chapters that I ignored until now.

8.1 Definitions

Let P be a point on a unit circle $x^2+y^2=1$. Let θ be the length of the arc from $(1,0)$ to P, measured counterclockwise along the circle. Then the coordinates of P are $(\cos\theta, \sin\theta)$.[12]



The measure of angles by the length of the arc is in units called *radians*. Recall the circumference of a circle is $C = 2\pi r$, and so the circumference of a unit circle is 2π. Thus π represents a $180°$ angle. Some common angles in radians are $2\pi, \pi, \frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{4}, \frac{\pi}{6}$, and $\frac{3\pi}{2}$. To convert these to degrees simply replace π with 180, and compute the fraction.

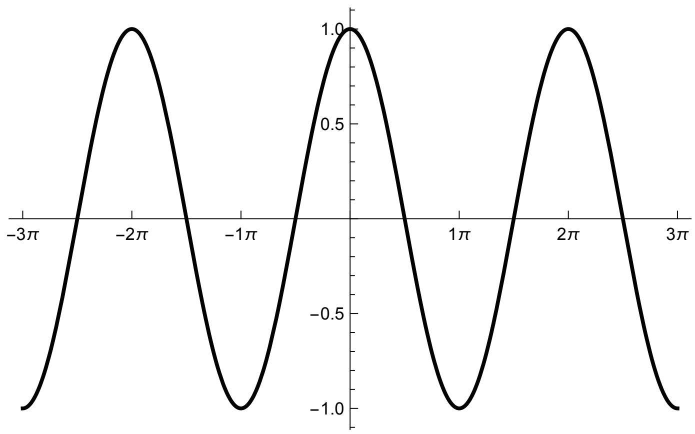
It should be self-evident that adding 2π to an angle results in the angle itself; and that adding $\frac{\pi}{2}$ to an angle shifts it by $90°$. Further:

$$(\cos 0, \sin 0) = (1,0) \qquad \text{and} \qquad (\cos\frac{\pi}{2}, \sin\frac{\pi}{2}) = (0,1)$$

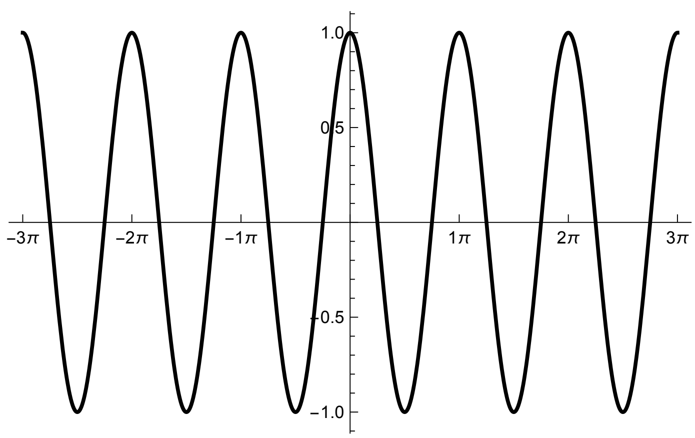
[12]The order is easy to remember– it's alphabetical.

8.2 Plotting

It is not too difficult to plot trigonometric functions. Consider some properties of cosine we've already seen (or can easily deduce): $\cos 0 = 1$, $\cos \frac{\pi}{2} = 0$, $\cos \pi = -1$. We've also seen that $\cos(x + 2\pi) = \cos x$. The x-axis below covers $[-3\pi, 3\pi]$ (i.e. a total length of 6π). Since cosine repeats every 2π, we should expect the graph to repeat thrice. And this is exactly what we see.



We can easily increase the frequency by plotting $y = \cos cx$. Here we double the frequency with $c = 2$:



8.3 Limits

Claim 1a: Let $f(x) = x \sin \frac{1}{x}$. Then $\lim_{x \to 0} f(x) = 0$.
Proof. Let $\epsilon > 0$ be given. We must show $0 < |x - 0| < \delta$ implies $|f(x) - 0| < \epsilon$. This is easy if you recall the co-domain of sin is $[-1, 1]$. This implies:

$$\left| x \sin \frac{1}{x} \right| \leq |x|$$

Thus fixing $|x| < \epsilon$ implies $\left| x \sin \frac{1}{x} \right| < \epsilon$ as desired.

—

Claim 1b: Let $f(x) = x^2 \sin \frac{1}{x}$. Then $\lim_{x \to 0} f(x) = 0$.
Proof. By the same reasoning as above, $\left| x^2 \sin \frac{1}{x} \right| \leq x^2$. Thus fixing $|x| < \sqrt{\epsilon}$ implies $\left| x^2 \sin \frac{1}{x} \right| < \epsilon$ as desired.

—

Claim 1c: Let $f(x) = \sqrt{|x|} \sin \frac{1}{x}$. Then $\lim_{x \to 0} f(x) = 0$.
Proof. By the same reasoning as above, $\left| \sqrt{|x|} \sin \frac{1}{x} \right| \leq \sqrt{|x|}$. Thus fixing $|x| < \epsilon^2$ implies $\left| \sqrt{|x|} \sin \frac{1}{x} \right| \leq \epsilon$ as desired.

—

Claim 2: Let $f(x) = \sin \frac{1}{x}$. Then $\lim_{x \to 0} f(x)$ does not exist.
Proof. Let $\epsilon = \frac{1}{2}$. For any δ there exists $x = \frac{1}{\pi/2 + 2\pi \cdot n} < \delta$. Observe that $f(x) = 1 > \epsilon$. Thus $\lim_{x \to 0} f(x)$ does not exist as desired.

—

Claim 3: Let $f(x) = \sin \frac{1}{x} = 0$. Then $\lim_{x \to \infty} f(x) = 0$.
Proof. Spivak states this without proof. The crux is that $|\sin t| < |t|$ for all real t (which we cannot prove until sin is defined later). But using this inequality the proof of the broader claim is simple.

Let $\epsilon > 0$ be given. We must show there exists N such that for all $x > N$, $|f(x)| < \epsilon$. Since $\left| \sin \frac{1}{x} \right| < \left| \frac{1}{x} \right|$, it suffices to ensure $\frac{1}{x} < \epsilon$, or $x > \frac{1}{\epsilon}$.

8.4 Continuity

Claim 1: $f(x) = \sin \frac{1}{x}$, $g(x) = x \sin \frac{1}{x}$ are not continuous at 0.
Proof. Neither function is defined at 0, and thus cannot be continuous at 0.

—

Claim 2: Let

$$G(x) = \begin{cases} x \sin \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Then G is continuous at 0.

Proof. This is easy. We saw that $\lim_{x \to 0} x \sin \frac{1}{x} = 0$, and $G(0) = 0$. Thus $\lim_{x \to 0} x \sin \frac{1}{x} = G(0)$, and thus G is continuous at 0.

—

Claim 3: G is continuous at all a.
Proof. We already saw that G is continuous at 0 in claim 2. By using continuity theorems (and assuming sin is continuous), it's easy to see $x \sin \frac{1}{x}$ is continuous for all $x \neq 0$. Thus f is continuous for all a.

—

Claim 4: Let

$$F(x) = \begin{cases} \sin \frac{1}{x}, & x \neq 0 \\ a, & x = 0 \end{cases}$$

Then F is *not* continuous at 0, for any choice of a.

Proof. We saw that $\lim_{x \to 0} x \sin \frac{1}{x}$ does not exist, thus F cannot be continuous at 0.

9 Derivatives, Part II (Differentiation)

9.1 Basic proofs

We now prove theorems that make differentiation of a large class of functions easy.

Theorem 1. If $f(x) = c$ then $f'(a) = 0$ for all a.

Intuitively derivatives measure the rate of change. A constant function doesn't change, thus the derivative is zero.

Proof: we already proved this in the previous chapter.

—

Theorem 2. If $f(x) = x$ then $f'(a) = 1$ for all a.

Intuitively $f(x)$ grows at exactly the same rate as x, thus the derivative is 1.

Proof:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = \lim_{h \to 0} \frac{a+h-a}{h} = 1$$

—

Theorem 3. If f, g are differentiable at a, then $(f+g)'(a) = f'(a) + g'(a)$.

Examples:

- You have two functions, each modeling growth of some bank account. You want to understand the rate of growth of both accounts.
- You have two different assembly lines producing the same product. $c_1(x)$ and $c_2(x)$ model the cost of producing x units on each assembly line. You want to understand total cost changes as production across both assembly lines increases.

Proof:

$$\begin{aligned}
(f+g)'(a) &= \lim_{h \to 0} \frac{(f+g)(a+h) - (f+g)(a)}{h} \\
&= \lim_{h \to 0} \frac{f(a+h) + g(a+h) - f(a) - g(a)}{h} \\
&= \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} + \lim_{h \to 0} \frac{g(a+h) - g(a)}{h} \\
&= f'(a) + g'(a)
\end{aligned}$$

—

Theorem 3a. If $f_1, \ldots, f_n$ are differentiable at a, then:

$$f_1 + \ldots_+ f_n)'(a) = f_1'(a) + \ldots_+ f_n'(a)$$

Proof. This is a fairly straightforward proof by induction. Skipping it here as I've already spent enough time on this chapter.

—

Theorem 4. If f, g are differentiable at a, then

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

Examples:

- Let $r_1(t), r_2(t)$ model the length of each side of a rectangle over time. You want to understand the change in area at time t.

Proof:

$$\begin{aligned}
(f \cdot g)'(a) = &= \lim_{h \to 0} \frac{(f \cdot g)(a+h) - (f \cdot g)(a)}{h} \\
&= \lim_{h \to 0} \frac{f(a+h)g(a+h) - f(a)g(a)}{h} \\
&= \lim_{h \to 0} \frac{f(a+h)g(a+h) - f(a)g(a) + f(a+h)g(a) - f(a+h)g(a)}{h} \\
&= \lim_{h \to 0} \frac{f(a+h)(g(a+h) - g(a)) + g(a)(f(a+h) - f(a))}{h} \\
&= \lim_{h \to 0} \left(f(a+h)\frac{g(a+h) - g(a)}{h} + g(a)\frac{f(a+h) - f(a)}{h} \right) \\
&= \lim_{h \to 0} f(a+h) \cdot \lim_{h \to 0} \frac{g(a+h) - g(a)}{h} + \lim_{h \to 0} g(a) \cdot \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \\
&= \lim_{h \to 0} f(a+h) \cdot g'(a) + g(a) \cdot f'(a)
\end{aligned}$$

Recall from 7.2 that if f is differentiable at a, then $\lim_{h \to 0} f(a+h) = f(a)$. Thus

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

—

Theorem 4a. If $f_1, \ldots, f_n$ are differentiable at a, then:

$$f_1 \cdot \ldots \cdot f_n)'(a) = \sum_{i=1}^{n} f_1(a) \cdot f_i'(a) \cdot f_n(a)$$

Proof. This is a fairly straightforward proof by induction. Skipping it here as I've already spent enough time on this chapter.

—

Theorem 5. If $g(x) = cf(x)$ then $g'(a) = c \cdot f'(a)$.

Examples:

- Let h be a height of a rectangle that's constant, and let $b(t)$ model the length of the base of a rectangle over time. You want to understand the change in area at time t.

Proof: Let $h(x) = c$ so $g = h \cdot f$. Then by theorem 4:

$$\begin{aligned} g'(x) &= h'(x)f(x) + f'(x)g(x) \\ &= 0 \cdot f(x) + cf'(x) \\ &= cf'(x) \end{aligned}$$

—

Theorems 1-5 imply:

$$(-f)'(a) = (-1 \cdot f)'(a) = -f'(a)$$

and

$$(f - g)'(a) = (f + (-g))'(a) = f'(a) + (-g)'(a) = f'(a) - g'(a)$$

—

Theorem 6. If $f(x) = x^n$ for $n \in \mathcal{N}$, then $f'(a) = na^{n-1}$ for all a.

Examples:

- Let $s(t)$ model the length of the side of a cube over time. You want to understand the change in volume at time t.

Proof. We prove this by induction. For $n = 1$, $f'(a) = 1$ by theorem 2.

Assume if $f(x) = x^n$ then $f'(a) = na^{n-1}$ for all a.

Let $I(x) = x$ and let $g(x) = x^{n+1} = xx^n$. Then $g(x) = I(x) \cdot f(x)$, i.e. $g = I \cdot f$. By theorem 4:

$$\begin{aligned} g'(a) &= (I \cdot f)'(a) \\ &= I'(a)f(a) + I(a)f'(a) \\ &= 1 \cdot a^n + a \cdot na^{n-1} \\ &= a^n + na^n \\ &= a^n(1 + n) \\ &= (n + 1)a^n \end{aligned}$$

—

Theorem 6b. If $f(x) = x^n$ for $n < 0$, then $f'(a) = na^{n-1}$ for all a. (In other words, we extend theorem 6 to negative exponents.)

Proof. We use theorem 7 below (putting 6b here for learning convenience).

$$\begin{aligned} f'(a) &= \left(\frac{1}{a^{-n}}\right)' \\ &= \frac{nx^{-n-1}}{x^{-2n}} \\ &= nx^{n-1} \end{aligned}$$

—

Theorem 7. If g is differentiable at a and $g(a) \neq 0$, then

$$\left(\frac{1}{g}\right)'(a) = \frac{-g'(a)}{[g(a)]^2}$$

Examples:

- Let $i(d) = \frac{1}{d^2}$ model the intensity of light, which is inversely proportional to the square of the distance from the source. You want to know how intensity changes with distance.

Proof. We will prove this by using the derivative definition. However, we must first show $\left(\frac{1}{g}\right)(a+h)$ is defined for sufficiently small h. This is easy.

Since g is differentiable at a it is continuous at a. Thus by nonzero neighborhood lemma (see 4.1) there exists $\delta > 0$ such that $|h| < \delta$ implies $g(a+h) \neq 0$ for all h. Thus $\left(\frac{1}{g}\right)(a+h)$ is defined for sufficiently small h.

We are now ready to prove the core of the theorem.

$$\begin{aligned} \lim_{h\to 0} \frac{\left(\frac{1}{g}\right)(a+h) - \left(\frac{1}{g}\right)(a)}{h} &= \lim_{h\to 0} \left(\frac{1}{g(a+h)} - \frac{1}{g(a)}\right)/h \\ &= \lim_{h\to 0} \left(\frac{g(a) - g(a+h)}{g(a)\cdot g(a+h)}\right)/h \\ &= \lim_{h\to 0} \frac{g(a) - g(a+h)}{h \cdot g(a) \cdot g(a+h)} \\ &= \lim_{h\to 0} \frac{-[g(a+h) - g(a)]}{h} \cdot \frac{1}{g(a)\cdot g(a+h)} \\ &= \lim_{h\to 0} \frac{-[g(a+h) - g(a)]}{h} \cdot \lim_{h\to 0} \frac{1}{g(a)\cdot g(a+h)} \end{aligned}$$

Recall from 7.2 that if f is differentiable at a, then $\lim_{h\to 0} f(a+h) = f(a)$. Thus:

$$\lim_{h\to 0} \frac{-[g(a+h)-g(a)]}{h} \cdot \lim_{h\to 0} \frac{1}{g(a)\cdot g(a+h)} = -g'(a) \cdot \frac{1}{[g(a)]^2}$$

as desired.

—

Theorem 8. If f, g are differentiable at a and $g(a) \neq 0$, then

$$\left(\frac{f}{g}\right)'(a) = \frac{g(a)\cdot f'(a) - f(a)\cdot g'(a)}{[g(a)]^2}$$

Examples:

- Let $e(t), s(t)$ model the number of engineers and sales people at a company over time. You want to understand the change in the ratio between the two.

Proof.

$$\begin{aligned}
\left(\frac{f}{g}\right)'(a) &= \left(f\cdot\frac{1}{g}\right)'(a) \\
&= f(a)\cdot\left(\frac{1}{g}\right)'(a) + f'(a)\cdot\left(\frac{1}{g}\right)(a) \\
&= \frac{-g'(a)\cdot f(a)}{[g(a)]^2} + \frac{f'(a)}{g(a)} \\
&= \frac{-g'(a)\cdot f(a)\cdot g(a) + f'(a)\cdot [g(a)]^2}{[g(a)]^3} \\
&= \frac{f'(a)\cdot g(a) - g'(a)\cdot f(a)}{[g(a)]^2}
\end{aligned}$$

9.2 Chain rule

The derivative of composed functions is considerably more complicated, and so deserves its own section. We'll prove this in two stages. First, we'll attempt a proof with a few false starts that will point us in the direction of a real proof. Once the direction becomes clear, we'll abandon our first draft and write a clean proof from scratch.

Theorem 9 (the chain rule). If g is differentiable at a, and f is differentiable at $g(a)$, then

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

Examples:

- Let $a(t)$ model altitude of a rocket over time, and let $p(a)$ model air pressure at a particular altitude. You want to know how air pressure changes over time.

Proof, first draft.

As usual, we start with the definition of the derivative:

$$\begin{aligned}
(f \circ g)'(a) &= \lim_{h \to 0} \frac{(f \circ g)(a+h) - (f \circ g)(a)}{h} \\
&= \lim_{h \to 0} \frac{f(g(a+h)) - f(g(a))}{h} \\
&= \lim_{h \to 0} \left(\frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} \cdot \frac{g(a+h) - g(a)}{h} \right) \\
&= \lim_{h \to 0} \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} \cdot \lim_{h \to 0} \frac{g(a+h) - g(a)}{h} \\
&= \left(\lim_{h \to 0} \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} \right) \cdot g'(a)
\end{aligned}$$

This is a bit of a false start as we now have two problems:

- To get $f'(g(a))$ in the first term, we need $\lim_{h \to 0} \frac{f(g(a)+h) - f(g(a))}{h}$, but instead we have $\lim_{h \to 0} \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)}$.
- $g(a+h) - g(a)$ may be zero for $h \neq 0$, so the division may be illegal.

However it isn't a total waste. Our false start gives us an idea for how we may proceed– we'll replace $\frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)}$ with something better. What could be the replacement? Let's hypothesize existance of a function $\phi(h)$ with the following property (we will soon prove such a function exists):

$$\frac{f(g(a+h)) - f(g(a))}{h} = \phi(h) \cdot \frac{g(a+h) - g(a)}{h}$$

We can then rewrite our initial equations as follows:

$$\begin{aligned}
(f \circ g)'(a) &= \lim_{h \to 0} \frac{(f \circ g)(a+h) - (f \circ g)(a)}{h} \\
&= \lim_{h \to 0} \frac{f(g(a+h)) - f(g(a))}{h} \\
&= \lim_{h \to 0} \left(\phi(h) \cdot \frac{g(a+h) - g(a)}{h} \right) \\
&= \lim_{h \to 0} \phi(h) \cdot \lim_{h \to 0} \frac{g(a+h) - g(a)}{h} \\
&= \lim_{h \to 0} \phi(h) \cdot g'(a)
\end{aligned}$$

To get to $(f \circ g)'(a) = f'(g(a)) \cdot g'(a)$ we need $\phi(h)$ to possess one more property:

$$\lim_{h \to 0} \phi(h) = f'(g(a))$$

Given this additional property, we can now finish our reasoning:

$$(f \circ g)'(a) = \lim_{h \to 0} \phi(h) \cdot g'(a) = f'(g(a)) \cdot g'(a)$$

Thus proving the chain rule reduces to proving there exists a function $\phi(h)$ with the two properties above. For cleanliness, let's start a new proof from scratch and demonstrate the existance of such a function.

Proof.

Suppose there exists a function $\phi(h)$ with the following properties:

$$\frac{f(g(a+h)) - f(g(a))}{h} = \phi(h) \cdot \frac{g(a+h) - g(a)}{h} \tag{1}$$

$$\lim_{h \to 0} \phi(h) = f'(g(a)) \tag{2}$$

Then

$$\begin{aligned}
(f \circ g)'(a) &= \lim_{h \to 0} \frac{(f \circ g)(a+h) - (f \circ g)(a)}{h} \\
&= \lim_{h \to 0} \frac{f(g(a+h)) - f(g(a))}{h} \\
&= \lim_{h \to 0} \left(\phi(h) \cdot \frac{g(a+h) - g(a)}{h} \right) && \text{by property 1} \\
&= \lim_{h \to 0} \phi(h) \cdot \lim_{h \to 0} \frac{g(a+h) - g(a)}{h} \\
&= \lim_{h \to 0} f'(g(a)) \cdot g'(a) && \text{by property 2}
\end{aligned}$$

To complete the proof we must construct such a function and prove our construction has properties 1 and 2. We will do so now. Define ϕ as follows:

$$\phi(h) = \begin{cases} \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} & \text{if } g(a+h) - g(a) \neq 0 \\ f'(g(a)) & \text{if } g(a+h) - g(a) = 0 \end{cases}$$

We will prove properties 1 and 2 hold for ϕ.

Property 1 proof.
We now show $\frac{f(g(a+h))-f(g(a))}{h} = \phi(h) \cdot \frac{g(a+h)-g(a)}{h}$. There are two cases: either $g(a+h) - g(a) \neq 0$ or $g(a+h) - g(a) = 0$. Suppose $g(a+h) - g(a) \neq 0$. Then

$$\begin{aligned}
\phi(h) \cdot \frac{g(a+h) - g(a)}{h} &= \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} \cdot \frac{g(a+h) - g(a)}{h} \\
&= \frac{f(g(a+h)) - f(g(a))}{h}
\end{aligned}$$

Alternatively, suppose $g(a+h) - g(a) = 0$. Then

$$\begin{aligned}\phi(h) \cdot \frac{g(a+h) - g(a)}{h} &= f'(g(a)) \cdot \frac{g(a+h) - g(a)}{h} \\ &= f'(g(a)) \cdot \frac{0}{h} \\ &= 0\end{aligned}$$

But $g(a+h) - g(a) = 0$ means $g(a+h) = g(a)$, and thus $\frac{f(g(a+h)) - f(g(a))}{h} = 0$. Thus in both cases property 1 holds, as desired.

Property 2 proof.
We now show $\lim_{h \to 0} \phi(h) = f'(g(a))$. Put differently:

- *Intuitively*, we're trying to show that when h is small, the top piece of ϕ piecewise definition approaches the bottom piece (which we chose to be $f'(g(a))$).
- Here is another way to frame it. Observe that $\phi(0) = f'(g(a))$. Thus showing $\lim_{h \to 0} \phi(h) = f'(g(a))$ is equivalent to showing $\lim_{h \to 0} \phi(h) = \phi(0)$, i.e. that ϕ is continuous at 0.
- Formally, we must show that given $\epsilon > 0$ there exists $\delta > 0$ such that $|h| < \delta$ implies $|\phi(h) - f'(g(a))| < \epsilon$.

So, let $\epsilon > 0$ be given.

Firstly, since f is differentiable at $g(a)$, by definition of the derivative we have:

$$f'(g(a)) = \lim_{k \to 0} \frac{f(g(a)+k) - f(g(a))}{k}$$

Inlining the limit defition, for all $\epsilon > 0$ there exists $\delta' > 0$ such that $0 < |k| < \delta'$ implies

$$\left| \frac{f(g(a)+k) - f(g(a))}{k} - f'(g(a)) \right| < \epsilon$$

Secondly, since g is differentiable at a, it continuous at a. Thus:

$$\lim_{h \to 0} g(a+h) = g(a)$$

Or put differently, there exists $\delta > 0$ such that $|h| < \delta$ implies:

$$|g(a+h) - g(a)| < \delta'$$

Finally, we now have everything we need to prove property 2. Consider any h with $|h| < \delta$.

- If $g(a+h) - g(a) = 0$ then $\phi(h) = f'(g(a))$ so $|\phi(h) - f'(g(a))| < \epsilon$.

- If $g(a+h) - g(a) \neq 0$ we can fix $k = g(a+h) - g(a)$ as both aren't 0 and are less than δ'. Thus we get:

$$\begin{aligned}
\epsilon &> \left| \frac{f(g(a)+k) - f(g(a))}{g(a+h) - g(a)} - f'(g(a)) \right| \\
&= \left| \frac{f(g(a)+g(a+h)-g(a)) - f(g(a))}{g(a+h) - g(a)} - f'(g(a)) \right| \\
&= \left| \frac{f(g(a+h)) - f(g(a))}{g(a+h) - g(a)} - f'(g(a)) \right| \\
&= |\phi(h) - f'(g(a))|
\end{aligned}$$

I.e. $|\phi(h) - f'(g(a))| < \epsilon$ as desired.

—

Theorem 9a. Let f_i be differentiable at $f_{i+1}(\ldots f_n(x)\ldots)$. Then:

$$(f_1 \circ \ldots \circ f_n)'(x) = \prod_{i=1}^{n} f_i' \left(f_{i+1}(\ldots f_n(x)\ldots)\right)$$

Proof. This is a fairly straightforward proof by induction. Skipping it here as I've already spent enough time on this chapter.

9.3 Derivatives of polynomials

We can easily find derivatives of polynomials using theorems 1-6. It turns out to be an interesting enough form that it's worth mentioning explicitly. Consider

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0$$

Then:

$$f'(x) = na_nx^{n-1} + (n-1)a_{n-1}x^{n-2} + \ldots + 2a_2x + a_1$$

Continuing:

$$f''(x) = n(n-1)a_nx^{n-2} + (n-1)(n-2)a_{n-1}x^{n-3} + \ldots + 2a_2$$

Repeatedly continuing this process we get:

$$f^{(n)}(x) = n!a_n$$

And of course for $m > n$ it's easy to see $f^{(m)} = 0$.

9.4 Differentiation practice

Spivak spends a lot of the chapter covering concrete differentiation examples. I work through these here. First, a summary of the nine differentiation theorems proved above:

1. If $f(x) = c$ then $f'(a) = 0$.
2. If $f(x) = x$ then $f'(a) = 1$.
3. $(f+g)'(a) = f'(a) + g'(a)$.
4. $(f \cdot g)'(a) = f'(a) \cdot g(a) + f(a) \cdot g'(a)$.
5. If $g(x) = cf(x)$ then $g'(a) = c \cdot f'(a)$.
6. If $f(x) = x^n$ for $n \in \mathcal{N}$, then $f'(a) = na^{n-1}$.
7. $\left(\frac{1}{g}\right)'(a) = \frac{-g'(a)}{[g(a)]^2}$.
8. $\left(\frac{f}{g}\right)'(a) = \frac{g(a) \cdot f'(a) - f(a) \cdot g'(a)}{[g(a)]^2}$.
9. $(f \circ g)'(a) = f'(g(a)) \cdot g'(a)$.

You also need to know two trig derivatives presented below without proof (proper proofs will show up in a later chapter when sin and cos are formally defined):

$$\begin{aligned} \sin'(a) &= \cos a \\ \cos'(a) &= -\sin a \end{aligned}$$

We are now ready to practice example problems.

$$\begin{aligned}
f(x) = \frac{x^2-1}{x^2+1} &\implies f'(x) = \frac{(x^2+1)2x - (x^2-1)2x}{(x^2+1)^2} = \frac{4x}{(x^2+1)^2} \\
f(x) = \frac{x}{x^2+1} &\implies f'(x) = \frac{1-x^2}{(x^2+1)^2} \\
f(x) = \frac{1}{x} = x^{-1} &\implies f'(x) = -\frac{1}{x^2} = (-1)x^{-2} \\
f(x) = x \sin x &\implies f'(x) = \sin x + x \cos x \\
&\implies f''(x) = 2\cos x - x \sin x \\
g(x) = \sin^2 x = \sin x \sin x &\implies g'(x) = 2 \sin x \cos x \\
&\implies g''(x) = 2\cos^2 x - 2\sin^2 x \\
h(x) = \cos^2 x = \cos x \cos x &\implies h'(x) = -2\sin x \cos x \\
&\implies h''(x) = 2\sin^2 x - 2\cos^2 x
\end{aligned}$$

Note $g'(x) + h'(x) = 0$. This is something we could have guessed– $(g+h)(x) = \sin^2 x + \cos^2 x = 1$, thus by theorem 1, $(g+h)'(x) = 0$.

$$\begin{aligned}
&f(x) = x^3 \sin x \cos x \\
&\implies f'(x) = 3x^2 \sin x \cos x + x^3 \cos^2 x - x^3 \sin^2 x
\end{aligned}$$

The next set of examples uses the chain rule (where sometimes the product rule could be used instead). For example, $\sin^2 x$ could be interpreted either as $\sin x \sin x$, or as $s(\sin x)$ where $s(x) = x^2$.

$$\begin{aligned}
f(x) = \sin x^2 &\implies f'(x) = \cos x^2 \cdot 2x \\
f(x) = \sin^2 x &\implies f'(x) = 2 \sin x \cdot \cos x \\
f(x) = \sin x^3 &\implies f'(x) = \cos x^3 \cdot 3x^2 \\
f(x) = \sin^3 x &\implies f'(x) = 3 \sin^2 x \cdot \cos x \\
f(x) = \sin \frac{1}{x} &\implies f'(x) = cos \frac{1}{x} \cdot \frac{-1}{x^2} \\
f(x) = \sin(\sin x) &\implies f'(x) = \cos(\sin x) \cdot \cos x \\
f(x) = \sin(x^3 + 3x^2) &\implies f'(x) = \cos(x^3 + 3x^2) \cdot (3x^2 + 6x) \\
f(x) = (x^3 + 3x^2)^{53} &\implies f'(x) = 53(x^3 + 3x^2)^{52} \cdot (3x^2 + 6x)
\end{aligned}$$

We now consider a composition of three functions:

$$\begin{aligned}
f(x) = \sin^2 x^2 = s \circ (\sin \circ s) &\implies f'(x) = 2 \sin x^2 \cdot \cos x^2 \cdot 2x \\
f(x) = \sin(\sin x^2) = \sin \circ (\sin \circ s) &\implies f'(x) = \cos(\sin x^2) \cdot \cos x^2 \cdot 2x
\end{aligned}$$

And finally a composition of four functions:

$$\begin{aligned}
&f(x) = \sin^2(\sin^2 x) = s \circ (\sin \circ (s \circ \sin)) \\
&\implies f'(x) = 2 \sin(\sin^2 x) \cdot \cos(\sin^2 x) \cdot 2 \sin x \cdot \cos x \\
&f(x) = \sin((\sin x^2)^2) = \sin \circ s \circ \sin \circ s \\
&\implies f'(x) = \cos((\sin x^2)^2) \cdot 2 \sin x^2 \cdot \cos x^2 \cdot 2x \\
&f(x) = \sin^2(\sin(\sin x)) = s \circ \sin \circ \sin \circ \sin \\
&\implies f'(x) = 2 \sin(\sin(\sin x)) \cdot \cos(\sin(\sin x)) \cdot \cos(\sin x) \cdot \cos x
\end{aligned}$$

9.5 Sine polynomials

I don't think "sine polynomials" is a real name, but I needed a clever name for this section. Here we explore derivatives of functions of the form $x^k \sin \frac{1}{x}$.

Claim 1: Let

$$f(x) = \begin{cases} x \sin \frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Then f is not differentiable at 0.

Proof. Using derivative definition:

$$\lim_{h\to 0} \frac{f(0+h)-f(0)}{h} = \lim_{h\to 0} \frac{h\sin\frac{1}{h}-0}{h} = \lim_{h\to 0} \sin\frac{1}{h}$$

We saw in 8.3 that $\lim_{h\to 0} \sin\frac{1}{h}$ does not exist. Thus f is not differentiable at zero.

—

Claim 2: Let

$$f(x) = \begin{cases} x^2\sin\frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Then f is differentiable at 0.

Proof. Using derivative definition:

$$\lim_{h\to 0} \frac{f(0+h)-f(0)}{h} = \lim_{h\to 0} \frac{h^2\sin\frac{1}{h}-0}{h} = \lim_{h\to 0} h\sin\frac{1}{h} = 0$$

Thus $f'(0) = 0$.

—

Claim 3: Let

$$f(x) = \begin{cases} x^2\sin\frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Then f' is not differentiable at 0.

Proof. Observe that:

$$f'(x) = \begin{cases} 2x\sin\frac{1}{x} - \cos\frac{1}{x}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

Observe that $\lim_{x\to 0} \cos\frac{1}{x}$ does not exist (for the same reason $\lim_{x\to 0} \sin\frac{1}{x}$ does not exist). Thus $\lim_{x\to 0} f'(x)$ does not exist. And thus f' is not continuous, let alone differentiable at 0.

10 Derivatives, Part III (Leibniz notation)

The notation f' that we've used so far is called the Lagrange notation.[13] However, there is another notation for the derivative in common use. You may have already seen something like $\frac{dy}{dx}$. This is called the Leibniz notation.

The Leibniz notation has many of what Spivak calls "vagaries". It has multiple interpretations– formal and informal. The informal interpretation doesn't map to modern mathematics, but can *sometimes* be useful (while at other times misleading). The full, unambigous Leibniz notation, at least as Spivak defines it, is verbose, so in practice people end up taking liberties with it. As a consequence, its meaning must often be discerned from the context.

This flexibility makes the notation very useful in science and engineering, but also makes it difficult to learn. Spivak chose to standardize on the Lagrange notation to maximize clarity, and banished Leibniz notation to problem sections. But since the Leibniz notation is so common, I take a different approach and explore it here in a dedicated chapter.

10.1 Historical motivation

We start with the historical interpretation, where the notation began. Leibniz didn't know about limits. He thought the derivative is the value of the quotient

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

when h is "infinitesimally small". He denoted this infinitesimally small quantity of h by dx, and the corresponding difference $f(x+dx)-f(x)$ by $df(x)$. Thus for a given function f the Leibniz notation for its derivative f' is:

$$\frac{df(x)}{dx} = f'$$

Intuitively, we can think of d in a historical context as "delta" or "change". Then we can interpret this notation as Leibniz did– a quotient of a tiny change in $f(x)$ and a tiny change in x. However, this explanation comes with two important disclaimers.

First, d is not a value. If it were a value, you could cancel out d's in the numerator and the denomenator. But you can't. Instead think of d as an operator. When applied to $f(x)$ or x, it produces an infinitesimally small quantity. Alternatively you can think of $df(x)$ and dx as one symbol that happens to look like multiplication, but isn't.[14]

[13] Wikipedia claims the notation was invented by Euler and Lagrange only popularized it.

[14] I read somewhere that in his notebooks Leibniz experimented with extending d with a squiggle on top that went over x to indicate that d is not a value, but I haven't been able to verify if that's true.

Second, note that $\frac{df(x)}{dx}$ denotes a function equivalent to f', *not* a value equivalent to $f'(x)$. To denote the image of the derivative function at a we use the following notation:

$$\left.\frac{df(x)}{dx}\right|_{x=a} = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h} = f'(a)$$

10.2 Modern interpretation

To summarize, the **full and unambiguous Leibniz notation** in modern interpretation is:

$$\frac{df(x)}{dx} = f' \qquad \text{and} \qquad \left.\frac{df(x)}{dx}\right|_{x=a} = f'(a)$$

Real numbers do not have a notion of infinitesimally small quantities. Thus in a modern interpretation we treat $\frac{df(x)}{dx}$ as a symbol denoting f', *not* as a quotient of numbers. Nothing here is being divided, nothing can be canceled out. In a modern interpretation $\frac{df(x)}{dx}$ is just one thing that *happens to look* like a quotient but isn't, anymore than f' is a quotient.

10.3 Second derivative

A question arises for how to express the second (or nth) derivative in the Leibniz notation. Let $g(x) = \frac{df(x)}{dx}$ (i.e. let g be the first derivative of f). Then it follows that the second derivative in Leibniz notation is $\frac{dg(x)}{dx} = g' = f''$. Substituting the definition of g we get:

$$\frac{d\left(\frac{df(x)}{dx}\right)}{dx} = f''$$

Of course this is too verbose and no one wants to write it this way. This is where the vagaries begin. For convenience people use the usual algebraic rules to get a simpler notation, eventhough formally everything is one symbol and you can't actually do algebra on it:

$$\frac{d\left(\frac{df(x)}{dx}\right)}{dx} = \frac{d^2f(x)}{dx^2}$$

Two questions arise here.

First, why dx^2? Shouldn't it be $(dx)^2$? One way to answer this question is to remember that dx is one symbol, *not* a multiplication (because d is not a value). And so we're just squaring that one symbol dx, which doesn't require parentheses.

Another probably more honest way to answer this question is to recall that this isn't real algebra– we just use a simularcum of algebra out of convenience. But convenience is a morally flexible thing, and people decided to drop parentheses because they're a pain to write. So $(dx)^2$ became dx^2.

Second, we said before that $df(x)$ can be thought of as one symbol. Then what is this d^2 business? The answer here is the same– we aren't doing real algebra, but a simularcum of algebra out of convenience. We aren't really squaring anything; we're overloading exponentiation to mean "second derivative". The symbol $d^2f(x)$ is again one symbol.

10.4 Liberties and ambiguities

There are a few more liberties people take with the Leibniz notation. Let $f(x) = x^2$. If we want to denote the derivative of f we can do it in two ways:

$$\frac{df(x)}{dx} \qquad \text{or} \qquad \frac{dx^2}{dx}$$

Here $\frac{dx^2}{dx}$ is new, but the meaning should be clear. We're just replacing $f(x)$ in $df(x)$ with the definition of $f(x)$. This is a little confusing because in the particular case of $f(x) = x^2$, it's visually similar to the notation for second derivative. There are no ambiguities here so far– it's just a visual artifact of the notation we have to learn to ignore. But now the liberties come.

Suppose we wanted to state what the derivative of f is. In Lagrange notation we say $f'(a) = 2a$. In Leibniz notation the proper way to say it would be as follows:

$$\left.\frac{df(x)}{dx}\right|_{x=a} = 2a$$

But this is obviously a pain, so people end up taking two liberties. First, everyone drops the vertical line that denotes the application at a. So in practice the form above becomes:

$$\frac{df(x)}{dx} = 2x$$

This shouldn't "compile" because $\frac{df(x)}{dx} = f'$. Thus this statement is equivalent to saying $f' = 2x$, which doesn't make sense. But this is the notation most people use, and you have to get used to it.

Second, people decided that writing $\frac{df(x)}{dx}$ is too painful, and in practice everyone writes $\frac{df}{dx}$. This also shouldn't compile (it would be something like writing $\lim_{x\to a} f$, which also doesn't make sense). But again, it's the notation most people use.

To summarize what we have so far:

$$\left.\frac{df(x)}{dx}\right|_{x=a} = 2a \qquad \text{becomes} \qquad \frac{df}{dx} = 2x$$

10.5 Chain rule

How do we express the chain rule $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$ in Leibniz notation? In the full and unambiguous version the chain rule ought to look like this:

$$\frac{df(g(x))}{dx} = \left.\frac{df(y)}{dy}\right|_{y=g(x)} \cdot \frac{dg(x)}{dx}$$

But, surprise, nobody does it this way. Usually people say that if $y = g(x)$ and $z = f(y)$ then:

$$\frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx}$$

Let's go through some examples of using this formula, and then see what's going on here. Let $z = \sin y, y = \cos x$. Then

$$\begin{aligned}\frac{dz}{dx} &= \frac{dz}{dy} \cdot \frac{dy}{dx} \\ &= \cos y \cdot (-\sin x) \\ &= -\cos(\cos x) \cdot \sin x\end{aligned}$$

How about $z = \sin u, u = x + x^2$? Well,

$$\begin{aligned}\frac{dz}{dx} &= \frac{dz}{du} \cdot \frac{du}{dx} \\ &= \cos u \cdot (2x+1) \\ &= \cos(x + x^2) \cdot (2x+1)\end{aligned}$$

How about a more complicated chain $z = \sin v, v = \cos u, u = \sin x$?

$$\begin{aligned}\frac{dz}{dx} &= \frac{dz}{dv} \cdot \frac{dv}{dx} \\ &= \frac{dz}{dv} \cdot \frac{dv}{du} \cdot \frac{du}{dx} \\ &= \cos v \cdot (-\sin u) \cdot \cos x \\ &= -\cos(\cos(\sin x) \cdot \sin(\sin x) \cdot \cos x\end{aligned}$$

Now, there are a bunch of notational liberties here:

- $y = \ldots$ implicitly defines a function $y(x)$ which is then used in e.g. $\frac{dy}{dx}$. But y can also be references as a value (e.g. "plot y when x is …"). So the deliniation between functions and the values they take on is blurred.
- dz on the left side of the equations (e.g. in $\frac{dz}{dx}$) denotes $f \circ g$. But dz on the right side of the equations (e.g. in $\frac{dz}{dy}$) denotes f. In other words, the denomenator has a bearing on the meaning of the numerator.

- $\frac{dz}{dy}$ denotes the derivative function, but is also understood to be "an expression involing y" that must be substituted with the value of y in the final answer. E.g. in the first example $\frac{dz}{dy}$ is equal to $\cos y$, and we must then substitute y with $\cos x$.

Despite all these quirks and ambiguities, with some practice we begin to see how easy and useful the Leibniz notation is. In the next two sections we will refine this understanding as we deal with physical problems involving the derivative.

10.6 Implicit differentiation

10.7 Notation practice

11 Derivatives, Part IV (Consequences)