LECTURE 1

MOTIVATION & INEQUALITIES

Luke Collins*
luke.collins.mt

What's the point of doing "Analysis"?

In mathematics, the term "analysis" does not mean the same thing as the colloquial term, i.e., it does not refer to the process of analysing. Rather, it refers to the branch of mathematics where we study inequalities and limits.[1] It is related to, but distinct from, calculus.

When first encountering calculus at A-level, you probably spent a lot of time learning different techniques for finding derivatives and integrals, and the notion of a limit would have been mentioned briefly, if at all. Calculus has prepared you, the student, for using mathematics without actually understanding why what you learned is true. But if you want to use, or teach, mathematics effectively, you cannot simply know *what* is true, but you must know *why* it is true as well. This course (and its sequels Analysis 2, 3 & 4) shows you, among other things, why calculus is true.

Think of it this way: engineers use calculus, but pure mathematicians use real analysis. By way of analogy: a car mechanic who knows how to change the oil, fix broken headlamps and change the battery, will only be able to do these simple tasks. He will be unable to work independently to diagnose and fix problems. A sixth form teacher who does not understand the definition

---

*Department of Mathematics, Faculty of Science, University of Malta.

[1] The reason analysis is called "analysis" is because historically, the terms analysis and synthesis were both used by Pappus of Alexandria in his *Synagoge* as the stages of formulation of a mathematical problem, and when Greek mathematics was revived in the Renaissance, many texts used the term in the title, such as L'Hôptial's *Analyse des Infiniment Petits pour l'Intelligence des Lignes Courbes*.

of the derivative or the Riemann integral may not be able to properly answer students' questions. To this day, I remember several nonsensical statements I heard from my teacher, who simply did not understand the concept of a limit, although he could technically "do" all the problems from A-level past papers.

Course structure. *Real* analysis chiefly deals with sequences, series and functions of real numbers. Other courses which fall under "analysis" include complex analysis, where you instead look at sequences, series and functions of complex numbers (these are surprisingly nicer and easier to deal with than their real counterparts), functional analysis, where instead functions are the basic object (rather than numbers), and harmonic analysis, which focuses mainly on waves (such as the sine function) and things like the Fourier transform.

A typical undergraduate training in real analysis includes:

- i. Basic properties of $\mathbb{R}$
- ii. Infinite sets
- iii. Sequences
- iv. Series
- v. Basic topology of $\mathbb{R}$
- vi. Limits and continuity
- vii. Derivatives
- viii. Local extrema
- ix. Riemann integration
- x. Sequences of functions
- xi. Functions of several variables
- xii. Higher order derivatives
- xiii. Differentials
- xiv. Implicit functions

In this course, we cover i–v. Analysis 2 covers vi–viii, Analysis 3 covers ix–x, and Analysis 4 covers xi–xiv.

Exercises. It is important to work through all the exercises provided, not only to reinforce what you have learned, but to also garner sufficient instincts for what is to come. It is not enough to be able to do the exercises—by the end of them, you should be able to do similar exercises *easily*, almost without thinking. This way, when you go on to more advanced topics, your focus will be entirely on the new material, and you will not sacrifice any of your brain's "processing power" to understand more basic steps.

Annotations. When a paragraph is annotated with a ☡ symbol, it means that you should read it slowly and carefully, since it contains some tricky ideas. When an exercise is annotated with a ☕ symbol, it is instructing you to pour yourself a cup of tea and dedicate some time to think about the problem, it might be harder than the others.

0. Basic Properties of The Real Numbers

In formal mathematics, we try to encode everything using sets—and when we say everything, we mean *everything*. There are very important reasons for this: we like to think that all mathematical statements live in the same axiomatic system (usually called the ZF(C) universe), this way, all of mathematics is built on only 8(+1) axioms. For instance, take the natural numbers,

$$\mathbb{N} = \{0, 1, 2, 3, 4, \dots\}.$$

What are these numbers exactly? Can we think of them as being "constructed" from a more basic kind of object? The answer is yes, and in mathematics, we choose sets to be our building blocks for everything else.

What are numbers exactly?

Formally, we define the set $\mathbb{N}$ as the *closure* of the set $\{\emptyset\}$ under the successor operation $x \mapsto x \cup \{x\}$. This means that $\mathbb{N}$ is the smallest set containing $\emptyset$, together with any other set(s) which can be obtained by repeatedly applying the operation $x \mapsto x \cup \{x\}$.[2] The successor of each natural number represents the next natural number in the usual order, so identifying 0 with the empty set, we have

$$\begin{aligned}
0 &= \emptyset \\
1 &= 0 \cup \{0\} = \emptyset \cup \{0\} = \{0\} = \{\emptyset\} \\
2 &= 1 \cup \{1\} = \{0, 1\} = \{\emptyset, \{\emptyset\}\} \\
3 &= 2 \cup \{2\} = \{0, 1, 2\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} \\
4 &= 3 \cup \{3\} = \{0, 1, 2, 3\},
\end{aligned}$$

and so on. In general, each natural number k ends up being structurally represented as the set of all its predecessors:

$$k = \{0, 1, 2, \dots, k-1\}.$$

There is nothing inherently special about this way of encoding $\mathbb{N}$. This is just one of many ways to construct a set which captures the behaviour of $\mathbb{N}$, namely, a set where we have

- a least element 0, (i.e., $\emptyset$), and
- a way to always "add 1" (i.e., the successor operation $x \mapsto x \cup \{x\}$).

[2] That this set exists is actually one of the 8(+1) axioms.

FIGURE 1: Set theoretic "meme"

When encoded this way, the natural numbers are called von Neumann ordinals. The two properties above capture entirely the essence of the natural numbers, and by constructing an object which behaves this way, we showed that it is possible to encode $\mathbb{N}$ with sets alone, and we don't need to think of numbers as autonomous "objects" in their own right.[3] One can define a notion of "addition" and "multiplication" and so on, all in terms of the underlying set representations, and prove things like $1 + 1 = 2$ (refer to figure 1). But of course in practice, we still think of the natural numbers as usual, only now safer in the knowledge that they are built on a more fundamental idea which we understand well. Similarly, we can construct $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ in terms of sets, but we will not get into that here. If you want to learn about the details of all this, the best way is to take the third year course on set theory.

So, since we're doing real analysis, we should begin by asking: what is a real number? Well, the (modern) answer is that a real number is some sort of elaborately constructed set which has all the desired properties as we explained above, but for our purposes, we can just say that a real number is a mathematical "object" which lives in the set $\mathbb{R}$ of real numbers. We don't really care (in this course) what they are *per se*, what we care about mainly is *how they behave*. For instance, we care about the fact that $1 + 1 = 2$, or that $x^2 \geqslant 0$ for all $x \in \mathbb{R}$. This is information about the behaviour of real numbers, not about the objects themselves: we are interested in descriptive questions, not ontological ones.

It turns out that all we need to know about $\mathbb{R}$ to fully understand its behaviour (and to start "doing" analysis) is that it is a *complete*, *totally ordered*, *field*.

Let's deal with these terms one by one, in reverse order, starting with fields: what is a field?

[3]This is analogous to how any photo, video or text on a computer is just 1s and 0s at the lowest level, but we still don't think about them in terms of 1's and 0's!

Fields

Definition 1 (Field)**.** Let F be a set, and let $+ \colon F \times F \to F$, $\cdot \colon F \times F \to F$ be two functions, usually called *addition* and *multiplication* respectively. The triple $(F, +, \cdot)$ is called a *field* if the following properties hold:

i. for all $x, y \in F$, $x + y \in F$, (Closure under $+$)
ii. for all $x, y, z \in F$, $x + (y + z) = (x + y) + z$, (Associativity under $+$)
iii. there exists $0 \in F$ s.t. for all $x \in F$, $x + 0 = x$, (Identity for $+$)
iv. for all $x \in F$, there exists $-x \in F$ s.t. $x + (-x) = 0$, (Inverses for $+$)
v. for all $x, y \in F$, $x + y = y + x$, (Commutativity under $+$)
vi. for all $x, y \in F$, $x \cdot y \in F$, (Closure under $\cdot$)
vii. for all $x, y, z \in F$, $x \cdot (y \cdot z) = (x \cdot y) \cdot z$, (Associativity under $\cdot$)
viii. there exists $1 \in F$ s.t. for all $x \in F$, $x \cdot 1 = x$, (Identity for $\cdot$)
ix. for all $x \in F \setminus \{0\}$, there exists $x^{-1} \in F$ s.t. $x \cdot x^{-1} = 1$, (Inverses for $\cdot$)
x. for all $x, y \in F$, $x \cdot y = y \cdot x$, (Commutativity under $\cdot$)
xi. for all $x, y, z \in F$, $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$, (Distributivity)

where the infix notation $x + y$ denotes the image $+(x, y)$ of x and y under the function $+$, and similarly $x \cdot y$ denotes $\cdot(x, y)$.

Remark 2 (Closure is obvious). Notice that i and vi are obvious from the definition of $+$ and $\cdot$, we state them here for emphasis. (All they say is that the images of $+$ and $\cdot$ are always in the codomain, which must be the case since they are functions.)

Remark 3 (Field = double abelian group). We can summarise properties i–v in definition 1 by saying that $(F, +)$ is an abelian group. Similarly, vi–x is just telling us that $(F, \cdot)$ is an abelian group (with the added caveat that the additive identity 0 is the only element which doesn't have an inverse).

The property xi is the only property which tells us how the operations of addition and multiplication behave together.

If you read definition 1 with $F = \mathbb{R}$ in mind, then most of the properties there are obviously true about the real numbers that we are used to working with; but what is important about this specific handful of properties is that they are the *only* properties we need to obtain all other algebraic properties we are used to using all the time (such as the fact that $(x + y)^2 = x^2 + 2xy + y^2$).

Remark 4 ($0 \neq 1$). Technically, in a field, there is nothing preventing us from having $0 = 1$, but the properties in the definition will force all other elements to be zero, so the whole field will just be the set $\{0\}$, which isn't very interesting.

So to be clear, if we let $F = \{0\}$, and define the functions $+\colon F \times F \to F$ and $\cdot : F \times F \to F$ by $0 + 0 = 0$ and $0 \cdot 0 = 0$, this is a legitimate (albeit trivial) example of a field. (This is usually called the *trivial field* or the *zero field*.) Going forward, however, we always assume 1 and 0 are not equal, so our fields contain at least two elements (eventually, for this course, we will only care about $\mathbb{R}$ anyway, where this is definitely the case!).

Notation (Algebraic conventions)**.** We relax the rigid notation introduced in definition 1 as follows.

- The product $x \cdot y$ is written simply as xy.
- The sum $x + -y$ is written simply as $x - y$.
- The product $x \cdot y^{-1}$ is written as $\frac{x}{y}$. We call x the *numerator* and y the *denominator*.
- $x + (y + z) = (x + y) + z$ is denoted simply as $x + y + z$, and similarly $x \cdot (y \cdot z) = (x \cdot y) \cdot z$ by xyz. Other similar relaxations are made when brackets are not necessary.

So for example, we write

$$\frac{3x - 5y}{1 + xyz} \qquad \text{instead of} \qquad (3 \cdot x + -(5 \cdot y)) \cdot (1 + x \cdot (y \cdot z))^{-1}.$$

Example 5 (A finite field). The real numbers are not the only field.[4] Consider the set of elements $\{o, i, a, b, c\}$, and define "addition" and "multiplication" by the following tables:

$+$	o	i	a	b	c
o	o	i	a	b	c
i	i	a	b	c	o
a	a	b	c	o	i
b	b	c	o	i	a
c	c	o	i	a	b

$\times$	o	i	a	b	c
o	o	o	o	o	o
i	o	i	a	b	c
a	o	a	c	i	b
b	o	b	i	c	a
c	o	c	b	a	i

[4] It would be pointless to have such a definition if $(\mathbb{R}, +, \cdot)$ was the only field!

Although it would be quite tedious, you can verify that each of the properties I–XI is satisfied by these two operations, so this defines a field, where $0 = o$ and $1 = i$, since from the table, we see that $x + o = x$ and $x \cdot i = x$ for all x. We also have $i^{-1} = i$, since $i \cdot i = i$, $a^{-1} = b$ since $a \cdot b = i = 1$, and similarly $b^{-1} = a$ and $c^{-1} = c$.

Thus we have (for example) that $\frac{c}{b} = c \cdot b^{-1} = c \cdot a = b$, and

$$\begin{aligned}
(a+b)(b-c) &= (a+b)(b+(-c)) && \text{(by definition of } x-y\text{)}\\
&= (a+b)(b+i) && (-c = i \text{ since } c+i=o)\\
&= (a+b)b + (a-b)i && \text{(by XI of definition 1)}\\
&= b(a+b) + i(a+b) && \text{(by X of definition 1)}\\
&= ba + bb + a + b && \text{(by XI and VIII of definition 1)}\\
&= i + c + a + b && \text{(from multiplication table)}\\
&= o + a + b && \text{(from addition table)}\\
&= a + b && \text{(by III of definition 1)}\\
&= o && \text{(from addition table)}
\end{aligned}$$

In fact, there is a more direct way to notice this, since the first bracket $a+b = o$, which gives us directly that $(a + b) \cdot (b - c) = o \cdot (b - c) = o$, since from the table we see that $o \cdot x = o$ for all x.

Fields obey some easy general facts (which $\mathbb{R}$ automatically inherits if we can prove them in general for any field F). We've already pointed out one of them in the example we saw: that $0 \cdot x = 0$ for all x. This is quite interesting: it is telling us how the *additive identity*, 0, behaves under *multiplication*. How can we prove that it is true in any field?

Well, the only property in definition 1 which links + and $\cdot$ is XI, so that must play some role in the proof. Indeed, even though the fact is pretty easy and familiar to us, it's not trivial to prove it using just the defining properties of a field.

Let us give a proposition which lists a few basic properties of fields. We prove that $0x = 0$ in (vi).

Proposition 6 (Field Properties)**.** *Suppose F is a field, and let $a, x, y \in F$. Then*

(i) The numbers 0 and 1 are unique,[5]

[5] In the sense that there is only one element (which we label as 0) in F satisfying III of definition 1, and similarly only one element (which we label as 1) in F satisfying VIII.

(ii) *For each* x*, there is only one* $-x$ *and* x^{-1} *(where* $x \neq 0$ *for the latter),*

(iii) $-(-x) = x$,

(iv) $x + a = y + a \implies x = y$,

(v) *If* $a \neq 0$*, then* $xa = ya \implies x = y$,

(vi) $x \cdot 0 = 0 \cdot x = 0$,

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

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

(ix) $(-1) \cdot x = -x$,

(x) $(-1) \cdot (-1) = 1$.

Proof. For (i), suppose that there are two zeros, 0 and $\hat{0}$ both satisfying III, where $0 \neq \hat{0}$. Then by III,

$$0 = 0 + \hat{0} = \hat{0} + 0 = \hat{0},$$

contradicting that $0 \neq \hat{0}$. Replacing 0 and $\hat{0}$ with 1 and $\hat{1}$ above, the proof for 1 is the same by VIII.

For (ii), again suppose we have two different minus x's, $-x$ and $\ominus x$, both satisfying IV. Then by II, III and IV,

$$-x = -x + 0 = -x + (x + \ominus x) = (-x + x) + \ominus x = 0 + \ominus x = \ominus x,$$

contradicting that $-x \neq \ominus x$. A similar argument proves the uniqueness of x^{-1} for each non-zero $x \in F$.

For (iii), we have

$$\begin{aligned}
-(-x) &= -(-x) + 0 && \text{(by III)}\\
&= -(-x) + (-x + x) && \text{(by IV)}\\
&= (-(-x) + -x) + x && \text{(by II)}\\
&= 0 + x && \text{(by IV)}\\
&= x && \text{(by III)}
\end{aligned}$$

as required. For (iv), suppose $x + a = y + a$. Then

$$\begin{aligned}
x &= x + 0 && \text{(by III)}\\
&= x + (a + -a) && \text{(by IV)}\\
&= (x + a) - a && \text{(by II)}\\
&+ (y + a) - a && \text{(by assumption)}\\
&= y + (a - a) && \text{(by II)}\\
&= y + 0 && \text{(by IV)}\\
&= y && \text{(by III)}
\end{aligned}$$

as required. A similar argument proves (v).

For (vi), we have

$$0 + x \cdot 0 = x \cdot 0 = x \cdot (0 + 0) = (x \cdot 0) + (x \cdot 0)$$

by XI. Hence by (iv), $0 = x \cdot 0 = 0 \cdot x$ by X.

For (vii), observe that

$$x \cdot y + x \cdot (-y) = x \cdot (y + -y) = x \cdot 0 = 0$$

by XI and (vi). Hence by (ii), $x \cdot (-y)$ is the unique inverse $-(x \cdot y)$ of $x \cdot y$. Similarly $(-x) \cdot y = -(x \cdot y)$.

Now for (viii), we have

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

by (ii) and by (iii).

For (ix), observe that

$$x + (-1) \cdot x = x \cdot 1 + x \cdot (-1) = x \cdot (1 + -1) = x \cdot 0 = 0$$

by VIII, XI and (vi). Thus by (ii), $(-1) \cdot x$ is the unique inverse $-x$ of x.

Finally (x) follows by (ix) with $x = -1$, and we get

$$(-1) \cdot (-1) = -(-1) = 1$$

by (iii). □

Notice that proving these "obvious" results is quite similar to playing chess in some sense; we know what tile (on the chessboard) we want to get to, but we can only make valid moves according to the rules of the game. In our case, the "valid moves" are the properties in definition 1, and subsequent results we established from them.

Exercise 7. Suppose F is a field. Prove the following for all $x, y \in F$ where $x, y \neq 0$. You may use any of the facts from proposition 6.

(i) $-0 = 0$,
(ii) $1^{-1} = 1$,
(iii) $(x^{-1})^{-1} = x$,
(iv) $(x \cdot y)^{-1} = x^{-1} \cdot y^{-1}$.

Some consequences of our fractional notation $\frac{x}{y}$ are the familiar properties of proposition 8. Notice that the proofs of these are quite a bit easier than those of proposition 6: it's the more basic facts which are harder to prove.

Proposition 8. *Suppose F is a field, and suppose $w, x, y, z \in F$. Then*

(i) $\dfrac{x}{y} \cdot \dfrac{w}{z} = \dfrac{xw}{yz}$, *(ii)* $\dfrac{xy}{xz} = \dfrac{y}{z}$,

(iii) $\dfrac{x}{z} + \dfrac{y}{z} = \dfrac{x+y}{z}$, *(iv)* $\dfrac{x}{y} + \dfrac{w}{z} = \dfrac{xz+yw}{yz}$,

where the variables are never zero when they are in the denominator.

Proof. (i) follows from exercise 7(iv) since

$$\begin{aligned} \frac{x}{y} \cdot \frac{w}{z} &= x \cdot y^{-1} \cdot w \cdot z^{-1} \\ &= (xw) \cdot (y^{-1} \cdot z^{-1}) = xw \cdot (yz)^{-1} = \frac{xw}{yz}. \end{aligned}$$

(ii) then follows easily from (i):

$$\frac{xy}{xz} = \frac{x}{x} \cdot \frac{y}{z} = x \cdot x^{-1} \cdot \frac{y}{z} = 1 \cdot \frac{y}{z} = \frac{y}{z}.$$

(iii) uses proposition 6(XI):

$$\begin{aligned} \frac{x}{z} + \frac{y}{z} &= x \cdot z^{-1} + y \cdot z^{-1} \\ &= z^{-1} \cdot x + z^{-1} \cdot y \\ &= z^{-1} \cdot (x+y) \\ &= (x+y) \cdot z^{-1} = \frac{x+y}{z}. \end{aligned}$$

Finally (iv) follows by (ii) and (iii):

$$\frac{x}{y} + \frac{w}{z} = \frac{zx}{zy} + \frac{yw}{yz} = \frac{xz}{yz} + \frac{yw}{yz} = \frac{xz+yw}{yz}.$$ □

Exercise 9. Suppose F is a field, and suppose $x, y \in F$, where $y \neq 0$. Prove that $-\frac{x}{y} = \frac{-x}{y} = \frac{x}{-y}$.

Notation (Decimal Notation)**.** Notice that 0 and 1 are symbols which are defined *by their behaviour* in II and VIII respectively. What about the other numbers like 2, 3, -5 and $\frac{4}{5}$, how do they fit in to our framework?

Well, unsurprisingly, the positive whole numbers are defined as follows:

$$\begin{aligned} 2 &= 1 + 1 \\ 3 &= 2 + 1 = 1 + 1 + 1 \\ 4 &= 3 + 1 = 1 + 1 + 1 + 1 \\ 5 &= 4 + 1 = 1 + 1 + 1 + 1 + 1 \\ 6 &= 5 + 1 = 1 + 1 + 1 + 1 + 1 + 1 \\ &\vdots \end{aligned}$$

Thus -5 is just

$$-(1 + 1 + 1 + 1 + 1),$$

and $\frac{4}{5}$ is

$$(1 + 1 + 1 + 1) \cdot (1 + 1 + 1 + 1 + 1)^{-1}.$$

But what *is* this? Like we mentioned before: we don't care! Or rather: for the purposes of this course, this *is* $\frac{4}{5}$ *by definition*. Indeed, all we care about is how it behaves: algebraically, the most important thing this number achieves is that, when we multiply it by 5 (i.e., $1 + 1 + 1 + 1 + 1$), the answer is 4 (i.e., $1 + 1 + 1 + 1$). In other words, unless we get into the set theory details, the number $\frac{4}{5}$ is as simple as it gets: it is just a short hand for the expression above, which uses 1s alone.

You might say "this is 0.8", but if you think back to primary school when you first learnt decimal notation, you quickly realise that this is just $\frac{8}{10}$, i.e., $8 \cdot 10^{-1}$. Since $8 = 4 \cdot 2$ and $5 = 10 \cdot 2$,[6] this is just $\frac{4 \cdot 2}{5 \cdot 2} = \frac{4}{5}$ by proposition 8. In general, the decimal number

$$a_n a_{n-1} \ldots a_1 a_0 \,.\, a_{-1} a_{-2} \ldots$$

i.e., the decimal number with digits $a_n \cdots a_0$ before the decimal point, and $a_{-1} a_{-2} \cdots$ after, represents the number

$$10^n a_n + 10^{n-1} a_{n-1} + \cdots + 10 a_1 + a_0 + \frac{a_{-1}}{10} + \frac{a_{-2}}{10^2} + \cdots = \sum_{k=-\infty}^{n} 10^k a_k.$$

[6] two facts you can prove by expanding $(1+1+1+1)(1+1)$ and $(1+1)(1+1+1+1+1)$ respectively.

We will get into the details of when the series has infinitely terms (and what that means exactly) later on, but for finitely many terms, this should be straightforward. For instance, 3.142 is

$$3 + \frac{1}{10} + \frac{4}{100} + \frac{2}{1000}$$

which can be expressed using only sums of 1's. In fact, the rationals, seen as a subset of $\mathbb{R}$, are just

$$\mathbb{Q} = \left\{ \pm \frac{\overbrace{1 + \cdots + 1}^{\text{any number} \geqslant 0 \text{ of 1's}}}{\underbrace{1 + \cdots + 1}_{\text{any number} > 0 \text{ of 1's}}} \right\}.$$

But what about the so-called irrational numbers, like π? We'll get to those in a moment. But before, we need to discuss the second term: we said that $\mathbb{R}$ is a *totally ordered* field. Let's explore this idea next.

Remark 10. Notice that fractional and decimal notation can be used in *any field*, not just $\mathbb{R}$ or $\mathbb{Q}$. For instance, in example 5, $0 = o$, $1 = i$, $2 = a$, $3 = b$ and $4 = c$. Similarly, $5 = 4 + 1 = c + i = o$. In other words, $5 = 0$. Consequently, $6 = 1$, $7 = 2$, etc.[7] Thus, in this field, $\frac{4}{5}$ doesn't exist, since $5 = 0$, but we can make sense of, for example, $\frac{7}{3} = \frac{2}{3} = 2 \cdot 3^{-1} = a \cdot b^{-1} = a \cdot a = b$.

Exercise 11. Unless told otherwise, you may only use the fact that $\mathbb{R}$ is a field (i.e., the properties in definition 1) in your answers.

(i) Why is $x + x = 2x$?

(ii) Solve the following equations, justifying each step you make by referencing one of definition 1, proposition 6 or 8.

(a) $2x + 1 = 3$, (b) $\frac{x}{2} + \frac{x}{3} = 3 + \frac{1}{3}$, (c) $x^2 - 5x + 6 = 0$.

(iii) Prove that $(a + b)(c + d) = ac + bc + ad + bd$.

(iv) In remark 4, I said that the properties in definition 1 force all other elements to be zero. Prove it: i.e., prove that if $1 = 0$, $F = \{0\}$.

(v) Rigorously prove that $\frac{6}{3} = 2$.

[7] In fact, this is isomorphic to the field $\mathbb{Z}/5\mathbb{Z}$ which you probably have mentioned in the study-unit you're doing on groups.

Order

Now we're going to talk about the order properties of $\mathbb{R}$. We need the following definition first:

Definition 12 (Ordered field)**.** Let F be a field having operations $+$ and $\cdot$ and additive identity 0. We say that F is an *ordered field* if there exists a non-empty set $P \subseteq F$ for which the following properties hold:

(i) if $x, y \in P$, then $x + y \in P$ and $x \cdot y \in P$, (Closure of P)

(ii) for each $x \in F$, exactly one of the following is true: (Trichotomy)
◇ $x \in P$ ◇ $x = 0$ ◇ $-x \in P$

The set P is called the *positive cone* of F. If $x \in P$, we say that x is *positive*.

When F is $\mathbb{R}$, we imagine that P is the usual set of positive numbers. We'll see that this is indeed the case. A few consequences of this definition:

Proposition 13. *Suppose F is a totally ordered field with additive identity 0, multiplicative identity 1, and positive cone P. Then:*

(i) *For every $x \in F$ with $x \neq 0$, $x \neq -x$,*

(ii) *For every $x \in F$ with $x \neq 0$, we have $x^2 \in P$,*

(iii) $1 \in P$*,*

(iv) $1 \neq 0$*.*

Proof. (i) follows immediately from definition 12(ii), since if $x \neq 0$, then we cannot have $x = -x$ since both would be in P (*exactly one* should be true).

For (ii), if $x \neq 0$, then either $x \in P$ or $-x \in P$. If $x \in P$, then by closure (definition 12(i)), $x^2 \in P$. Otherwise, $-x \in P$, but then again by closure, $(-x)(-x) = x^2 \in P$ (proposition 6(viii)).

Now for (iii), $1 \in P$ since $1 = 1^2$, and we apply (ii).

Even though we've been assuming (iv) throughout (recall remark 4), we don't need to assume it to conclude that $1 \neq 0$ in this case, since this follows from the fact that $0 \notin P$ by definition 12, and that P is non-empty; so it must contain some non-zero element, and therefore $1 \neq 0$ (recall exercise 11(iv)). □

Now that we have established the notion of positive numbers, we can introduce the familiar order relation symbols.

Notation (Order relations)**.** Suppose F is an ordered field, and let $x, y \in F$. We write $x > y$ or $y < x$ if $x - y$ is positive. Similarly, we write $x \geqslant y$ or $y \leqslant x$ if $x - y$ is positive or $x = y$.

A few other consequences of this notation:

Proposition 14. *Let F be an ordered field, and suppose $a, x, y \in F$. Then:*

(i) $x > 0$ *if and only if* $x \in P$,

(ii) $x > 0$ *or* $x = 0$ *or* $x < 0$,

(iii) if $x < y$, *then* $x + a < y + a$,

(iv) if $x \leqslant y$, *then* $x + a \leqslant y + a$,

(v) if $x > 0$ *and* $y > 0$ *then* $xy > 0$,

(vi) if $x > 0$, *then* $x^{-1} > 0$,

(vii) if $a > 0$, $x < y \iff ax < ay$,

(viii) if $0 < a < b$, *then* $0 < \frac{1}{b} < \frac{1}{a}$.

Exercise 15. (i) Prove proposition 14.

(ii) Is the field from example 5 an ordered field?

Before we discuss the final property of the reals, *completeness*, let's see an important inequality. We shall state in terms of the following useful function.

Definition 16 (Absolute Value)**.** Let F be an ordered field. Then for $x \in F$, we define

$$|x| = \begin{cases} x & \text{if } x \geqslant 0, \\ -x & \text{otherwise.} \end{cases}$$

Remark 17. The absolute value function is convenient since it allows us to write things such as $-r < x < r$ more succinctly as $|x| < r$, or $a - \epsilon < x < a + \epsilon$ as $|x - a| < \epsilon$.

A few basic properties:

Proposition 18. *Let F be an ordered field, and let $x, y \in F$. Then*

(i) $|xy| = |x||y|$,

(ii) $x \leqslant |x|$,

(iii) $x^2 = |x|^2$,

(iv) if $x^2 < y^2$ *then* $|x| < |y|$.

Proof. This proof is a bit tedious, but straightforward: we simply split into cases for x, y being positive, negative or zero.

CASE I: $x > 0$.
In this case, $|x| = x$ by definition of $|\cdot|$, so (ii) and (iii) are obvious. For (i) and (iv), we break into subcases.

I.a. $y > 0$. In this case, $xy > 0$ by closure of P, so $|xy| = xy = |x||y|$, which proves (i). Similarly, if $x^2 < y^2$, then $(y+x)(y-x) > 0$. Now $y+x > 0$, so we can cancel (by proposition 14(vii)) to get $y - x > 0$, i.e., $|x| < |y|$, proving (iv).

I.b. $y = 0$. In this case, $xy = 0$, so $|xy| = |0| = 0 = |x| \cdot 0 = |x||y|$, proving (i), and the hypothesis (iv) is never true, since it becomes $x^2 < 0$, contradicting proposition 13(ii). Thus (iv) is true vacuously.[8]

I.c. $y < 0$. In this case, $xy < 0$, so $|xy| = -xy$ and $|y| = -y$, which gives $|xy| = -xy = x(-y) = |x||y|$, proving (i). For (iv), $x^2 < y^2$ becomes $(y-x)(y+x) > 0$, and this time we know that $y - x < 0$, so cancellation gives $y + x < 0$, i.e., $|x| < |y|$, proving (iv).

CASE II: $x = 0$.
In this case, $|xy| = 0 = |x||y|$, proving (i), $x = 0 = |x|$, proving (ii), and $x^2 = 0 = |x|^2$, proving (iii). The conclusion of (iv) is obvious since it just says that $|y| > 0$, which is true by definition of $|\cdot|$.

CASE III: $x < 0$. We leave this case as an exercise. □

The following inequality is essential in analysis, and will play a role in many upcoming proofs:

Theorem 19 (The Triangle Inequality)**.** *Let F be an ordered field, and let $x, y \in F$. Then*

$$|x+y| \leqslant |x| + |y|.$$

Proof. We could also prove this by splitting into cases, but there is a nice way which avoids it. Consider:

$$\begin{aligned} & xy \leqslant |xy| \\ \implies \quad & 2xy \leqslant 2|x||y| \\ \implies \quad & 2xy + x^2 + y^2 \leqslant 2|x||y| + x^2 + y^2 \\ \implies \quad & (x+y)^2 \leqslant (|x|+|y|)^2 \\ \implies \quad & |x+y| \leqslant |x| + |y|. \end{aligned}$$

□

In some contexts, we would like to bound the quantity $|x+y|$ from below. That's when the following proves useful:

[8] An implication $p \Rightarrow q$ is *vacuously true* if the antecedent p is never true.

Proposition 20 (Reverse Triangle Inequality)**.** *Let F be an ordered field, and let $x, y \in F$. Then*

$$|x| - |y| \leqslant |x + y|.$$

Proof. If we let $X = x + y$ and $Y = -y$, then $X + Y = x$, and the thing we want to prove is $|X + Y| - |Y| \leqslant |X|$, i.e., $|X + Y| \leqslant |X| + |Y|$. This is just the ordinary triangle inequality. □

We can express the "distance" between two numbers x and y on the number line as $d(x, y) = |x-y|$. A set M together with a function $d\colon M \times M \to [0, \infty)$ is called a *metric space*, there's a whole study-unit dedicated to them in the third year of your degree to look forward to.

Proposition 21 (Metric space properties)**.** *Let F be an ordered field, and let $x, y, z \in F$. If $d(x, y)$ denotes $|x - y|$, then we have:*

(i) $d(x, y) = 0$ *if and only if* $x = y$*,*

(ii) $d(x, y) = d(y, x)$*,*

(iii) $d(x, y) \leqslant d(x, z) + d(z, y)$*.*

Proof. For (i), we have

$$d(x, y) = 0 \iff |x - y| = 0 \iff x - y = 0 \iff x = y,$$

and for (ii),

$$d(x, y) = |x - y| = |y - x| = d(y, x).$$

Finally for (iii), we want to show that $|x - y| \leqslant |x - z| + |z - y|$. If we let $X = x - z$ and $Y = z - y$, notice that $X + Y = x - y$, so the inequality we want to prove is $|X + Y| \leqslant |X| + |Y|$, but this is the triangle inequality. □

Let's conclude this section with a proposition which starts to capture the way we think about "closeness" in analysis.

We say that x and y are *arbitrarily close* if for all $\epsilon > 0$, their distance is less than ϵ, i.e., $d(x, y) < \epsilon$. If two numbers are arbitrarily close, then they must be equal:

Proposition 22. *Let F be an ordered field, and let $x, y \in F$. Then $x = y$ if and only if $|x - y| < \epsilon$ for all $\epsilon > 0$.*

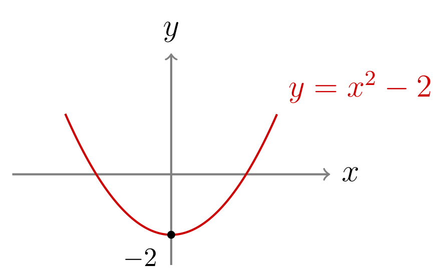
Proof. In one direction, this is obvious: if $x = y$, then clearly $|x - y| = 0 < \epsilon$ for any $\epsilon > 0$. For the converse, we proceed by contradiction: suppose that x and y are arbitrarily close, i.e., we have $|x - y| < \epsilon$ for any $\epsilon > 0$ that we want, but that $x \neq y$. Since $x \neq y$, the distance $|x - y|$ between them is positive. Now for the contradiction, if we let ϵ be this distance, then we get that $|x - y| < |x - y|$, which is nonsense. □

Completeness

So far, it seems that the fact that $\mathbb{R}$ is a field gives us all the algebraic properties we are used to working with, and the fact that it is ordered gives us an idea of where numbers are relative to each other on a number line. So why do we need this third thing, completeness?

The best way to answer this question is to look at an ordered field which isn't complete: namely, the rational numbers $\mathbb{Q}$. Indeed, in every result we've seen so far, we haven't assumed that we are working with real numbers; instead we've been deliberately general, starting everything with "let F be an ordered field". This way, everything we've seen so far also applies to $\mathbb{Q}$ (and any other ordered fields we may come across). $\mathbb{Q}$ is in fact the "smallest" ordered field.

Now, the issue with $\mathbb{Q}$ is that it has "holes", and it is because of these holes that we say the rationals are not complete. The holes are not as gaping and obvious to spot as they are in a set like $\mathbb{Z}$; indeed, between any two rationals p and q, there is another rational (an easy one would be $\frac{p+q}{2}$), so the holes aren't something we can immediately visualise. To detect them, it's best to consider something like this:



If we only worked with rational numbers, then this graph *doesn't intersect the x-axis*, since the intercepts are at $x = \pm\sqrt{2}$, which aren't rational numbers. We can get rational numbers very close to $\sqrt{2}$, e.g., by taking the first few digits of the decimal representation of $\sqrt{2} \approx 1.41421356237$. This is rational because it equals

$$\frac{141\,421\,356\,237}{100\,000\,000\,000},$$

but it is not exactly $\sqrt{2}$. It is because of the discomfort of these situations (and others where these holes cause problems) that we choose to work with a larger number system where these holes are “filled in”. Now we will develop the theory and mathematical tools which allows us to be more precise about what we mean when we say they are “filled in”.

Let’s start with a few definitions.

Definition 23 (Bounds)**.** Let F be an ordered field, and let $A \subseteq F$. Then we say that A is:

(i) *bounded above* if there exists $M \in F$ such that $x < M$ for all $x \in A$,

(ii) *bounded below* if there exists $M \in F$ such that $M < x$ for all $x \in A$,

(iii) *bounded* if it is both bounded above and below.

In case (i) we call the number M an *upper-bound*, in case (ii) we call it a *lower-bound*, and in case (iii), we simply call it a *bound*.

Definition 24 (Max and min)**.** Let F be an ordered field, and let $A \subseteq F$. Then we say that:

(i) $M \in A$ is a *maximal element* or a *maximum* of A if $x \leqslant M$ for all $x \in A$,

(ii) $m \in A$ is a *minimal element* or a *minimum* of A if $x \geqslant m$ for all $x \in A$.

Remark 25. Maximum or minimum elements do not always exist in a set, but when they do exist, they are unique. Indeed, suppose A has two maximum elements: M_1 and M_2. Then by definition they are both in A, and therefore $M_1 \leqslant M_2$ and $M_2 \leqslant M_1$ by definition of maxima, which means they must be the same element. A similar argument proves the uniqueness of minima (when they exist).

You might be tempted to guess that maxima exist when a set is bounded above: but this isn’t always the case, for instance, take the open interval $A = (0, 1)$. This set has no maximum nor minimum. By contrast, the closed interval $[0, 1]$ maximum 1 and minimum 0.

Notation (Max and min)**.** Since when they exist, they are well-defined, we introduce the notation $\max(A)$ and $\min(A)$ for the maximum and minimum of the set A respectively.

If we return our attention to the open interval $(0, 1)$, we see that it doesn’t have a maximum or minimum, but 0 and 1 are distinguished lower- and upper-bounds respectively. This leads to the following more subtle definition.

Definition 26 (Supremum)**.** Let F be an ordered field, and let $A \subseteq F$. The number $\alpha \in \mathbb{R}$ is said to be a *supremum* or a *least upper-bound* of A if:

(i) α is an upper-bound of A, and

(ii) $\alpha \leqslant u$ for all upper-bounds u of A.

Adapting the argument we used for maxima and minima, we easily see that, when a supremum exists, it must be unique. The analogous definition for lower-bounds is:

Definition 27 (Infimum)**.** Let F be an ordered field, and let $A \subseteq F$. The number $\beta \in \mathbb{R}$ is said to be an *infimum* or a *greatest lower-bound* of A if:

(i) β is a lower-bound of A, and

(ii) $\beta \geqslant \ell$ for all lower-bounds ℓ of A.

Unsurprisingly, just as with suprema, when an infimum exists, it is unique.

Notation (Sup and Inf)**.** For a subset A of an ordered field F, we introduce the notation $\sup(A)$ and $\inf(A)$ for the supremum and infimum of A respectively.

So notice, even though the set $(0, 1)$ doesn't have a maximum or a minimum, it has $\sup(0, 1) = 1$ and $\inf(0, 1) = 0$.

Exercise 28. In an ordered field F, show that when a set $A \subseteq F$ has a maximum (resp. minimum), then A also has a supremum (resp. infimum) and they are equal.

Finally, we introduce the definition which will fill in the "gaps" for us:

Definition 29 (Least upper-bound property)**.** An ordered field F is said to have the *least upper-bound property* if every non-empty subset $A \subseteq F$ which is bounded above has a supremum.

The axiom we add to our framework is:

Completeness Axiom. $\mathbb{R}$ has the least upper-bound property.

This axiom about $\mathbb{R}$ will fill in all the gaps for us. For instance, if we consider the set

$$A = \{x \in \mathbb{R} : x^2 < 2\},$$

then the completeness axiom implies that $\alpha = \sup A$ exists. In a bit, we'll show that $\alpha^2 = 2$ (so that $\alpha = \sqrt{2}$ by definition). But we will need to use the

Archimedean property, which we shall see in the next section.

Before we conclude the section, let's give an equivalent formulation of these ideas in terms of infima, as well as a few useful propositions.

Definition 30 (Greatest lower-bound property)**.** An ordered field F is said to have the *greatest lower-bound property* if every non-empty subset $A \subseteq F$ which is bounded below has an infimum.

Proposition 31. *Let F be an ordered field, let $A \subseteq F$, and let $\alpha \in F$ be an upper-bound of A. Then $\alpha = \sup A$ if and only if for every $\epsilon > 0$, there exists $x \in A$ such that $\alpha - \epsilon < x$.*

Proof. If $\alpha = \sup A$, then $\alpha - \epsilon$ cannot be an upper-bound of A for any $\epsilon > 0$ (otherwise α won't be the supremum). Thus there must be some $x \in A$ such that $\alpha - \epsilon < x$.

Conversely, suppose that for every $\epsilon > 0$, there exists $x \in A$ such that $\alpha - \epsilon < x$, and for contradiction, suppose that α is not the supremum. Then there exists $u < \alpha$ such that u is an upper-bound of A. Taking $\epsilon = \alpha - u > 0$, we get that $u < x$ for some $x \in A$, which contradicts that u is an upper-bound. □

Proposition 32. *Let F be an ordered field, let $A \subseteq F$, and let $\beta \in F$ be a lower-bound of A. Then $\beta = \inf A$ if and only if for every $\epsilon > 0$, there exists $x \in A$ such that $x < \beta + \epsilon$.*

Notation (Minkowski Operations)**.** Let F be an ordered field. For subsets $A, B \subseteq F$ and $c \in F$, we define the following notations:

$$\begin{aligned} A + B &= \{a + b : a \in A \text{ and } b \in B\} \\ A - B &= \{a - b : a \in A \text{ and } b \in B\} \\ -A &= \{-a : a \in A\} \\ cA &= \{ca : a \in A\} \\ AB &= \{ab : a \in A \text{ and } b \in B\} \end{aligned}$$

Careful not to confuse $A - B$ and $A \setminus B$. Moreover, despite the fact that the definitions are straightforward, some facts we might be used to applying do not hold, e.g., $A + A \neq 2A$.

Proposition 33. *Let F be an ordered field, let $A, B \subseteq F$ and $c \in F$. Then, assuming the sets are such that the necessary suprema and infima exist,*

(i) If $A \subseteq B$ then $\sup A \leqslant \sup B$ and $\inf A \geqslant \inf B$,

(ii) $\sup(A+B) = \sup A + \sup B$ *and* $\inf(A+B) = \inf A + \inf B$*,*

(iii) $\sup(A-B) = \sup A - \inf B$ *and* $\inf(A-B) = \inf A - \sup B$*,*

(iv) $\sup(-A) = -\inf A$ *and* $\inf(-A) = -\sup A$*,*

(v) If $c > 0$*,* $\sup(cA) = c\sup A$ *and* $\inf(cA) = c\inf A$*,*

(vi) If A *and* B *consists of non-negative elements,* $\sup(AB) = (\sup A)(\sup B)$ *and* $\inf(AB) = (\inf A)(\inf B)$*.*

Proof. We prove (ii), and leave the rest as exercises. Let $s = \sup(A)$ and $t = \sup(B)$. Then, being upper-bounds, we have $a \leqslant s$ for all $a \in A$ and similarly $b \leqslant t$ for all $b \in B$. Consequently, $x \leqslant s + t$ for all $x \in A + B$, so $s + t$ is an upper-bound.

Now, suppose $s + t$ is not the least upper-bound, so by proposition 31, there exists $\epsilon > 0$ such that $x \leqslant s + t - \epsilon$ for all $x \in A + B$. Thus, for any $a \in A$ and $b \in B$, we have

$$a = a + b - b \leqslant s + t - \epsilon - b \leqslant s - \epsilon,$$

i.e., $s-\epsilon$ is an upper-bound of A, contradicting that s is the least upper-bound. Hence, $s + t = \sup(A + B)$. □

Proposition 34 (LUP ⇔ GLP)**.** *An ordered field* F *has the least upper-bound property if and only if it has the greatest lower-bound property.*

Exercise 35. (i) Prove that if a set has a supremum, then it is unique.

(ii) Prove proposition 32.

(iii) Prove proposition 33.

(iv) Prove proposition 34.

The Archimedean Property

The idea behind the Archimedean property is to exclude the possibility of elements which are "infinitely large" or "infinitesimally small". Indeed, we say that x is *infinitesimal* with respect to y, or that y is *infinite* with respect to x, if for every $n \in \mathbb{N}$,

$$\underbrace{x + \cdots + x}_{n \text{ times}} < y.$$

In words, x is so small (or y is so big) that no matter how many x's we add, we can never exceed the value of y.

An ordered field F is Archimedean if this doesn't happen for any $x, y \in F$. For our purposes, we will prefer to formulate the Archimedean property as follows:

Theorem 36 (Archimedean Property for $\mathbb{R}$)**.** *For every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $\frac{1}{N} < \epsilon$.*

This is equivalent to what we mentioned earlier since it says that we can always find N such that $N\epsilon > 1$, i.e., N copies of the small number ϵ exceeds 1.

Proof. Let $\epsilon > 0$, and consider the set $A = \{n\epsilon : n \in \mathbb{N}\}$. If the theorem is false, then 1 is an upper-bound of A. Thus $\alpha = \sup A$ exists. Now by proposition 31, $\alpha - \epsilon < N\epsilon$ for some $N\epsilon \in A$. But then $\alpha < (N+1)\epsilon \in A$, which is a contradiction, since α is an upper-bound of A. □

Remark 37 ($\sqrt{2} \in \mathbb{R}$). Let's show that $\alpha = \sup\{x \in \mathbb{R} : x^2 < 2\}$ satisfies $\alpha^2 = 2$.

Indeed, suppose (for contradiction) that $\alpha^2 < 2$. Then $2 - \alpha^2 > 0$, and

$$\left(\alpha + \tfrac{1}{N}\right)^2 = \alpha^2 + \tfrac{2}{N}\alpha + \tfrac{1}{N^2} \leqslant \alpha^2 + \tfrac{2\alpha+1}{N},$$

which means that by the Archimedean property, we can find N large enough such that $\frac{2\alpha+1}{N} > 2 - \alpha^2$, giving us that

$$\left(\alpha + \tfrac{1}{N}\right)^2 < 2,$$

i.e., $\alpha + \frac{1}{N} \in A$. This contradicts that $\alpha = \sup A$.

Next, suppose (again, for contradiction) that $\alpha^2 > 2$. Then

$$\left(\alpha - \tfrac{1}{N}\right)^2 = \alpha^2 - \tfrac{2}{N}\alpha + \tfrac{1}{N^2} > \alpha^2 - \tfrac{2\alpha}{N},$$

and if N is large enough such that $\frac{2\alpha}{N} < \alpha^2 - 2$ (by the Archimedean property),

$$\left(\alpha - \tfrac{1}{N}\right)^2 > 2,$$

in other words, $\alpha - \frac{1}{N}$ is an upper-bound of A, contradicting that $\alpha = \sup A$.

Thus we see that $\alpha \in \mathbb{R}$, and α^2 is neither > 2 nor < 2; thus it must be the case that $\alpha^2 = 2$. □