3 Axioms for the real number system

Calculus is concerned with differentiation and integration of functions of the real numbers. Before understanding differentiation and integration, we need to understand functions, and before understanding functions, we need to understand the real numbers.

3.1 Why the axiomatic approach?

We all have an intuitive idea of the various number systems of mathematics:

1. the *natural numbers*, $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$ — the ordinary counting numbers, that can be added, multiplied, and sometimes divided and subtracted, and the slightly more sophisticated version of the natural numbers, that includes 0 — we'll denote this set as $\mathbb{N}^0$;

2. the *integers*, $\mathbb{Z} = \{\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots\}$ — the natural numbers together with their negatives, which allow for subtraction in all cases;

3. the *rationals*, $\mathbb{Q}$, the set of all numbers of the form a/b where a and b are integers[34], and b is not 0, which allows for division in all cases; and

4. the *real numbers*, which we denote by $\mathbb{R}$, and which "fill in the gaps" in the rational numbers — π, for example, or $\sqrt{2}$, are not expressible as the ratio of two integers, but they are important numbers that we need to have in our number system. Mathematical life is much easier when we pass from

   - the rational numbers, which may be thought of as a large (actually infinite) collection of marks on the number line:

     $$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,$$

     that's very dense, but doesn't cover the whole line,

   to

   - real numbers, which may be thought of as the *entire* number line:

     $$\overline{\hspace{8cm}},$$

[34] There is a subtlety here. Any given fraction appears infinitely often among the set of all numbers of the form a/b; for example, two-thirds appears as 2/3, 4/6, 6/9, et cetera.To be proper, we should say that each rational number is an *infinite set* of expressions of the form a/b, where a and b are integers, $b \neq 0$, satisfying $a/b = a'/b'$ for all pairs a, b and a', b' in the set. The specific rational number represented by the infinite set is the common value of all these ratios.

It will be helpful for us to keep these intuitive ideas in mind as we go through the formal definition of the real numbers; we will use them to provide (hopefully helpful) illustrative examples as we go along. But it is very important to remember that in this section our goal is to

> *rigorously derive all important properties of the real numbers*;

and so our intuitive ideas will only ever be used for illustration.

The approach we are going to take to understanding the real numbers will be *axiomatic*: we will write down a collection of axioms, that we should all agree that the real numbers should satisfy, and then we will *define* the real number system to be the set of objects that satisfy all the axioms.

This begs four questions:

- **Q1**: What axioms should we choose?
- **Q2**: How do we know that there is actually a set of objects that satisfies the axioms?
- **Q3**: Even if we know there is such a set, how do we know that it is the *only* such set (as the language we've used above — "*the* real number system [is] *the* set of objects that satisfy all the axioms" — strongly suggests?
- **Q4**: If there is a unique set of objects that satisfies the axioms, why don't we approach the real numbers by actually *constructing* that set? Why the more abstract axiomatic approach?

We'll briefly discuss possible answers now.

- **A1**: What axioms should we choose? We'll try to capture what we believe are the essential properties on out "intuitive" real numbers, with a collection of axioms that are as simple as possible. Most of the axioms will turn out to be very obvious uncontroversial; a few will be less obvious, motivated by considering our intuitive understanding of the reals, but still uncontroversial; and only one will be non-obvious. This last will be the axiom that separates the rational numbers from the reals, and will be the engine that drives almost every significant result of calculus.
- **A2**: Is there a set of objects that satisfies the axioms? Starting from the basic rules of set theory, it *is* possible to construct a set of sets, and to define addition and multiplication on that set, in such a way that the result behaves exactly as we would expect the real numbers to behave. We could therefore take a *constructive* approach to the real numbers, and indeed Spivak devotes the last few chapters of his text to explaining much of this construction.
- **A3**: Is there a *unique* such a set? Essentially, yes; again, this is discussed in the final chapters of Spivak's text.

- **A4**: Why the more abstract axiomatic approach, then? A good question: if we can *construct* the reals, why not do so?

  One reason not to is that the construction is quite involved, and might well take a whole semester to fully describe, particularly since it requires understanding the axioms of set theory before it can get started.

  Another reason is a more practical, pedagogical one: most mathematical systems don't have the luxury, that the reals have, of having an essentially *unique* model. In a little while in your mathematical studies, for example, you will see the incredibly important notion of a *vector space*. It will turn out that there are many, many (infinitely many) essentially different instances of a vector space. That means that it hopeless to try to study vector spaces constructively. Instead we have to approach axiomatically — we set down the basic properties that we want a "vector space" to satisfy, then derive all the further properties that it must satisfy, that follow, via rules of logic, from the basic properties, and then know that *every* one of the infinitely many instances of a vector space must satisfy all these further properties.

  The same goes for most of the other basic mathematical objects, such as groups, metric spaces, rings, fields, ... Since so many mathematical objects need to be studied axiomatically, it's good to get started on the axiomatic approach as early as possible!

3.2 The axioms of addition

From now on, whenever we talk about the *real numbers*, we are going to mean the following: the real numbers, which we denote by $\mathbb{R}$, is a set of objects (which we'll call *numbers*)

- including two special numbers, 0 and 1 ("zero" and "one")

together with

- an operation, $+$ ("addition"), which can combine any two numbers a, b to form another (not necessarily different) number, $a + b$, and
- another operation, $\cdot$ ("multiplication"), which can combine any two numbers a, b to form another number, $a \cdot b$,

and which satisfies a collection of 13 axioms, which we are going to label P1 through P13, and which we will introduce slowly over the course of the next few sections.

The first four axioms, P1 through P4, say that addition behaves in all the obvious ways, and that 0 plays a special role in terms of addition:

- **P1, Additive associativity**: For all a, b, c,

$$a + (b + c) = (a + b) + c.$$

- **P2, Additive identity**: For all a,

$$a + 0 = 0 + a = a.$$

- **P3, Additive inverse**: For all a there's a number $-a$ such that

$$a + (-a) = (-a) + a = 0.$$

- **P4, Additive commutativity**: For all a, b,

$$a + b = b + a.$$

Comments on axiom P1

Axiom P1 says that when we add together three numbers, the way in which we parenthesize the addition is irrelevant (this property is referred to as *associativity*); so whenever we work with real numbers, we can *unambiguously* write expressions like

$$a + b + c.$$

But what about adding together *four* numbers? There are five different ways in which we can add parentheses to a sum of the form $a + b + c + d$, to describe the order in which the addition should take place:

- $(a + (b + c)) + d$
- $((a + b) + c) + d$
- $a + (b + (c + d))$
- $a + ((b + c) + d)$
- $(a + b) + (c + d)$.

Of course, if the set of "real numbers" we are axiomatizing here is to behave as we expect the real numbers to behave, then we want all five of this expressions to be the same. Do we need to add an axiom, to declare that all five expressions are the same? And then, do we need to add another axiom to say that all 14 ways of parenthesizing $a + b + c + d + e$ are the same? And one that says that all 42[35] ways of parenthesizing $a + b + c + d + e + f$ are the

[35] How many different ways are there to parenthesize the expression $a_1 + a_2 + \ldots + a_n$? For $n = 2, 3, 4, 5, 6, \ldots$, the answer is $1, 2, 5, 14, 42, \ldots$, as can be verified by a brute-force search. The sequence of numbers that comes up in this problems is very famous (among mathematicians ...): it is the answer to literally hundreds of different problems, has been the subject of at least five books, and has been mentioned in at least 1400 mathematical papers. Although it's not obvious, the terms of the sequence follow a very simple rule.

same? And ... you see where I'm going — do we need to add infinitely many axioms, just to say that for every n,

$$a_1 + a_2 + \ldots + a_n$$

is an unambiguous expression, whose value doesn't depend on the way in which parenthesize?

Fortunately, no! Using the rules of inference, we can *deduce* that **if**

$$a + (b + c) = (a + b) + c$$

for all possible choices of a, b, c, **then**

$$(a+(b+c))+d = ((a+b)+c)+d = a+(b+(c+d)) = a+((b+c)+d) = (a+b)+(c+d)$$

for all possible choices of a, b, c, d. We formulate this as a claim; it will be the first proper proof of the course.

Claim 3.1. *If a, b, c and d are real numbers, then each of $(a + (b + c)) + d$, $((a + b) + c) + d$, $a + (b + (c + d))$, $a + ((b + c) + d)$ and $(a + b) + (c + d)$ are the same.*

Proof: We first show that $(a + (b + c)) + d = ((a + b) + c) + d$. By axiom P1, $(a + (b + c)) = ((a + b) + c)$, and so $(a + (b + c)) + d = ((a + b) + c) + d$ follows immediately from the rules of equality (specifically, from **E3**).

By virtually the same reasoning[36], $a + (b + (c + d)) = a + ((b + c) + d)$.

We now consider $(a + (b + c)) + d$ and $a + ((b + c) + d)$. We apply axiom P1, but with a twist: P1 says that for any A, B and C, $(A + B) + C = A + (B + C)$. We apply this with $A = a$, $B = b + c$ and $C = d$ to conclude that $(a + (b + c)) + d = a + ((b + c) + d)$.

All this shows that first four expressions are all equal to each other. So what is left to show is that fifth equals *any one* of the first four. We leave this as an exercise to the reader.[37]
$\square$[38]

Note that this was an example of a *direct* proof.

It would take much more work to show that all 14 ways of parenthesizing $a+b+c+d+e$ lead to the same answer, but this too can be shown to follow from P1. Sadly, using this approach it would take infinitely much work to show that for *all* n, and all $a_1, a_2, \ldots, a_n$, all ways of parenthesizing $a_1 + a_2 + \cdots + a_n$ lead to the same answer. We could get over this problem by adding infinitely many P1-like axioms to our set of axioms; fortunately, we will

[36]If two parts of a proof are basically identical, it's quite acceptable to describe one of them in detail, and then say that the other is essentially the same. **But**: you should *only* do this if the two arguments really are basically identical. You should *not* use this to worm your way out of writing a part of a proof that you haven't fully figured out!

[37]You'll see this expression — "we leave this as an exercise to the reader" — a lot throughout these notes. I *strongly* encourage you to do these exercise. They will help you understand to concepts that are being discussed, and they are good practice for the quizzes, homework and exams, where some of them will eventually appear.

[38]It's traditional to use this symbol — $\square$ — to mark the end of a proof.

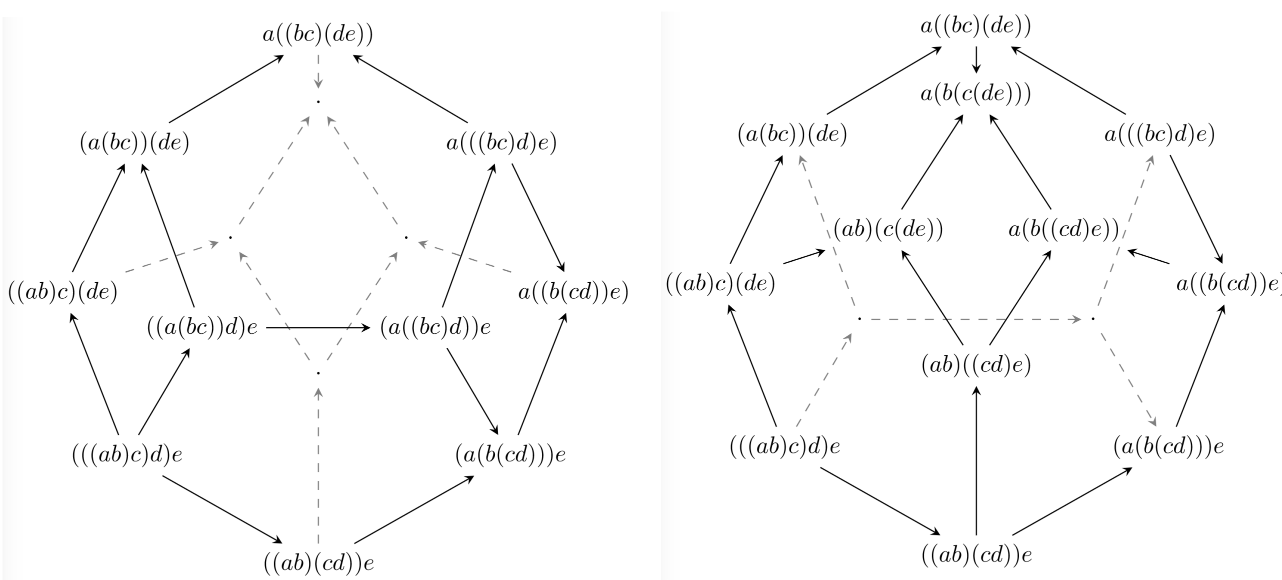
soon come to a method — proof by induction — that allows us to prove statements about *all* natural numbers in a finite amount of time, and we will use this method to show that it is indeed the case that the expression

$$a_1 + a_2 + \cdots + a_n$$

doesn't depend on the way in which it is parenthesized. So from here on, we will allow ourselves to assume this truth.

Aside: The *associahedron* K_n is an object that consists of points and edges. The points are all the different ways of parenthesizing the expression $a_1 + a_2 + \cdots + a_n$, and there is an edge between two points, if it is possible to show that the two associated expressions are equal, by applying the associativity axiom $a + (b + c) = (a + b) + c$ (with appropriate choices of a, b and c). So, for example, the associahedron K_4 has five points, namely $(a_1 + (a_2 + a_3)) + a_4$, $((a_1{+}a_2){+}a_3){+}a_4$, $a_1{+}(a_2{+}(a_3{+}a_4))$, $a_1{+}((a_2{+}a_3){+}a_4)$ and $(a_1{+}a_2){+}(a_3{+}a_4)$, and there is an edge joining $(a_1{+}(a_2{+}a_3)){+}a_4$ and $a_1{+}((a_2{+}a_3){+}a_4)$, because one application of axiom P1 (with $a = a_1$, $b = a_2{+}a_3$ and $c = a_4$ to conclude that $a_1{+}((a_2{+}a_3){+}a_4) = (a_1{+}a_2){+}(a_3{+}a_4)$.

It's evident that the associahedron K_3 is just a line segment joining two points (and so is 1-dimensional), and after a little bit of work you should be able to convince yourself that K_4 is a pentagon (and so is 2-dimensional). The associahedron K_5, corresponding to the different ways of parenthesizing the sum of 5 terms, is a pretty 3-dimensional shape, built out of squares and pentagons, shown in the pictures below (front and back):



The associahedron K_5

The fact that the associahedron K_5 is connected — it's possible to move from any one point to any other, along edges — shows that all 14 ways of parenthesizing the sum of 5 things end up yielding the same final value.

In higher dimensions — when "5 things" is replaced with "6 things", "7 things", and so on — the associahedra continue to be connected, showing that the way that an arbitrary sum is parenthesized doesn't change the final answer; but this isn't (or at least shouldn't be) obvious!

I'm mentioning associahedra, because there is a Notre Dame connection: much of what we know about associahedra was discovered by Jim Stasheff, who worked on them while he was a professor here from 1962 to 1968. See https://en.wikipedia.org/wiki/Associahedron for more information, including a nice animation of the three-dimensional associahedron.

Comments on axioms P2 and P3

Axiom P2 says that the number 0 is special, in that when it is added to anything, or when anything is added to it, nothing changes. We of course know (from our intuitive understanding of real numbers) that 0 should be the *unique* number with these special properties. The axiom doesn't say that, but fortunately this extra property of 0 can be *deduced* (proven) from the axioms as presented. If fact, something a little stronger is true:

Claim 3.2. *If x and a are any numbers satisfying $a + x = a$, then $x = 0$.*

Proof: We simply "subtract a" from both sides of the equation $a + x = a$:

Since $a + x = a$, we know that

$$-a + (a + x) = (-a) + a.$$

Using P1 on the left and P3 on the right, this says that

$$((-a) + a) + x = 0.$$

Using P3 on the left, this says that $0 + x = 0$, and using P2 on the left we finally conclude that $x = 0$. $\square$

Notice that this allows us to deduce the uniqueness of 0: if x is such that $a+x = x+a = a$ for all a, then in particular $a + x = a$ for some particular number a, so by the above claim, $x = 0$.

Notice also that we had to use all three of axioms P1, P2 and P3 to prove an "obvious" fact; this will be a fairly common feature of what follows. We're trying to produce as simple as possible a set of axioms that describe what we think of as the real numbers. So it makes sense that we're going to have to make wide use of this simple set of axioms to verify the more complex, non-axiomatic properties that we would like to verify.

The proof above proceeded by adding $-a$ to both sides of the equation, which we though of as "subtracting a". We formalize that idea here:

for any numbers a, b, we define "$a - b$" to mean $a + (-b)$.

Axiom P3 says that every number a has an *additive inverse* (which we denote by $-a$): a number which, when added to a, results in the answer 0. Of course, the additive inverse of each number should be *unique*. We leave it as an exercise to the reader to verify this: for any numbers a and b, if $a+b=0$, or if $b+a=0$, then $b=-a$.

Another property we would expect to be true of addition is the *cancellation* property. We leave it as an exercise to prove that if a, b, c are any numbers, and if $a+b=a+c$, then $b=c$.

Comments on axiom P4

Axiom P4 tells us the order in which we add two numbers doesn't affect the sum. This property of addition is referred to as *commutativity*. This is not true of all operations that we will perform on numbers — we don't expect $a-b$ to be equal to $b-a$ in general, for example — so for addition, it really needs to be explicitly said.

We know that in fact if we add n numbers, for any n, the order in which we add the numbers doesn't impact the sum. It should be fairly clear that we don't need to add any new axioms to encode this more general phenomenon. For example, while there are six different ways of ordering three numbers, a, b, c, to be added[39], namely

- $a+b+c$
- $a+c+b$
- $b+a+c$
- $b+c+a$
- $c+a+b$
- $c+b+a$,

it's easy to see that all six of them are equal to $a+b+c$. All we need to do is to repeatedly apply commutativity to neighboring pairs of summands, first to move the a all the way to the left, then to move the b to the middle position. For example,

$$c+b+a=c+(b+a)=c+(a+b)=(c+a)+b=(a+c)+b=a+(c+b)=a+(b+c)=a+b+c$$

(overkill: since we have already fully discussed associativity, I could have just written

$$c+b+a=c+a+b=a+c+b=a+b+c).$$

As with associativity, once we have proof by induction we will easily prove that when we add any n terms, $a_1, \ldots, a_n$, the sum doesn't depend on the pairwise order in which the pairs are added. So from here on, we will allow ourselves to assume this truth.

[39] In an earlier footnote we asked the question, "How many different ways are there to parenthesize the expression $a_1+a_2+\ldots+a_n$?". We can ask the analogous question here: "How many different ways are there to order the n summands $a_1, a_2, \ldots, a_n$?" This question is much easier than the one for parenthesizing. For $n=2,3,4,5,6,\ldots$ the sequence of answers is $2,6,24,120,720,\ldots$, and you should quickly be able to see both the pattern, and the reason for the pattern.

3.3 The axioms of multiplication

The next four axioms, P5 through P8, say that multiplication behaves in all the obvious ways, and that 1 plays a special role in terms of multiplication:

- **P5, Multiplicative associativity**: For all a, b, c,
$$a \cdot (b \cdot c) = (a \cdot b) \cdot c.$$
- **P6, Multiplicative identity**: For all a,
$$a \cdot 1 = 1 \cdot a = a.$$
- **P7, Multiplicative inverse**: For all a, if $a \neq 0$ there's a number a^{-1} such that
$$a \cdot a^{-1} = a^{-1} \cdot a = 1.$$
- **P8, Multiplicative commutativity**: For all a, b,
$$a \cdot b = b \cdot a.$$

These axioms look almost identical to those for addition. Indeed, replace "$\cdot$" with "$+$" and "1" with "0" in P5 through P8, and we have almost exactly P1 through P4. *Almost* exactly: we know that *every* number should have a negative (an additive inverse), but it it only *non-zero* numbers that have a reciprocal (a multiplicative inverse), and so in P7 we explicitly rule out 0 having an inverse.

Looking at what we did for addition, you should be able to prove the following properties of multiplication:

- 1 is unique: if x is such that $a \cdot x = a$ for all a, then $x = 1$, and
- the multiplicative inverse is unique: if $a \neq 0$, and if x is such that $a \cdot x = 1$, then $x = a^{-1}$,

and you should also be able to convince yourself that associativity and commutativity of multiplication extend to the product of more than two terms, without the need for additional axioms

With regards the first bullet point above: it will not be possible to prove for multiplication, the analog of the most general statement we proved for addition: it is *false* that "if a and x are any numbers, and $a \cdot x = a$, then $x = 1$". Indeed, if $a = 0$ then a counterexample to this statement is provided by any $x \neq 0$. Instead, the most general statement one can possibly prove is "if a and x are any numbers, with $a \neq 0$, and if $a \cdot x = a$, then $x = 1$".

We have mentioned the cancellation property of addition. There is a similar property of multiplication, that says we can "cancel" a common factor on both sides of an equation, *as long as that factor is not* 0.

Claim 3.3. *If a, b, c are any numbers, and $a \cdot b = a \cdot c$, then* either $a = 0$ or $b = c$.

Proof: Suppose that $a \cdot b = a \cdot c$. We want to argue that either $a = 0$ or $b = c$ (or perhaps both).

There are two possibilities to consider. If $a = 0$, then we are done (since if $a = 0$ it is certainly the case that either $a = 0$ or $b = c$). If $a \neq 0$, then we must argue that $a \cdot b = a \cdot c$ implies $b = c$. We get to use that there is a number a^{-1} such that $a^{-1} \cdot a = 1$. Using this, the argument goes like:

$$\begin{array}{cc}
a \cdot b = a \cdot c & \text{implies (using P7, valid since } a \neq 0\text{)} \\
a^{-1} \cdot (a \cdot b) = a^{-1} \cdot (a \cdot c) & \text{which implies (using P5)} \\
(a^{-1} \cdot a) \cdot b = (a^{-1} \cdot a) \cdot c & \text{which in turn implies (using P7)} \\
1 \cdot b = 1 \cdot c & \text{which finally implies (using P6)} \\
b = c. &
\end{array}$$

□

Notice that

- *the proof above is presented in full sentences.* We did not simply write

$$\begin{array}{c}
\text{“}a \cdot b = a \cdot c \\
a^{-1} \cdot (a \cdot b) = a^{-1} \cdot (a \cdot c) \\
(a^{-1} \cdot a) \cdot b = (a^{-1} \cdot a) \cdot c \\
1 \cdot b = 1 \cdot c \\
b = c,\text{”}
\end{array}$$

  and notice also that

- *every step of the proof was justified* (in this case, by reference to a particular axiom).

My expectation is that you will always present your proofs in complete sentences, and initially with every step justified (this condition will get relaxed soon, but for now it is the expectation!)

The proof above proceeded by multiplying both sides of the equation by a^{-1}, which we think of as "dividing by a". We formalize that idea here:

for any numbers a, b, with $b \neq 0$, we define "a/b" to mean $a \cdot (b^{-1})$.

We've mentioned two special properties of 0 — it is the additive inverse, and it is the unique number that we do not demand has a multiplicative inverse. There is a third special property of 0, that we know from our intuitive understanding of real numbers, namely that $a \cdot 0 = 0$ for any a. This hasn't been mentioned in the axioms so far, but we definitely want it to be true. There are two possibilities:

- *either* we can deduce $(\forall a)(a \cdot 0 = 0)$ from the axioms so far,
- *or* we can't, in which case we really need another axiom!

It turns out that we are in the second situation above — it is not possible to prove that for all a, $a \cdot 0 = 0$, just using Axioms P1 through P8. And in fact, we can *prove* that we can't prove this! We won't bother to make the digression and do that here, but I'll explain how it works. Suppose that we can find a set X of "numbers", that include numbers "0" and "1", and we can define operations "+" and "·" on this set of numbers, in such as way that all of the axioms P1 through P8 hold, *but for which also there is some number a with $a \cdot 0 \neq 0$.* Then that set X would act as a witness to prove that P1 through P8 alone are not enough to prove that for all a, $a \cdot 0 = 0$. (It's actually quite simple to find such a set X. Consider it a challenge!)

An obvious choice of new axiom is simply the statement "for all a, $a \cdot 0 = 0$" — if we want this to be true, and it doesn't follow from the axioms so far, then let's force it to be true by adding it as a new axiom. The route we'll take is a little different. We'll add an axiom that talks in general about how addition and multiplication interact with each other.

3.4 The distributive axiom

The next axiom, that links addition and multiplication, lies at the heart of almost every algebraic manipulation that we will ever do.

- **P9, Distributivity of multiplication over addition**: For all a, b, c,

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c).$$

As a first substantial consequence, let's use P9 to prove that $a \cdot 0 = 0$.

Claim 3.4. *If a is any number then $a \cdot 0 = 0$.*

Proof: By P2,

$$0 + 0 = 0.$$

Multiplying both sides by a, we get that

$$a \cdot (0 + 0) = a \cdot 0.$$

By P9, this implies

$$a \cdot 0 + a \cdot 0 = a \cdot 0. \quad (\star)$$

Adding $-(a \cdot 0)$ to the left-hand side of $(\star)$, applying P1, then P3, then P2, we get

$$a \cdot 0 + a \cdot 0 + (-a \cdot 0) = a \cdot 0 + (a \cdot 0 - a \cdot 0) = a \cdot 0 + 0 = a \cdot 0. \quad (\star\star)$$

Adding $-(a \cdot 0)$ to the right-hand side of $(\star)$, and applying P3, we get

$$a \cdot 0 + (-a \cdot 0) = 0. \quad (\star\star\star)$$

Since the left- and right-hand sides of $(\star)$ are equal, they remain equal on adding $-(a \cdot 0)$ to both sides, so combining $(\star\star)$ and $(\star\star\star)$ we get

$$a \cdot 0 = 0.$$

□

Another important consequence of P9 is the familiar property of real numbers, that if the product of two numbers if zero, then at least one of the two is zero.

Claim 3.5. *If $a \cdot b = 0$ then either $a = 0$ or $b = 0$.*

Proof: If $a = 0$, then there is no work to do, so from here on we assume that $a \neq 0$, and we argue that this forces $b = 0$.

Since $a \neq 0$ there is a^{-1} with $a^{-1} \cdot a = 1$. Multiplying both sides of $a \cdot b = 0$ by a^{-1}, we get

$$\begin{aligned} a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0, & \quad \text{which implies (by P5, P7 and Claim 3.4) that} \\ 1 \cdot b = 0, & \quad \text{which implies (by P6) that} \\ b = 0. & \end{aligned}$$

□

Notice that as well as using the axioms in this proof, we have also used a previously proven theorem, namely Claim 3.4. As the results we prove get more complicated, this will happen more and more. Notice also that we condensed three lines — applications of P5, P7 and Claim 3.4 — into one. This is also something that we will do more and more of as we build more proficiency at constructing proofs.

To illustrate the use of P9 in algebraic manipulations, consider the identity

$$x^2 - y^2 = (x - y) \cdot (x + y),$$

valid for all real x, y, where "x^2" is shorthand for "$x \cdot x$". (If you are not familiar with this identity, you should familiarize yourself with now; it will prove to be very useful.) To verify that it is a valid identity, note that, by P9, we have

$$\begin{aligned} (x - y) \cdot (x + y) &= (x - y) \cdot x + (x - y) \cdot y \\ &= (x + (-y)) \cdot x + (x + (-y)) \cdot y \quad \text{(definition of } -y) \\ &= x \cdot x + (-y) \cdot x + x \cdot y + (-y \cdot y) \quad \text{(P9)} \\ &= x \cdot x - y \cdot x + y \cdot x - y \cdot y \quad \text{(P8)} \\ &= x \cdot x - y \cdot y \quad \text{(various axioms, applied in obvious ways).} \end{aligned}$$

Now defining x^2 to mean $x \cdot x$, and $y^2 = y \cdot y$, we get the result. (Note: in the third line we are using a version of P9 that follows immediately from P9 using P8: for all a, b, c, $(b + c) \cdot a = (b \cdot a) + (c \cdot a)$.)

But wait! In the middle of the proof, we slipped in

$$(-y) \cdot x = -(y \cdot x)$$

(the product of (the additive inverse of y) and (x), is the same as the additive inverse of (the product of y and x)). This isn't an axiom, so needs to be proved! As a last example of the power of P9, we present a proof that suggests the rule "negative times negative equals positive", and along the way verifies the sneaky step in the above proof.

Claim 3.6. *For all numbers a, b, $(-a) \cdot (-b) = ab$*[40].

Proof: We begin by arguing that $(-a) \cdot (b) = -(a \cdot b)$. We know that $-(a \cdot b)$ is the additive inverse of $a \cdot b$, and that moreover it is the *unique* such inverse (a previous exercise for the reader). So if we could show $a \cdot b + (-a) \cdot (b) = 0$, then we could deduce $(-a) \cdot (b) = -(a \cdot b) = 0$. But by P9[41],

$$a \cdot b + (-a) \cdot (b) = (a + (-a)) \cdot b = 0 \cdot b = 0.$$

So indeed, $(-a) \cdot (b) = -(a \cdot b)$.

But now we have

$$\begin{aligned}
(-a) \cdot (-b) + (-(a \cdot b)) &= (-a) \cdot (-b) + (-a) \cdot (b) \quad \text{(by what we just proved above)} \\
&= (-a)((-b) + b) \quad \text{(by distributivity)} \\
&= (-a) \cdot 0 = 0.
\end{aligned}$$

But also, directly from P3,

$$a \cdot b + (-(a \cdot b)) = 0.$$

It follows, either by uniqueness of additive inverses, or by cancellation for addition, that $(-a) \cdot (-b) = a \cdot b$. $\square$

This is an example of a proof that is simple, in the sense that every step is easy to justify; but not *easy*, because to get the proof right, it is necessary to come up with just the right steps!

When we come (very shortly) to introduce positive and negative numbers, we will see that Claim 3.6 really can be interpreted to say that

[40] The seemingly similar statement that $-(-a) = a$ is much simpler, and follows from P1, P2 and P3. It's left as an exercise

[41] P9 says that $X \cdot (Y + Z) = (X \cdot Y) + (X \cdot Z)$. But using P8 (commutativity of multiplication), this is exactly the same as $(Y + Z) \cdot X = (Y \cdot X) + (Z \cdot X)$, and this is the form in which P9 is being used here. As we become more familiar with the concepts of commutativity, associativity, and distributivity, will we start to make this shortcuts more and more, without explicitly saying so.

in the real numbers, negative times negative is positive. $(\diamond)$

This is a rule of numbers that's hard to make intuitive sense of, but it is an unavoidable one. If we believe axioms P1 through P9 to be true statements about real numbers (and they all seem uncontroversial), then the proof of Claim 3.6 tells us that we must, inevitably, accept $(\diamond)$ as a true fact.

Before moving on we make one more (attempt at a) proof. If it clearly true that in the real numbers, the only solution to the equation

$$a - b = b - a$$

is $a = b$. We "prove" this by starting from $a - b = b - a$, adding $a + b$ to both sides, reordering terms, and applying the additive inverse axiom to get $a + a = b + b$, using multiplicative identity and distributivity to deduce $(1 + 1) \cdot a = (1 + 1) \cdot b$, and then multiplying both sides by the multiplicative inverse of $1 + 1$ to deduce $a = b$.

What's wrong with this "proof"? What is wrong is that we multiplied by the inverse of $1 + 1$. But we can only do this if $1 + 1 \neq 0$. Of course, we know that in the real numbers, $1 + 1$ is *not* 0. But, how do we know this in our axiomatic approach?

It turns out that we *cannot* prove $1 + 1 \neq 0$ using axioms P1 through P9 only. There is set of "numbers", including "0" and "1", together with operations "+" and "·", that satisfy all of axioms P1 through P9, but that also has $1 + 1 = 0$! To capture the real numbers, we therefore need more axioms. In the next section, we introduce the three axioms of *order*, that together rule out the possibility $1 + 1 = 0$.

3.5 The axioms of order

Nothing in the axioms so far have captured the notion that there is an *order* on the real numbers — that for any two distinct numbers a, b, one of a, b is bigger than the other. One way to rectify this is to introduce a relation "$<$" (with "$a < b$" meaning "a is less than b", or "b is greater than a"), and then add some axioms that describe how "$<$" should behave.

Another approach, the one we will take, is to declare a subset of the real numbers to be the "positive" numbers, add axioms that describe how positivity behaves with respect to addition and multiplication, and then *define* an order relation in terms if positivity.

The axioms of order that we will use say that there is a collection $\mathbb{P}$ (of *positive* numbers) satisfying

- **P10, Trichotomy law**: For every a exactly one of

  1. $a = 0$
  2. $a \in \mathbb{P}$
  3. $-a \in \mathbb{P}$.

  holds.

- **P11, Closure under addition**: If $a, b \in \mathbb{P}$ then
$$a + b \in \mathbb{P}.$$
- **P12, Closure under multiplication**: If $a, b \in \mathbb{P}$ then
$$ab \in \mathbb{P}.$$

Numbers which are neither positive nor zero are referred to as *negative*; there is no special notation for the set of negative numbers. This last definition immediately says that each number is exactly one of positive, negative or 0. The trichotomy axiom also fairly immediately implies the following natural facts:

Claim 3.7. *If a is positive then $-a$ is negative; and if a is negative then $-a$ is positive.*

Proof: We start with the second point. Suppose a is negative. Then, by definition, a is neither positive nor 0, so by Trichotomy $-a$ is positive.

Now for the first point. Suppose a is positive. Then $-a$ is not positive (that violates the Trichotomy axiom). Also, $-a$ is not 0, because then $-(-a)$ would be 0 too, and (exercise) $-(-a) = a$, so $a = 0$, violating trichotomy. So by definition of negativity, $-a$ is negative. $\square$.

Quickly following on from this, we get the familiar, fundamental, and quite non-intuitive statement that the product of two negative numbers is positive.

Claim 3.8. *If a and b are negative, then ab is positive.*

Proof: Since a and b are negative we have that $-a$ and $-b$ are positive, so by P12 $(-a)(-b)$ is positive. But by Claim 3.6 $(-a)(-b) = ab$, so ab is positive. $\square$

A fundamental and convenient property of the real numbers is that it is possible to put an *order* on them: there is a sensible notion of "greater than" ("$>$") and "less than" ("$<$") such that for any two numbers a, b with $a \neq b$, either a is greater than b or a is less than b. We now define one such notion of order, using the positive-negative-zero trichotomy.

Order definitions:

- "$a > b$" means $a - b \in \mathbb{P}$
- "$a < b$" means $b > a$
- "$a \geq b$" means that either $a > b$ or $a = b$ (i.e, "$a \geq b$" is the same as "$(a > b) \vee (a = b)$")
- "$a \leq b$" means that either $a < b$ or $a = b$

Note that "$a < b$" means the same as "$b - a \in \mathbb{P}$", so the same as "$-(b - a)$ is negative", which is easily seen to be the same as "$a - b$ is negative".

Applying the trichotomy law to $a - b$, and using the definitions of $<$ and $>$, we easily determine that for every a, b, exactly one of

- $a = b$
- $a < b$
- $a > b$

holds.

In Spivak's text (Chapter 1), many properties of $<$ are derived. You should look over these, and treat them as (excellent) exercises in working with axioms and definitions. There will also be some of them appearing on the homework. You shouldn't have to memorize them (and you *shouldn't* memorize them), because they are all obvious properties, that you are already very familiar with. Nor should you be memorizing proofs. Your goal should be to do enough of these types of proofs that they become instinctive.

We'll give two examples; there will be plenty more in the homework. In the sequel, we will freely use many properties of inequalities that we have not formally proven; but we will use nothing that you couldn't prove, if you chose to, using the ideas of this section.

Claim 3.9. *If $a < b$ and $b < c$ then $a < c$.*

Proof: Since $a < b$ we have $b - a \in \mathbb{P}$ and since $b < c$ we have $c - b \in \mathbb{P}$, so by P11, closure under addition, we get that $(b-a) + (c-b) = c - a$[42] $\in \mathbb{P}$, which says $a < c$. $\square$

Claim 3.10. *If $a < b$ and $c > 0$ then $ac < bc$.*

Proof: Since $a < b$ we have $b - a \in \mathbb{P}$ and since $c > 0$ we have $c \in \mathbb{P}$, so by P12, closure under multiplication, we get that $(b-a)c = bc - ac \in \mathbb{P}$, which says $ac < bc$. $\square$

Related to Claim 3.10 is the fact that when an inequality is multiplied by a *negative* number, the direction of the inequality is *reversed*:

$$\text{If } a < b \text{ and } c < 0 \text{ then } ac > bc.$$

We leave the proof of this as an exercise to the reader.

We now highlight an important consequence of Claim 3.8, and also use this to introduce for the first time the word "Corollary": a result that follows in a quite direct way as an application of a previous result.

Corollary 3.11. *(Corollary of Claim 3.8) If $a \neq 0$ then $a^2 > 0$.*

[42]Note that we're using associativity, commutativity, additive inverse and additive identity axioms here, without saying so. At this point these steps should be so obvious that they can go without saying.

Proof: If $a \neq 0$ then either a is positive, in which case $a^2 = a \cdot a$ is positive by P12, of a is negative, in which case $a^2 = a \cdot a$ is also positive, this time by Claim 3.8.[43] $\square$

A few more important corollaries tumble out now. The first is the (obvious?) fact that

- $1 > 0$.

Indeed, we have $1^2 = 1 \cdot 1 = 1$ by P6, so since $1 \neq 0$[44] Corollary 3.11 applies to conclude that $1^2 = 1$ is positive.

The second is that $1 + 1 \neq 0$; this follows from the facts that 1 is positive and that positivity is closed under addition. This is the fact that we need to go back and complete our earlier "proof" that $a - b = b - a$ only if[45] $a = b$.

Have we pinned down the real numbers with axioms P1 through P12? It seems not. Our intuitive notion of the rational numbers $\mathbb{Q}$ seems to satisfy all of P1 through P12, as does our intuitive notion of the reals $\mathbb{R}$; but we have a sense that the reals are "richer" than the rationals, containing "irrational" numbers like $\sqrt{2}$ and π. So it seems that more axioms are needed to precisely pin down the notion of real numbers — more on that in a short while.

For the moment, let us mention one more set of "numbers" that satisfy P1 through P9, but fail to satisfy the order axioms. This is the set $\mathbb{C}$ of *complex* numbers, numbers of the form $a + bi$ where a and b are real numbers and i is a newly introduced symbol that acts as a "square root" of -1 — it satisfies $i^2 = -1$ (since -1 is negative, and any non-zero number, when squared, is positive, there can be no real number whose square is -1).

The complex numbers are algebraically manipulated in all the obvious ways:

- $(a + bi) + (c + di) = (a + c) + (b + d)i$, and
- $(a + bi) \cdot (c + di) = ac + adi + bci + bdi^2 = ac + adi + bci - bd = (ac - bd) + (ad + bc)i$.

Their importance lies in the following fact: in the rationals, we can solve any equation of the form $ax + b = 0$ for $a \neq 0$, but we can't solve all quadratic equations, for example we can't solve $x^2 - 2 = 0$. Moving to the reals will allow us to solve that and many other quadratic equations, but not all of them, for example we can't solve $x^2 + 1 = 0$. We *can* solve this

[43] This is our first natural example of a *proof by cases*: the assertion to be proved is of the form $p \Rightarrow q$ where p:"$a \neq 0$". By trichotomy the premise p can be written as $p_1 \vee p_2$ where p_1:"$a > 0$" and p_2:"$a < 0$". Proof by cases says that to prove $(p_1 \vee p_2) \Rightarrow q$ it is necessary and sufficient to prove both $p_1 \Rightarrow q$ and $p_2 \Rightarrow q$, that is, to "break into cases", which is exactly what we have just done.

[44] Really? Is it true that $1 \neq 0$? You could try and prove this from the axioms, but you would fail, for the simple reason that there is a set of "numbers", together with special numbers 0 and 1, operations "+" and "$\cdot$" and a subset "$\mathbb{P}$" of positive numbers, that satisfies all of axioms P1 through P12, but for which $1 \neq 0$ *fails*, that is, for which $1 = 0$! The setup is simple: let 0 be the *only* number in the set (so $0 + 0 = 0$ and $0 \cdot 0 = 0$), let the special element 1 be that same number 0, and let $\mathbb{P}$ be empty. It's easy to check that all axioms are satisfied in this ridiculous setup. To rule out this giving a perfectly good model for real numbers, we actually have to build in to our definition of real numbers the fact that $0 \neq 1$. We'll say this explicitly when we summarize the axioms later.

[45] Our first "natural" use of "only if" for "implies".

quadratic in the complex numbers, via $x = i$ or $x = -i$. But presumably there are more complex polynomials that we can't even solve in complex numbers, no? No! Amazingly, once i is introduced to the number system, *every* polynomial

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$$

has a solution![46]

So, the complex numbers form a set that satisfies P1 through P9. What about P10, P11 and P12?

Claim 3.12. *It is not possible to find a subset $\mathbb{P}$ of the complex numbers for which axioms P10 through P12 hold.*

Proof: Suppose it was possible to find such a subset $\mathbb{P}$ of "positive" complex numbers. Consider the number i. We certainly have $i \neq 0$ (if $i = 0$ then $i^2 = 0$, but also $i^2 = -1$ by definition, so $-1 = 0$; and adding 1 to both sides gives $0 = 1$, a contradiction).

Since $i \neq 0$ we have $-1 = i^2 > 0$ by Corollary 3.11, so -1 is positive, so 1 is negative; but this contradicts the fact that 1 is positive.

This contradiction proves that no such a subset $\mathbb{P}$ can exist. $\square$

Axioms P1 through P12 describe a mathematical object called an *ordered field*; the above claim demonstrates that $\mathbb{C}$ is *not* an example of an ordered field.

3.6 The absolute value function

The *absolute value* of a number is a measure of "how far from 0" the number is, without regard for whether it is positive or negative.

- 0 itself has absolute value 0;
- if two positive numbers a and b satisfy $a < b$ (so b is bigger, "more positive" than a, "further from 0"), then the absolute value of a is smaller than the absolute value of b;
- if they are both negative and $a < b$ (so a is "more negative" than b) then the absolute value of a is bigger than the absolute value of b; and
- if a and b are negatives of each other ($a = -b$, $b = -a$) then they have the same absolute value.

The formal definition of the absolute value function is: for real a, the *absolute value* of a, denoted $|a|$, is given by

$$|a| = \begin{cases} a & \text{if } a > 0 \\ 0 & \text{if } a = 0 \\ -a & \text{if } a < 0. \end{cases}$$

[46] This is the *fundamental theorem of algebra.*

So, for example, $|2| = 2$, $|-\pi| = \pi$, and $|1-\sqrt{2}| = \sqrt{2}-1$. We will frequently define functions using this "brace" notation, so you have to get used to reading and using it. The brace notation above says that there are three different, disjoint regimes for the answer to the question "what is $|a|$?" — the regime $a > 0$ (where the answer is "a"); the regime $a = 0$ (where the answer is "0"); and the regime $a < 0$ (where the answer is "$-a$"). The three regimes have no overlap, so there is no possible ambiguity in the definition[47], and by trichotomy they cover all reals, so there are no gaps in the definition.

Noting that when $a = 0$ the numbers "a" and "0" coincide, we could have been a little more efficient, and broken into the two regimes $a \geq$ [48]0 and $a < 0$, to get:

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0. \end{cases}$$

It's a matter of taste which approach to take.

We will use the absolute value to create a notion of "distance" between numbers: the distance between a and b is $|a-b|$. Representing a and b on a number line, $|a-b|$ can be thought of as the length of the line segment joining a and b (a quantity which is positive, whether $a < b$ or $a > b$).

A fundamental principle of the universe is that "the shortest distance between two points is a straight line". A mathematical interpretation of this principle says that for any sensible notion of "distance" in a space, for any three points x, y, z

$$\text{(the distance from } x \text{ to } y\text{)}$$

is no larger than

$$\text{(the distance from } x \text{ to } z\text{) plus (the distance from } z \text{ to } y\text{)},$$

that is, it can never be *quicker* to get from x to y, if you demand that you must pass through a particular point z on the way (though it might be just as quick, if z happens to lie on a shortest path between x and y). The mathematical study of *metric spaces* explores these ideas.

In the context of using absolute value as a notion of distance between two real numbers, consider $x = a$, $y = -b$ and $z = 0$. The distance from x to y is $|a-(-b)| = |a+b|$, the distance from x to 0 is $|a-0| = |a|$, and the distance from 0 to $-b$ is $|0-(-b)| = |b|$. If we believe that absolute value is sensible as a notion of distance, then we would expect that $|a+b| \leq |a| + |b|$. This is indeed the case. The following, called the *triangle inequality* is one of the most useful tools in calculus.

[47]Sometimes we will present braced definitions in which there *is* overlap between regimes. As long as the two potentially conflicting clauses agree at the points of overlap (and they usually are just points), this is fine, if a little sloppy. As an example, this is an unambiguous and complete definition of absolute value, with two overlapping regimes:

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a \leq 0. \end{cases}$$

[48]Remember "$a \geq b$" is shorthand for "either $a > b$ or $a = b$"

Claim 3.13. *(Triangle inequality) For all reals a, b, $|a+b| \leq |a|+|b|$.*

Proof: Because the absolute value function is defined in cases, it makes sense to consider cases for a, b.

Case 1, $a, b \geq 0$ In this case, $|a| = a$, $|b| = b$, and (since $a+b \geq 0$), $|a+b| = a+b$, and so $|a+b| = |a|+|b|$.

Case 2, $a, b \leq 0$ In this case, $|a| = -a$, $|b| = -b$, and (since $a+b \leq 0$), $|a+b| = -a-b$, and so again $|a+b| = |a|+|b|$.

Case 3, $a \geq 0$**,** $b \leq 0$ Here we know $|a| = a$ and $|b| = -b$, but what about $|a+b|$?

If $a+b \geq 0$ then $|a+b| = a+b$, and to verify the triangle inequality in this case we need to establish

$$a+b \leq a-b,$$

or $b \leq -b$. Since $b \leq 0$, we have $-b \geq 0$ and $0 \leq -b$, so indeed $b \leq -b$.

If, on the other hand, $a+b < 0$ then $|a+b| = -a-b$, and to verify the triangle inequality in this case we need to establish

$$-a-b \leq a-b,$$

or $-a \leq a$. Since $a \geq 0$ (and so $0 \leq a$), we have $-a \leq 0$, so indeed $-a \leq a$.

Case 4, $a \leq 0$**,** $b \geq 0$ This is almost identical to Case 3, and we omit the details.

□

Another, more conceptual, proof of the triangle inequality appears in Spivak.

The absolute value function appears in two of the most important definitions of calculus — the definitions of limits and continuity — so it behooves us to get used to working with it. The standard approach to dealing with an expression involving absolute values is to break into cases, in such a way that within each case, all absolute value signs can be removed. As an example, let us try to find all real x such that

$$|x-1|+|x-2| > 1.$$

The clause in the absolute value definition that determines $|x-1|$ changes at $x = 1$, and the clause that determines $|x-2|$ changes at $x = 2$. It makes sense, then, to consider five cases: $x < 1$, $x = 1$, $1 < x < 2$[49], $x = 2$ and $x > 2$.

[49]This is shorthand for "$1 < x$ and $x < 2$."

Case 1: $x < 1$ Here $|x-1| = 1-x$ (since $x-1 < 0$ in this case), and $|x-2| = 2-x$, so $|x-1|+|x-2| = 3-2x$ and $|x-1|+|x-2| > 1$ is the same as $3-2x > 1$ or $1 > x$. So: in the regime $x < 1$, $|x-1|+|x-2| > 1$ is true exactly when $x < 1$, which it always is in this regime, and we conclude that the set of all $x < 1$ is one set of numbers satisfying the inequality.

Case 2: $x = 1$ Here $|x-1|+|x-2| = 1$, so the inequality is not satisfied.

Case 3: $1 < x < 2$ Here $|x-1|+|x-2| = x-1+2-x = 1$ and again the inequality is not satisfied.

Case 4: $x = 2$ Here $|x-1|+|x-2| = 1$, so again the inequality is not satisfied.

Case 5: $x > 2$ Here $|x-1|+|x-2| = 2x-3$ and the inequality becomes $x > 2$, which is true always in this regime, and we conclude that the set of all $x > 2$ is another set of numbers satisfying the inequality.

Having finished the case analysis, we conclude that the inequality is satisfied when x is less than 1 and when x is greater than 2.[50]

3.7 The completeness axiom

This section introduces the completeness axiom, which allows us to give a complete (no pun intended) description of the real numbers. Almost immediately after we are done with this section, the complete axiom will fade into the background. But in a few weeks, when we come to the major theorems of continuity — the intermediate value theorem and the extreme value theorem — it will come blazing back to the foreground, spectacularly.

Our intuition about the real numbers suggests that it cannot be the case that axioms P1 through P12 are *not* enough to pin down the real precisely, or uniquely: both $\mathbb{Q}$ and $\mathbb{R}$ (as we understand them, informally) satisfy all the axioms so far; but surely $\mathbb{R}$ contains more numbers than $\mathbb{Q}$ — numbers like $\sqrt{2}$, π, and e — so in particular $\mathbb{Q}$ and $\mathbb{R}$ should be different sets that both satisfy P1 through P12. We formalize this now, by presenting the ancient[51] proof that $\sqrt{2}$ is not a rational number.

Claim 3.14. *There do not exists natural numbers a, b with $\frac{a^2}{b^2} = 2$.*

Proof: Suppose, for a contradiction, that there are natural numbers a, b with $\frac{a^2}{b^2} = 2$. If a and b are both even, say $a = 2m$ and $b = 2n$ for some natural numbers m, n, then we have

$$\frac{m^2}{n^2} = \frac{4m^2}{4n^2} = \frac{(2m)^2}{(2n)^2} = \frac{a^2}{b^2} = 2,$$

[50] We will soon see the standard way to represent sets like this.

[51] Literally ancient — this proof is hinted at in Aristotle's *Prior Analytics*, circa 350BC.

so we could just as well use the pair of numbers m, n. By repeating this process, of dividing each number by 2 if both are even, until we can no longer do this, we reach a point where we have two natural numbers a', b', *not both even*, with $\frac{(a')^2}{(b')^2} = 2$.

We have $(a')^2 = 2(b')^2$, so $(a')^2$ is even. But that implies that a' is even (an odd number — say one of the form $2k+1$ for natural number k — can't square to an even number, since $(2k+1)^2 = 4k^2 + 4k + 1 = 2(2k^2 + 2k) + 1$, which is odd since $2k^2 + 2k$ is a whole number).

So $a' = 2c$ for some whole number c. Plugging in to $(a')^2 = 2(b')^2$, this yields $4c^2 = 2(b')^2$ or $2c^2 = (b')^2$. So $(b')^2$ is even, implying that b' is even.

This is a contradiction — we have that a', b' are not both even, and simultaneously are both even. It follows that there are *no* natural numbers a, b with $\frac{a^2}{b^2} = 2$. $\square$

Given that we would *like* there to be a square root of 2 in the real numbers, it makes sense to hunt for some more axioms. It turns out that we actually need just one more. Our intuitive sense is that though there are "gaps" in the rationals, the gaps are "small", and in fact the rationals are in some sense "dense" in the reals[52] — for every real r, there can be found a sequence of rational numbers that approaches arbitrarily close to r. Indeed, if we believe that every real r has a decimal expansion

$$r = a.d_1 d_2 d_3 d_4 \ldots,$$

then the sequence of rational numbers

$$a, a.d_1, a.d_1 d_2, a.d_1 d_2 d_3, \ldots$$

always lies below r, but eventually gets as close to r as one wishes. So, if we make sure that not only do the real contain all the rational numbers, but also contain all the "numbers" that can be approached arbitrarily closely by sequences of rationals, then it doesn't seem beyond the bounds of possibility that we would have "filled in all the gaps" in the rationals.

We'll formalize this idea with a single extra axiom. To state it sensibly, we need to introduce the notions of upper bounds and least upper bounds.

Informally, b is an upper bound for a set S of numbers if b is at least as large as everything in S. Formally, say that b is an *upper bound* for S if

$$\text{for all } s, \text{ if } s \text{ is an element of } S \text{ then } s \le b.$$

Some sets have upper bounds. For example,

- the set A of

  all numbers that are strictly bigger than 0 and strictly less than 1

  has 12 as an upper bound, and also 6, and 3, and 2.5, and 1; but not 1/2, or .99, or 0 or -10; and

[52] We will formalize this later.

- the set B of

  non-positive numbers (number that are 0 or less than 0)

  has 0 as an upper bound, and also *any* number bigger than 0, but not any number less than 0.

Other sets *don't* have upper bounds, such as

- the set of reals itself (there is no real number that is at least as large as all other real numbers, since for every real r the number $r + 1$ is a real that is bigger than r); and

- the natural numbers (for the same reason).

What about the empty set, the set that contains *no* numbers? We denote this set by $\emptyset$. Does $\emptyset$ have an upper bound? Yes! The number b is an upper bound for a set if

for all s, if s is an element of the set, then $s \leq b$.

Pick an arbitrary b, and then an arbitrary s. The premise of the implication "if s is an element of the empty set, then $s \leq b$" is "s is an element of the empty set". This premise is *false*. So the implication is true, for an arbitrary s and for all s. And so b is an upper bound for an arbitrary b, and so for all b.

This may seem counter-intuitive, but it is a consequence of the way we define implication. This is a situation where it might be very helpful to think of "p implies q" as meaning "either p is false, or q is true". Then the definition of b being an upper bound for S becomes

b is an upper bound for S if, for all s, either s is not a member of S, or $s \leq b$.

In this form, it is clear that *every* number is an upper bound for the empty set.

Having talked about upper bounds, we now introduce the concept of a least upper bound. A number b is a *least upper bound* for a set S if

- it is an upper bound for S (this is the "upper bound" part of the definition) and

- if b' is any other upper bound for S, then $b \leq b'$ (this is the "least" part of the definition).

The definition talks about "a" least upper bound; but it should be clear that if S has a least upper bound, then it has only one. Indeed, if b and b' are both least upper bounds then $b \leq b'$ (because b is a *least* upper bound), and also $b' \leq b$ (because b' is a *least* upper bound), so in fact $b = b'$.

Let's look at three examples: for the set A above, 1 and all numbers greater than 1 are upper bounds, and no other numbers are. So A has a least upper bound, and it is the number 1. Note that in this example, the least upper bound is *not* in the set A.

For the set B above, 0 and all numbers greater than 0 are upper bounds, and no other numbers are. So B has a least upper bound, and it is the number 0. Note that in this example, the least upper bound *is* in the set B.

It might seem like I'm working towards suggesting that *every* set that has an upper bound has a least upper bound. But our third example, the empty set, nixes that suggestion: every number is an upper bound for $\emptyset$, so it has upper bounds but no *least* upper bound.

But $\emptyset$ seems quite special. Maybe every *non-empty* set that has an upper bound, has a *least* upper bound? This is not a theorem that we can prove, using just axioms P1 through P12. Here's an informal reason: we sense that there is a "gap" in the rationals, where $\sqrt{2}$ should be. So, inside the rationals, consider the set C of all numbers x satisfying $x^2 < 2$. This set has an upper bound — 2, for example. But does it have a least upper bound? If b is any upper bound, then it must be that $b^2 > 2$ (we can't have $b^2 = 2$, since we are in the world of rationals; nor can we have $b^2 < 2$, because then we should be able to find another rational b', slightly larger than b, with $(b')^2$ still less than 2 — this using the idea that $\sqrt{2}$ can be approached arbitrarily closely by rationals). But (again using the idea that $\sqrt{2}$ can be approached arbitrarily closely by rationals) if $b^2 > 2$, then it can't be a *least* upper bound, because we should be able to find another rational b', slightly smaller than b, with $(b')^2$ still greater than 2, that acts as a lesser upper bound for the set C.

So, that every *non-empty* set that has an upper bound, has a *least* upper bound, is not a theorem we can prove in P1 through P12; but it seems like a good fact to have, because it seems to allow for the "filling in of the gaps" in the rationals — in the example discussed informally above, if C *had* a least upper bound b, it seems pretty clear that it should satisfy $b^2 = 2$, and so b should act as a good candidate for being the square root of 2.

This motivates the last, and most subtle, and most powerful, axiom of the real numbers, the *completeness axiom*:

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

The notation that is traditionally used for the (remember, it's unique if it exists) least upper bound of a set A of numbers is "l.u.b. A" or (much more commonly)

$$\sup A$$

(sup here is short for "supremum"). So the completeness axiom says:

If A is a non-empty set of numbers that has an upper bound, then $\sup A$ exists.

There's an equivalent form of the Completeness axiom, that involves lower bounds. Say that b is a *lower bound* for a set S if it is no bigger than any element of S, that is, if

for all s, if s is an element of S then $b \leq s$.

(As before with upper bounds, some sets have lower bounds, and some don't; and *every* number is a lower bound for the empty set.) A number b is a *greatest lower bound* for a set S if

- it is a lower bound for S and
- if b' is any other lower bound for S, then $b \geq b'$.

(As with least upper bounds, this number, if it exists, is easily seen to be unique). The notation that is traditionally used for the greatest lower bound of a set A of numbers is "g.l.b. A" or (much more commonly)

$$\inf A$$

(inf here is short for "infimum").

Following through our discussion about least upper bounds, but now thinking about greatest lower bounds, it seems reasonable clear that non-empty sets of numbers with lower bounds should have greatest lower bounds. This doesn't need a new axiom; it follows from (and in fact is equivalent to) the Completeness axiom.

Claim 3.15. *If A is a non-empty set of numbers that has a lower bound, then* $\inf A$ *exists.*

Proof: Consider the set $-A := \{-a : a \in A\}$. This is non-empty since A is non-empty, and it has an upper bound; if b is a lower bound for A, then $-b$ is an upper bound for $-A$. So $-A$ has a least upper bound, call it α.

We claim that $-\alpha$ is a greatest lower bound for A. Since α is an upper bound for $-A$, we have $\alpha \geq -a$ for all $a \in A$, so $-\alpha \leq a$, so $-\alpha$ is certainly a lower bound for A. Now suppose β is another lower bound for A. Then $-\beta$ is an upper bound for $-A$, so $-\beta \geq \alpha$, so $\beta \leq -\alpha$, so $-\alpha$ is the *greatest* lower bound for A. □

The key point in this proof is the following fact, which is worth isolating and remembering:

$$\inf A = -\sup(-A) \text{ where } -A = \{-x : x \in A\}.$$

3.8 Examples of the use of the completeness axiom

When we come to discuss continuity, we will see plenty of examples of the power of P13. For now, we give a few fairly quick examples. As a first example, we give a formalization of the informal discussion about the existence of $\sqrt{2}$ that came up earlier.

Claim 3.16. *There is a number x with $x^2 = 2$.*

Proof: Let C be the set of numbers a satisfying $a^2 < 2$.

C is non-empty — for example, 0 is in C. Also, C has an upper bound. For example, 2 is an upper bound. Indeed, we have $2^2 = 4 > 2$, and if $y \geq 2$ then $y^2 \geq 2^2 = 4 > 2$, so an element of C must be less than 2.

By the completeness axiom, C has a (unique) least upper bound. Call it x. We claim that $x^2 = 2$; we will prove this by ruling out the possibilities $x^2 < 2$ and $x^2 > 2$.

Suppose $x^2 < 2$. Consider the number

$$x' = \frac{3x+4}{2x+3}.$$

(It certainly is the case that $x > 0$, so x' really is a number — we are not guilty of accidentally dividing by zero.) Note that $x < x'$ is equivalent to $x < (3x+4)(2x+3)$, which is equivalent to $2x^3 + 3x < 3x + 4$, which is equivalent to $x^2 < 2$, which is true, so $x < x'$. And note also that $(x')^2 < 2$ is equivalent to

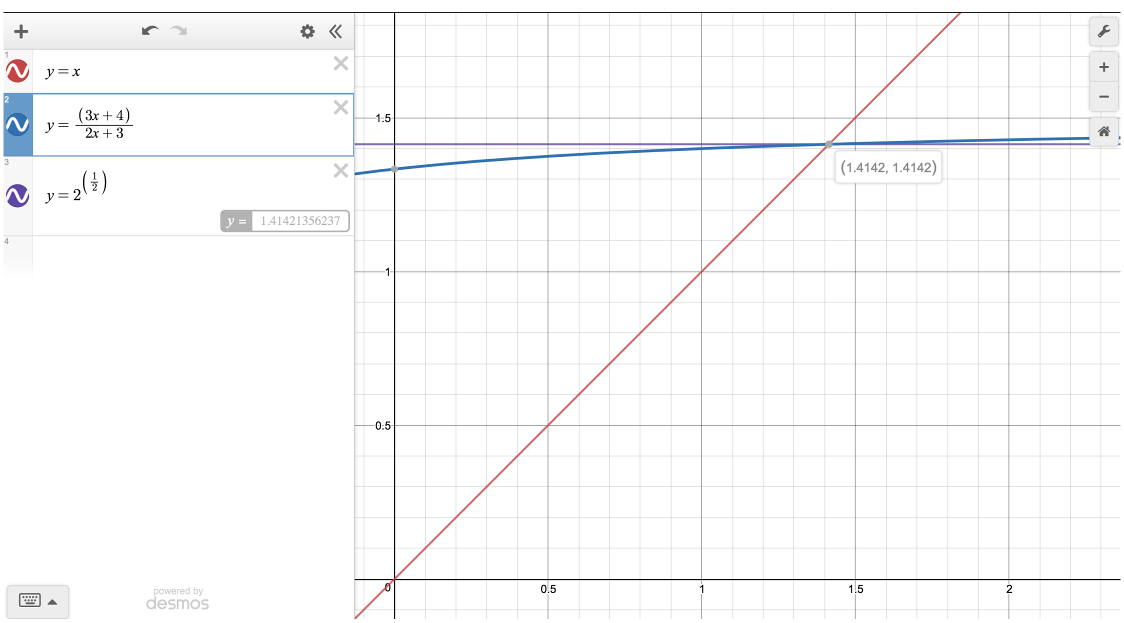
$$\left(\frac{3x+4}{2x+3}\right)^2 < 2,$$

which (after some algebra) is equivalent to $x^2 < 2$, which is true, so $(x')^2 < 2$. It follows that $x \in C$, but since $x < x'$ this contradicts that x is an upper bound for C. So we conclude that it is not the case that $x^2 < 2$.

Now suppose $x^2 > 2$. Again consider $x' = (3x+4)/(2x+3)$. Similar algebra to the last case shows that now $x' < x$ and $(x')^2 > 2$, so $y^2 > 2$ for any $y \geq x'$, so all elements of C are less than x', so x' is an upper bound for C, contradicting that x is the *least* upper bound for C. So we conclude that it is not the case that $x^2 > 2$.

We are left only with the possibility $x^2 = 2$, which is what we wanted to prove. $\square$

The picture below illustrates what's going on in the proof above. The red line is $y = x$, the blue curve is $y = (3x+4)/(2x+3)$, and the purple line is $y = \sqrt{2}$. All three lines meet at $(\sqrt{2}, \sqrt{2})$. The blue curve is between the two lines: above the red & below the purple before $\sqrt{2}$, and below the red, above the purple after $\sqrt{2}$.



So: if we take any number x with $x^2 < 2$ (so in regime where red is below blue is below purple), the three graphs illustrate that $(3x+4)/(2x+3)$ is bigger than x, but its square is still less than 2. So x can't be the l.u.b. for the set of numbers whose square is less than 2 — it's not even an upper bound, since it's smaller than $(3x+4)/(2x+3)$.

And: if we take any number x with $x^2 > 2$ (so in regime where purple is below blue is below red), the graphs illustrate that $(3x+4)/(2x+3)$ is smaller than x, but its square is still greater than 2. So x can't be the l.u.b. for the set of numbers whose square is less than 2 — it's an upper bound, but $(3x+4)/(2x+3)$ is a better (lesser) upper bound.

So the l.u.b. for the set of numbers whose square is less than 2 — which exists by the completeness axiom — has a square which is neither less than nor greater than 2. By trichotomy, it must be equal to 2.

The algebra in the proof is simply verifying that indeed red is below blue is below purple when $x^2 < 2$, and purple is below blue is below red when $x^2 > 2$.

Later we will prove the Intermediate Value Theorem (IVT), a powerful result that will make it essentially trivial to prove the existence of the square root, or cubed root, or any root, of 2, or any positive number. Of course, there is no such thing as a free lunch — we will need the completeness axiom to prove the IVT.

The next three examples are probably best looked at after reading the section on Natural numbers, as a few concepts from that section get used here.

The first of these is the use of completeness is to demonstrate the "obvious" fact that the natural numbers

$$\mathbb{N} = \{1, 1+1, 1+1+1, \dots\} = \{1, 2, 3, \dots\}$$

forms an *unbounded* set (a set with no upper bound). While this seems obvious, it should not actually be; there exist examples of sets of numbers satisfying P1-P12, in which $\mathbb{N}$ is *not* unbounded, meaning that P13 is absolutely necessary to prove this result.

The proof goes as follows. $\mathbb{N}$ is non-empty. Suppose it is bounded above. Then, by P13, it has a least upper bound, i.e., there's an $\alpha = \sup \mathbb{N}$. We have $\alpha \geq n$ for all $n \in \mathbb{N}$. Now if $n \in \mathbb{B}$, so is $n+1$, so this says that $\alpha \geq n+1$ for all $n \in \mathbb{N}$. Subtracting one from both sides, we get that $\alpha - 1 \geq n$ for all $n \in \mathbb{N}$. That makes $\alpha - 1$ an upper bound for $\mathbb{N}$, and one that is smaller than α, a contradiction! So $\mathbb{N}$ must not be bounded above.

Closely related to this is the *Archimedean property* of the real numbers:

> Let r be any positive real number (think of it as large), and let ε be any positive real number (think of it as small). Then there is a natural number n such that $n\varepsilon > r$.[53]

The proof is very quick: Suppose the property were false. Then there is some $r > 0$ and some $\varepsilon > 0$ such that $n\varepsilon \leq r$ for all $n \in \mathbb{N}$, so $n \leq r/\varepsilon$, so $\mathbb{N}$ is bounded above, and that's a contradiction.

[53] In his book *The sand-reckoner*, Archimedes put an upper bound on the number of grains of sand that could fit in the universe. Think of the Archimedean property as saying "no matter how small your grains of sand, or how large your universe, if you have enough grains of sand you will eventually fill the entire universe."

A simple and tremendously useful corollary of the Archimedean property is the special case $r = 1$:

> for all $\varepsilon > 0$ there is a natural number n such that $n\varepsilon > 1$, that is, so that $1/n < \varepsilon$.

The final application we give of the Completeness axiom we give in this quick introduction is to the notion of density.

> A set $S \subseteq \mathbb{R}$ is *dense in* $\mathbb{R}$ if for all $x < y$ in $\mathbb{R}$, there is an element of S in (x, y).[54]
> We also say that S is a *dense subset* of $\mathbb{R}$.

For example, the set of reals itself forms a dense subset of the reals, rather trivially, as does the set of reals minus one point. The set of positive numbers is *not* dense (there is no positive number between -2 and -1), and nor is the set of integers (there is no integer between 1.1 and 1.9).

Our intuition is that the rationals *are* dense in the reals. This is indeed the case.

Claim 3.17. *$\mathbb{Q}$ is dense in $\mathbb{R}$ — if x, y are reals with $x < y$, then there is a rational in the interval (x, y).*

Proof: We'll prove that for $0 \leq x < y$ there's a rational in (x, y). Then given $x < y \leq 0$, there's a rational r in $(-y, -x)$, so $-r$ is a rational in (x, y); and given $x < 0 < y$, any rational in $(0, y)$ is in (x, y)[55].

So, let $0 \leq x < y$ be given. By the Archimedean property, there's a natural number n with $1/n < y - x$. The informal idea behind the rest of the proof is that, because the gaps between consecutive elements in the "$1/n$" number line

$$\{\ldots, -3/n, -2/n, -1/n, 0, 1/n, 2/n, 3/n, \ldots\}$$

are all smaller than the distance between x and y, one (rational) number in this set must fall between x and y.

Formally: Because $\mathbb{N}$ is unbounded, there's $m \in \mathbb{N}$ with $m \geq ny$. Let m_1 be the least such (this is an application of the well-ordering principle). Note that $m_1 > 1$, because $y - x > 1/n$, so $y > 1/n$, so $1 < ny$. Consider $(m_1 - 1)/n$. We have $(m_1 - 1) < ny$ (or else m_1 would not have been the least integer at least as large as ny) and so $(m_1 - 1)/n < y$. If $(m_1 - 1)/n \leq x < y \leq m_1/n$ then $1/n \geq y - x$, a contradiction. So $(m_1 - 1)/n > x$, and thus $(m_1 - 1)/n \in (x, y)$. $\square$

We also should believe that the *irrational* numbers are dense in $\mathbb{R}$. There's a quite ridiculous proof of this fact, that used the irrationality of $\sqrt{2}$.

Claim 3.18. *$\mathbb{R} \setminus \mathbb{Q}$ is dense in $\mathbb{R}$.*

Proof: Let $x < y$ be given. There's a rational r in $(x/\sqrt{2}, y/\sqrt{2})$, by density of the rationals. But then $\sqrt{2}r \in (x, y)$, and $\sqrt{2}r$ is irrational! $\square$

[54] S stands for *Starbucks* — between any two points in New York City, there is a Starbucks.

[55] Or even more simply (as was pointed out by someone in class one day) $0 \in (x, y)$

3.9 A summary of the axioms of real numbers

To summarize this chapter: the real numbers, denoted $\mathbb{R}$, is a set of objects, which we call *numbers*, with

- two special numbers, 0 and 1, that are distinct from each other,[56]
- an operation $+$, *addition*, that combines numbers a, b to form the number $a + b$,
- an operation $\cdot$, *multiplication*, that combines a, b to form $a \cdot b$, and
- a set $\mathbb{P}$ of *positive* numbers,

that satisfies the following 13 axioms:

P1, Additive associativity For all a, b, c, $a + (b + c) = (a + b) + c$.

P2, Additive identity For all a, $a + 0 = 0 + a = a$.

P3, Additive inverse For all a there's a number $-a$ with $a + (-a) = (-a) + a = 0$.

P4, Additive commutativity For all a, b, $a + b = b + a$.

P5, Multiplicative associativity For all a, b, c, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

P6, Multiplicative identity For all a, $a \cdot 1 = 1 \cdot a = a$.

P7, Multiplicative inverse For all a, if $a \neq 0$ there's a number a^{-1} such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

P8, Multiplicative commutativity For all a, b, $a \cdot b = b \cdot a$.

P9, Distributivity of multiplication over addition For all a, b, c, $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

P10, Trichotomy law For every a exactly one of

- $a = 0$
- $a \in \mathbb{P}$
- $-a \in \mathbb{P}$.

holds.

P11, Closure under addition If $a, b \in \mathbb{P}$ then $a + b \in \mathbb{P}$.

P12, Closure under multiplication If $a, b \in \mathbb{P}$ then $ab \in \mathbb{P}$.

[56] See the footnote just after the proof of Corollary 3.11.

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

We can legitimately talk about *the* real numbers: there is (essentially) a *unique* structure that satisfies all the above axioms, and it can be explicitly constructed. We will say no more about this; but note that Spivak discusses this topic in Chapters 28 through 30 of his book.