Analysis I

Life from Math

Howard Xiao

Contents

1. The Alphabet

1.1 Sets

1.1.1 The Third Mathematical Crisis

Before communicating freely in any language, we first need to understand the words of that language. Even before this, we need to know the building blocks of words, the alphabet. For the mathematical language, we will first understand what the concept of *alphabet* means. For different purposes in math, we define different alphabets, hence different mathematical structures arise from these alphabets. The basic concept of alphabet is called **sets**.

> **Definition 1.1.1 — Sets.** A **set** is a collection of objects. We call an object x *"an element of a set S"* if x is contained in S, denoted $x \in S$.

■ Example 1.1 — Examples of Sets. ■

- $S = \{1,2,3\}$ is a **finite** set since S only has finite number (3) items.
- $S = \{\ldots,-2,-1,0,1,2,\ldots\}$ is an **infinite** set (often denoted as $\mathbb{Z}$, the integers).
- $S = \{$MAT157, MAT240, CSC148, CSC165$\}$ is a set of courses a typical student wishing to study math and CS would take at U of T.
- $S = \{a,b,c,d,\ldots,z\}$ is a set of letters used in English, which is exactly the alphabet of English.

> **R** Here, objects do not necessarily refer to mathematical objects, in fact it can be *anything*. A *set* is also an object itself.

> **Definition 1.1.2** The number of elements in a set S is called its **cardinality**, denoted $\#S$ or $|S|$.

■ Example 1.2 — Cardinality. ■

The examples in example 1.1 has cardinality $3,\infty,4,26$ respectively. We call a set S **finite** if $\$S$ is finite, and otherwise we call S to be **infinite**.

The above definition of sets gives rise to the following question: is the collection of all sets a set or not a set? Both answers will lead to a contradiction. This is known as the *Russel's paradox* and *the third mathematical crisis*. To address this problem, various set theories were proposed, the most famous one being what is known as the *"axiom of choice"*. However, we will now go into further details about these philosophical analysis on mathematics. We will focus more on the understanding of the math language.

1.1.2 Subsets

In fact, the mathematical language is much more general than any language we speak, for example English. One of the main reason is that we can extend or narrow our alphabet used while still having meaningful conversations, which is not possible for English (imagine using only a,b,c for daily lives). Narrowing down a set gives rise to the definition of subsets and intersection of sets. Extending a set gives rise to the definition of union of sets.

Definition 1.1.3 The sub-collection of objects (let us denote it C) in a set S is called a **subset** of S, denoted $C \subset S$.

Example 1.3 — Examples of subsets.

- $C = \{1\}$, $D = \{2,3\}$ are both subsets of $S = \{1,2,3\}$.
- We denote the empty collection of objects $\emptyset$. This is a subset of any set S.
- A proper subset C of a given set S is a subset $C \subset S$ with extra constraint that C and S are not the same collection. This can be emphasized using notation $C \subsetneq S$.

Definition 1.1.4 Given two sets A and B, we call the intersection of A and B another set $C : \{x : x \in A \text{ and } x \in B\}$, denoted $C = A \cap B$.

Example 1.4 — Set intersection.

- Let $S = \{1,2,3\}$, $T = \{1,3,5\}$, then $S \cap T = \{1,3\}$.
- Let $S = \{1,2,3\}$, $T = \{4\}$, then $S \cap T = \emptyset$. If two sets have empty intersection, we call them **disjoint** sets.
- From the definition of subsets, we can see that if $A \subset B$, $A \cap B = A$.

On the contrary of narrowing down to common elements of two sets which is the intersection described above, we have another operation on sets called the union of sets, which extends two sets to a set containing all appearing elements in them.

Definition 1.1.5 Given sets A,B, we define the union of A and B to be the set $C = \{x : x \in A \text{ or } x \in B\}$. We denote $C = A \cup B$.

Example 1.5 — Set union.

- Let $S = \{1,2,3\}$, $T = \{1,3,5\}$, then $S \cup T = \{1,2,3,5\}$.
- From the definition of subsets, we can see that if $A \subset B$, $A \cup B = B$.

There is sometimes another operation on sets that we care about, which involves dealing with what elements are not in a particular subset.

Definition 1.1.6 Given set A and set $S \subset A$, the complement of S in A, denoted S^c or $A \setminus S$ is the set $T = \{x : x \in A \text{ and } x \notin S\}$.

With the concept of subsets, intersection, union and complements defined, we can prove interesting properties on sets. In fact, there is an area of mathematics called *set theory*, and an introductory course to this is MAT409.
Different mathematical structures you will encounter often uses different alphabets, discussing relationships between these alphabets also allows us to discuss relationships between these mathematical structures.

1.1.3 Operations and Closedness

Beauty of mathematics begins to appear when we discuss how to combine letters in the alphabet to produce "words". The ways of combining letters to get words is known as **operations**.
An important difference between the mathematical language and natural language is that in some mathematical structures, combining letters in the alphabet using a defined operation can only give rise to other letters in the alphabet and *nothing outside* of this alphabet. In this case, this mathematical structure is said to be "closed" under the defined operation. We now will define *closedness* formally. This allows us to understand the mathematical language we would like to use in analysis later on. We will begin by defining the concept of operation.

Definition 1.1.7 An **operation** on a set S is an assignment of a unique element $s \in S$ to a fixed number of elements in S.

Example 1.6 — Operations. ■

- By definition, the addition "+" and multiplication "$\cdot$" defined on $\mathbb{R}$ is an operation that assigns a unique number $x \in \mathbb{R}$ given 2 elements $a, b \in \mathbb{R}$, denoted $x = a + b$ or $x = a \cdot b$, respectively.
- It is also an operation to assign any unique element $x \in S$ given the whole set of elements in S.

Definition 1.1.8 Given an operation $\cdot$ defined on a set S, a subset $A \subset S$ is said to be **closed** under this operation if for all $x, y \in A$, $x \cdot y \in A$.

R By definition of closedness and operation, we know that given an operation $\cdot$ defined on S, S is automatically closed under this operation.

Example 1.7 — Operation and Closedness. ■

- Using the normal definition of addition as the operation, integers $\mathbb{Z}$, rationals $\mathbb{Q}$ and real numbers $\mathbb{R}$ are all closed under this operation.
- With the normal definition of multiplication as the operation, integers $\mathbb{Z}$, rationals $\mathbb{Q}$ and real numbers $\mathbb{R}$ are also closed under this operation.
- With the normal definition of division, rationals $\mathbb{Q}$ and real numbers $\mathbb{R}$ are still closed under this operation. However, $\mathbb{Z}$ is not closed under division since $\frac{1}{2} \notin \mathbb{Z}$.

R In the above example, "normal" just mean that addition and multiplication are defined as what we always know, $1 + 2 = 3, 3 \cdot 2 = 6$ and so on.

1.2 Mathematical Structures

Most mathematical structures that we are ever interested in are sets closed under one or more kind of operations. Additional properties give rise to various mathematical structures we care about. The concept of mathematical structures lies in the area of *abstract algebra*, and the details will be covered in MAT347Y1.
As of now, it is important to understand the very basics of these structures as our primary interest, the real numbers $\mathbb{R}$, is assumed to be a member of an important type of mathematical structure.

1.2.1 Groups

Definition 1.2.1 A **magma** (binar, groupoid) $(M,\star)$ is a set M that is closed under operation $\star$ with no additional property assumptions.

■ Example 1.8 — Magma. ■

Since $\mathbb{R}$ is closed under addition and multiplication, we can define the operation $x\star y$ to be any combination of adding and multiplying x or y together and conclude that $\mathbb{R}$, together with the operation $\star$ is a magma, for example $x\star y = x^6y + x^3y^3$, then $(\mathbb{R},\star)$ in this case is a magma.

Definition 1.2.2 A **semigroup** $(M,\star)$ is a magma M that satisfies associative law: $a\star(b\star c) = (a\star b)\star c$, for all $a,b,c \in M$.

■ Example 1.9 — Semigroup. ■

The set $\{0,1\}$ can be equipped with operation "AND", "OR" and result in two semigroups. We define a AND $b = 1$ only when $a = b = 1$, and a AND $b = 0$ otherwise. We define a OR $b = 1$ when either a or b is 1, and a OR $b = 0$ otherwise, for all $a,b \in \{0,1\}$. An exercise is to try to verify that this is associative.

Definition 1.2.3 A **monoid** $(M,\star)$ is a semigroup M with an identity $e \in M$ such that for all $a \in M$, $a\star e = e\star a = a$.

■ Example 1.10 — Monoid. ■

The first example is a "flip-flop" monoid, which is the set $\{a,b,c\}$ where a represents operation "SET", b represents operation "RESET" and c represents operation "DO NOTHING". We know that $a\star b = c$, $a\star c = a$, $b\star c = b$, thus the identity element is c. An exercise is to try to verify that the "flip-flop" monoid is associative.

With grounds layed, we can talk about one of the most important algebraic structures - groups.

Definition 1.2.4 A **group** $(M,\star)$ is a monoid M with an additional operation *inv* on M, where we denote $inv(g) = g^{-1} \in M$ for all $g \in M$. This operation satisfies that $g\star g^{-1} = g^{-1}\star g = e$. If for all $a,b \in M$, we also have $a\star b = b\star a$, we call the group to be **abelian**.

■ Example 1.11 — groups. ■

- We call the Hamiltonian Quarternions $\mathbb{H} = \{\pm 1, \pm i, \pm j, \pm k\}$ to be the set of 8 elements with operation $\cdot$, where $1^2 = 1 = (-1)^2$, $i^2 = j^2 = k^2 = -1$, $ij = k = -ji$, $jk = i = -kj$ and

$ik = j = -ki$. You can check that it is indeed a group (it satisfies associative laws, has identity and inversion operation).

- We call Klein four group $\mathbb{K} = \{1, a, b, c\}$ to be the set of 4 elements with operation $\cdot$, where 1 is the identity and $a \cdot b = c = b \cdot a$, $b \cdot c = a = c \cdot b$, $a \cdot c = b = c \cdot a$. This is an example of an *abelian group*.

R It is easy to see once the inverse of an element $m \in M$ exists, then it is unique. Suppose m has two inverses m_1, m_2, then we know that $m \star m_1 = m \star m_2 = e$, hence $(m_1 \star m) \star m_1 = (m_1 \star m) \star m_2$, and we conclude $m_1 = m_2$. Therefore, the inverse is **unique** if it exists.

1.2.2 Fields and $\mathbb{R}$

Definition 1.2.5 A **ring** is an abelian group $(M, \star)$ equipped with another operation $\cdot$ different from $\star$, such that M is closed under $\cdot$ and associativity holds for $\cdot$. A ring is **commutative** if M is commutative under $\cdot$, i.e., for all $a, b \in M$, $a \cdot b = b \cdot a$.

R Note that the identity for $\cdot$ may be different from the identity for $\star$. Notation wise, we often define the inverse of $a \in M$ under $\star$ to be $-a$ and inverse of a under $\cdot$ to be a^{-1}.

■ **Example 1.12 — Rings.** ■

The integers $\mathbb{Z}$ is a classic example of a ring. It is easy to check that $(\mathbb{Z}, +)$ is an abelian group, and $(\mathbb{Z}, \cdot)$ is associative.

In fact the reason why $\mathbb{Z}$ is a ring is purely by definition (assumption). This assumption comes from a broader assumption about $\mathbb{R}$. Before we go into this assumption about the "words" which is also our alphabet for the rest of the notes, we need to understand the concept of *fields*.

Definition 1.2.6 A **field** is a commutative ring $(M, \star, \cdot)$ with an inversion operation defined on $\mathbb{F} \setminus \{e\}$ for $\cdot$, where e is the identity for operation $\star$.

It is important to note that up to this point, we know that field $\mathbb{F}$ with operations $\star, \cdot$ and identities e, e' respectively has the following properties. For all $a, b, c \in \mathbb{F}$, $a \neq e$:

- **Associative Law for $\star$:** $a \star (b \star c) = (a \star b) \star c$.
- **Existence of identity for $\star$:** $a \star e = e \star a = a$.
- **Existence of inverse for $\star$:** $a \star (-a) = e = (-a) \star a$.
- **Commutativity for $\star$:** $a \star b = b \star a$.
- **Associative Law for $\cdot$:** $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
- **Existence of identity for $\cdot$:** $a \cdot e' = e' \cdot a = a$.
- **Existence of inverse for $\cdot$:** $a \cdot a^{-1} = e' = a^{-1} \cdot a$.
- **Commutativity for $\cdot$:** $a \cdot b = b \cdot a$.

R We sometimes denote e by 0, and e' by 1.

The above 8 properties are also called **field axioms**. Now, we are finally able to understand the "words" of our mathematical language which is the real numbers $\mathbb{R}$. We begin by making assumptions on our words $\mathbb{R}$ as follows:

Theorem 1.2.1 $\mathbb{R}$ with operations addition $(+)$ and multiplication $(\cdot)$ is a field.

This is the end of this chapter, but the starting point of everything that follows. We assumed that our alphabet is also our words $\mathbb{R}$ and we equipped lots of interesting field axioms on $\mathbb{R}$. Just as the English words, we have nouns, verbs and different type of nouns, etc.. We will talk about different types of words in $\mathbb{R}$, including $\mathbb{Q}, \mathbb{Z}$, and $\mathbb{N}$ in the next chapter. We will also talk about logics and proofs that somewhat defines the "grammar" of the math language.

2. The Words $\mathbb{R}$

In this chapter, we will be talking and learning about our words for (Real) analysis, which is the real numbers $\mathbb{R}$. The words for complex analysis is the complex numbers $\mathbb{C}$, which we will not talk about it here. Introduction to complex analysis course MAT354 takes care of this topic. There are many types of words, and most of the time to learn a language is spent on learning nouns, and here in real analysis, majority of the time will also be about *nouns*. As we will see, the nouns for $\mathbb{R}$ are the rationals, $\mathbb{Q}$. There are also many types of nouns such as $\mathbb{Z}$ and $\mathbb{N}$, which we will introduce in this chapter. In mathematical language, we associate emotions with words. We will also talk about emotions in this chapter, so that we can begin our dialogue(conversation) in the next one.
In this chapter, we will start talking about building blocks of analysis, in fact of all mathematics, which are logics and proofs, and we will end off the chapter with a specific proof technique of mathematical induction which gives properties of our specific nouns $\mathbb{N}$ and an official definition of our words $\mathbb{R}$.

2.1 Logics and Proofs

This section might be a little boring with symbols and abstract logic, however it serves as the foundation for all analysis. In fact, both math language and our natural language can be thought as a formal system, where the "rules" of this system is the *grammar of the language*. A *proof* in math language intuitively is a paragraph written using math words, but it must follow the correct grammar of mathematical language, which is *logic.*
In this section, we aim to be able to understand the logic behind the proofs, and throughout the notes, I am describing the natural language analogy to build intuition behind almost all proofs. With good intuition and clear logic flows, successful proofs come out of good understanding of mathematics and the beauty of math arises from these proofs.

2.1.1 Implications and Negations

In this section, we will first explain the mathematical equivalent of the "if...then..." sentence, which is implication and the symbol $\Rightarrow$, $\Leftarrow$ and $\iff$. Before this, we need to understand that a **boolean variable** is a variable that can only have value "True" or "False", and a **logic statement** is a statement consisting of math symbols, numbers and boolean variables that can only be evaluated "True" or "False". We will often write "statement" instead of "logic statement".

Definition 2.1.1 We denote $A \Rightarrow B$ to represent "if A then B". Here, A, B are boolean variables or logic statements. The statement $A \Rightarrow B$ is true *precisely when* A is False or B is True. Similarly, the statement $A \Leftarrow B$ is equivalent to $B \Rightarrow A$. The statement $A \iff B$ is True only when both $A \Rightarrow B$ and $B \Rightarrow A$ are True. We sometimes call A and B to be **equivalent** if $A \iff B$ is True.

■ **Example 2.1 — Implications.** ■

- The statement "If U of T is easy, then $\sqrt{2}$ is irrational." is True since in this case $A =$"U of T is easy", which is *False*. However, this is *not a valid proof* of $\sqrt{2}$ being irrational, as we will see later on.
- The statement "If U of T is hard, then $\sqrt{2}$ is rational" is False since in this case although $A =$"U of T is hard" is True, $\sqrt{2}$ is not rational.

(R) We can have a *chain* of *same direction implications*, for example $A_1 \Rightarrow A_2, A_2 \Rightarrow A_3, \dots$. In this case, we can simply notation and write $A_1 \Rightarrow A_2 \Rightarrow A_3 \Rightarrow \dots$.

Next, we will talk about the mathematical equivalence of the word "not", which is negations.

Definition 2.1.2 A negation is a logical statement of the form "not A", denoted $\neg A$, is true *only when* A is False.

Exercise: Verify the following statements are True:

$$(A \Rightarrow B) \iff (\neg B \Rightarrow \neg A)$$
$$\neg(\neg A) \iff A$$

■

(R) A good method of verifying two logic statements to be equivalent is to compare the cases when they are each True/False, and verify that they are both True or they are both False.

2.1.2 Conjunctions and Disjunctions

In this section, we will describe two important logical connections – conjunctions and disjunctions. This is the mathematical equivalence of "and" and "or", respectively.

Definition 2.1.3 We use A AND B to denote the logical conjunction, which is True *only when* A and B are both True. We use A OR B to denote the logical disjunction, which is True if one of A or B is True, or if both of them are True.

Exercise: Verify that the following logic statements are True (**De Morgan's Laws**):

$$\neg(A \text{ OR } B) \iff \neg A \text{ AND } \neg B.$$

$$\neg(A \text{ AND } B) \iff \neg A \text{ OR } \neg B.$$

■

Exercise: Verify that the following logic statement is True:

$$(A \Rightarrow B) \iff \neg A \text{ OR } B$$

■

2.1.3 Quantifiers

In this section, we are going to talk about two important quantifiers, **existential quantifier** (which corresponds to "there exists") and **universal quantifier** (which corresponds to "for all"). Using quantifiers and previously described logic symbols, we can form comprehensive logical statements used in mathematical proofs.

Definition 2.1.4 We use symbol $\exists$ to denotes that "there exists" and the symbol $\forall$ to denote "for all". Both quantifiers need to combine with sets (called the **domains** of the quantifiers) to make sense. A statement of $\exists$ is True if there is an element in the domain that makes the statement True. A statement of $\forall$ is True if every element in the domain makes the statement True.

■ Example 2.2 — Quantifiers. ■

- Consider the statement $\forall x \in \mathbb{R}.\exists a \in \mathbb{Z}.a < x$. This statement is True only when for x being any real number, we can find a specific integer a smaller than this x. The choice of a *can depend* on the value of x.
- Consider the statement $\exists x \in \mathbb{Z}.\forall a \in \mathbb{R}^+.|x| < |a|$. This statement is True only when we can find an integer x such that any positive real number a satisfies $|x| < |a|$. The value of x does not depend on any value of a.

Exercise: Verify that the following statements are True (A is a domain and B is a statement):

$$\neg(\forall x \in A.B.) \iff \exists x \in A.\neg B.$$

$$\neg(\exists x \in A.B.) \iff \forall x \in A.\neg B.$$

■

After the above exercises, you can get a complete understanding of how to negate a specific logic statement. There are several steps to this process, including:

1. Change any implications into conjunctions and disjunctions by definition and the second exercise in 2.1.2.
2. Distribute $\neg$ inward, changing conjunctions and disjunctions accordingly by De Morgan's Laws, and swap $\exists$ and $\forall$ by the above exercises.
3. When there is double $\neg$, remove them, according to previous exercises in 2.1.1.

■ **Example 2.3** We would like to change the statement

$$A = \text{"}\forall \varepsilon > 0.\forall x < 1.\exists N \in \mathbb{N}.\forall n \in \mathbb{N}.n > N \Rightarrow x^n < \varepsilon.\text{"}$$

to $\neg A$.

1. Firstly, we rewrite A to be:

$$A = \text{"}\forall \varepsilon > 0.\forall x < 1.\exists N \in \mathbb{N}.\forall n \in \mathbb{N}.[\neg(n > N) \text{ OR } (x^n < \varepsilon)].\text{"}$$

2. Next, we distribute $\neg$ inward, to get:

$$\neg A = \text{"}\exists \varepsilon > 0.\exists x < 1.\forall N \in \mathbb{N}.\exists n \in \mathbb{N}.\neg(\neg(n > N)) \text{ AND } \neg(x^n < \varepsilon).\text{"}$$

3. Finally, we negate the $<$ and $>$, as well as removing double $\neg$ and get:

$$\neg A = \text{"}\exists \varepsilon > 0.\exists x < 1.\forall N \in \mathbb{N}.\exists n \in \mathbb{N}.n > N \text{ AND } x^n \geq \varepsilon.\text{"}$$

■

2.1.4 Proofs, finally

A proof is describing a *logical flow* to show that a statement is True via a *fixed set* of True statements. This is the most important technique we use to establish new math results (to add to the set of True statements) and to practise mathematical language. There are various types of proof techniques, and here in this chapter, we will only describe the basic ones. We will start by the most standard proof technique – *proof by logical deduction*.

Definition 2.1.5 A **logical deduction** is the method of showing a statement Y is True by showing that X is *True* and the statement $X \Rightarrow Y$ is *True*.

Let us denote the statement we would like to prove G. Proof by logical deduction is a proof technique that starts with known-True statement A in the fixed set of True statements (this can be axioms, definitions or proved theorems, propositions and statements) and use a series of logical deductions, to finally show that G is True.

■ **Example 2.4** Prove that if $\mathbb{F}$ is a field, then we have **cancellation law:** if $a \cdot b = c \cdot b$ where $a, b, c \in \mathbb{F}$, $b \neq 0 \in \mathbb{F}$, then $a = c$.
Proof. We will start with the known-True statement in the question, which is $a \cdot b = c \cdot b$.
Then, by properties of equality, we know that if $x = y$, then $x \cdot z = y \cdot z$ for any z. This allows us to make the logical deduction to show that:

$$a \cdot b \cdot b^{-1} = c \cdot b \cdot b^{-1}.$$

(Of course this also uses the fact that $b^{-1} \in \mathbb{F}$ exists since $\mathbb{F}$ is a field, which is a more subtle logical deduction.)
Then, we can use the associative law of field $\mathbb{F}$ and properties of equality to show that:

$$a \cdot b \cdot b^{-1} = a \cdot (b \cdot b^{-1}) = a \cdot e = a = c \cdot b \cdot b^{-1} = c \cdot (b \cdot b^{-1}) = c \cdot e = c.$$

(Above is another logical deduction.)
Finally, we have shown that $a = c$, which is the statement we would like to prove (G), hence above completes this proof. □

■

Next, we will talk about another very important technique of proof - *proof by contradiction*.

Definition 2.1.6 A **proof by contradiction** is an argument of showing that a statement G is True through assuming $\neg G$ is True, and prove $\neg Y$ is True via *logical deduction* for some known True statement Y in the fixed set of True statements. The fact that $\neg Y$ is proved and Y is True is called a **contradiction**.

R Intuitively, this is just assuming the statement we are proving to be False, then find a contradiction with a known True statement. Then, this implies that the statement we are proving is not False, hence True.

■ Example 2.5 ■

Prove that $\sqrt{2}$ is irrational (i.e. $\sqrt{2}$ cannot be written in the form of $\frac{p}{q}$ for any integers $p,q \in \mathbb{Z}$.)

Proof. Suppose that $\sqrt{2}$ is not irrational, i.e. $\sqrt{2} = \frac{p}{q}$ for some $p,q \in \mathbb{Z}$. (First step in proof by contradiction). We can assume that they are reduced into lowest terms, i.e. p,q have no common factors (since if they have common factors, we can further reduce them until they don't).
Then, by properties of equality, we know that $\sqrt{2}\cdot q = \frac{p}{q}\cdot q = p$. Next, by properties of equality again, we can square both sides of the equality and get:

$$2q^2 = p^2.$$

Next, we would like to form contradiction with the help of even/odd properties statements that are known to be True. Firstly, since $2q^2$ is even, $2q^2 = p^2$, p^2 is even. Since the square of an odd integer is odd, the only way for p^2 to be even is that p is even. By definition of "even", we can write $p = 2k$ for some $k \in \mathbb{Z}$, then we know that $2q^2 = p^2 = (2k)^2 = 4k^2$, hence by properties of equality, we can divide 2 on both sides and get:

$$q^2 = 2k^2.$$

With similar arguments as above, we know that q^2 is even, thus q is even. However, p,q are both even suggests that they have a common factor 2.
However, at the start of the proof, we assumed that p,q have no common factors, this gives us a contradiction. Therefore, we have proved by contradiction that $\sqrt{2}$ is irrational. □

The third proof technique that we will talk about in this section is proof by cases, which is very useful if adding in additional property assumptions (i.e., cases), the statement of interest is easier to prove.

Definition 2.1.7 A **proof by cases** is a proof for a statement of interest involving elements from a specific domain D, given by first choose finitely many subsets $D_1, D_2, \dots, D_n \subset D$ such that $D_1 \cup D_2 \cdots D_n = D$. Then, for each specific subset, prove the statement using proof by logical deduction or contradiction. Each subset $D_1, D_2, \dots$ are called **cases**.

This proof technique becomes very intuitive once we understand an example.

■ Example 2.6 ■

Prove that for all $a,b \in \mathbb{R}$, the absolute value $|ab| = |a||b|$. The definition of the absolute value is that for all $x \in \mathbb{R}$, $|x| = x$ if $x \geq 0$, and $|x| = -x$ otherwise.

Proof. We will consider four cases and do a proof by cases in the following way:

Case 1. $a,b \geq 0$.

Then, we know that $ab \geq 0$, thus $|ab| = ab = |a||b|$ by definition of absolute value. Hence, the statement is proved for Case 1.

Case 2. $a \geq 0, b < 0$.

Then, we know that $ab \leq 0$, thus $|ab| = -ab = a(-b) = |a||b|$ by definition of absolute value. Hence, the statement is proved for Case 2.

Case 3. $a < 0, b \geq 0$.

Similar to Case 2, we know that $ab \leq 0$, $|ab| = -ab = |a||b|$ by definition of absolute value. Hence, the statement is proved for Case 3.

Case 4. $a < 0, b < 0$.

In this case, we know that $ab > 0$, thus $|ab| = ab = (-a)(-b) = |a||b|$ by definition of absolute value. Hence, the statement is proved for Case 4.

Then, since the four cases above cover every possible case for $a,b \in \mathbb{R}$, we know that $|ab| = |a||b|$ is proved via a proof by cases. □

The proof by logical deduction and proof by contradiction are two classic proof techniques that are used frequently. Often times, when the domain of proof is too "general", we can break it down into cases and use logical deduction and contradiction to prove each case separately. This constitutes a proof by cases. There is another well-known and important proof technique - *proof by mathematical induction*. However, that depends on a bit more knowledge about the math language, especially the words that we are speaking.

2.2 Nouns

As I explained before, learning nouns are a core part of studying and understanding a language. In real analysis, we will treat $\mathbb{R}$ as the whole universe of words, and $\mathbb{Q}$, the rationals are the nouns here. Let us start by understanding various type of nouns and then understand what $\mathbb{Q}$ is.

2.2.1 The Proper Nouns

In natural language, proper nouns are nouns that represent specific *people, places and things.* In our daily language, they are often the items with lots of functionalities, memories or knowledge associated. In mathematics language, the proper nouns are very "nice" numbers in the universe of nouns $\mathbb{Q}$.

Definition 2.2.1 The **integers** $\mathbb{Z}$, or the *proper nouns* in the math language is a set of numbers containing the multiplicative identity $1 \in \mathbb{R}$, the additive identity $0 \in \mathbb{R}$, all additive multiples of 1 and their additive inverses. The notation is to denote $2 = 1+1$, $3 = 1+1+1$, $\ldots$ as we normally use. We use $-x$ to denote the additive inverses of x for all $x \in \mathbb{R}$. Thus, $\mathbb{Z} = \{\ldots,-3,-2,-1,0,1,2,3,\ldots\}$.

R $\mathbb{Z}$ has the same addition and multiplication defined for $\mathbb{R}$, and in section 1, we have already showed that $\mathbb{Z}$ is a *ring* but not a **field** since it does not have multiplicative inverses.

With the definition of the "proper nouns" $\mathbb{Z}$, we can define another very important type of nouns - the natural numbers $\mathbb{N}$.

Definition 2.2.2 The set of **natural numbers**, denoted $\mathbb{N}$ is the set of all additive multiples of $1 \in \mathbb{R}$, i.e. $\mathbb{N} = \{1, 2, \dots\}$.

Based on this definition, we can see that the set of natural numbers is a subset of the integers $\mathbb{Z}$ but is not even a group since we do not include 0 and additive inverses. However, with this definition, we are able to find very important properties of the natural numbers, and propose another proof technique later on. After our discussion of emotions and ordered fields, we can also form intuition on $\mathbb{N}$.

2.2.2 All Nouns from Proper Nouns

With the definition, or rather classification of proper nouns $\mathbb{Z}$ from all words, we can finally classify the universe of nouns, namely $\mathbb{Q}$.

Definition 2.2.3 A number $x \in \mathbb{R}$ is a *mathematical noun* (or we usually call it **rational number** if $x = \frac{p}{q}$ for some $p, q \in \mathbb{Z}$. We denote the set of all rational numbers $\mathbb{Q}$, and they are all nouns of the math language. q in $x = \frac{p}{q}$ is called the **denominator** of x, and p is called the **numerator** of x.

R We have briefly discussed that $\mathbb{Q}$ with the same addition and multiplication defined as in $\mathbb{R}$ is also a field. This can be shown since for every $x = \frac{p}{q} \in \mathbb{Q}$, $x^{-1} = \frac{q}{p}$. We say $\mathbb{Q}$ is a **subfield** of $\mathbb{R}$.

With the definition, we notice that there exists multiple ways to write a rational number x as the quotient of two integers. For example, $\frac{3}{4} = \frac{9}{12} = \frac{12}{16} = \dots$. In the proof that $\sqrt{2}$ is irrational before, we briefly talked about the concept of "*rational in lowest terms*". We will now make it formal.

Definition 2.2.4 A rational $x \in \mathbb{Q}$ is said to be written **in lowest terms** $\frac{p}{q}$ $(p, q \in \mathbb{Z})$ if $x = \frac{p}{q}$ and for any $p', q' \in \mathbb{Z}$ such that $x = \frac{p'}{q'}$, we know that $q' \geq q$.

R In the above definition, we are saying that $x = \frac{p}{q}$ is written in lowest terms if $\frac{p}{q}$ is the way to write x into quotients of integers for q to be the *smallest*.

2.3 Emotions

In fact, the most important intuition to have about the mathematical language used in analysis is the *"emotions"* of words. In the current real analysis domain, our universe of words is $\mathbb{R}$, our nouns are $\mathbb{Q}$ which are generated from proper nouns $\mathbb{Z}$. In English, we often have subtle emotions associated with each noun, "table", "professor", "mathematics", "U of T", "The Cows", or "Tim Hortons".

We put them into conversations where emotions move in. In fact, every math word has emotions associated with it and in this section, we will make sense of the idea of "emotion" in math words. Before we try to understand the strength of emotions and compare emotions of words, we first need to gain the sense of **order**. More generally, we will talk about the notion of **order** on *general fields*.

Definition 2.3.1 An **ordered field** is a field $\mathbb{F}$ with the notion of the set of positive elements P and the set of negative elements N, such that it satisfies three properties:
- $P \cap N = \emptyset$; $0 \notin P, N$; $P \cup N \cup \{0\} = \mathbb{F}$, where 0 is the additive identity of $\mathbb{F}$.
- For all $a, b \in P$, $a + b \in P$.
- For all $a, b \in P$, $ab \in P$.

With the notion of "order" defined as above, we can finally make *concrete* of the notion of "less than" and "greater than".

Definition 2.3.2 For all $a, b \in \mathbb{F}$ which is an ordered field, we say $a > b$ (a is greater than b) if $a - b \in P$, $a < b$ (a is less than b) if $a - b \in N$ and $a = b$ if $a - b = 0$.

R We often use the notation of $a \geq b$ to represent "$a > b$ OR $a = b$" is True, and $a \leq b$ likewise.

There are various properties of order and some of them are listed below. Proofs are left as exercises since almost all the properties directly follow from definition of order, or previously proved properties.

Proposition 2.3.1 For any $a, b, c, d \in \mathbb{F}$, which is an ordered field, we have:
- Exactly one of $a > b$, $a < b$, $a = b$ is True.
- $a > b \iff b < a$
- $a > b \iff a + c > b + c$
- $a > b, c > 0 \Rightarrow ac > bc$.
- $a > b, c < 0 \Rightarrow ac < bc$.
- $a > 0 \iff a^{-1} > 0$
- $a > 0 \iff -a < 0$
- $a \neq 0 \iff a^2 > 0$
- $a > b > 0 \iff b^{-1} > a^{-1} > 0$
- $a > b, c > d \Rightarrow a + c > b + d$
- $a > b > 0, c > d > 0 \Rightarrow ac > bd$.
- If $x \geq 0, y \geq 0$, $x \geq y \iff x^2 \geq y^2$.
- $1 > 0$. (*Hint:* Consider $1 = 1^2$)
- $\mathbb{F}$ cannot have finitely many elements. *Hint: consider 1, 1 + 1, 1 + 1 + 1, ...*

Theorem 2.3.2 $\mathbb{R}$, with the definition of **positive real numbers** $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$ and **negative real numbers** $\mathbb{R}^- = \{x \in \mathbb{R} : x < 0\}$ is an **ordered field**.

This theorem is very easy to verify by the definition of ordered fields, and I need to remark here that the of "order" for $\mathbb{R}$ is purely a **definition**, as we always know, such as $1.6 > 1$, $3 < 4.2$, ...

In the mathematics language, the presence of emotion is often made explicit, in fact, the ways to measure the emotions of mathematical words can be varied and are often defined explicitly (depending on the order). This gives rise to the definition of **norm**, which is the measurement of emotions for each word. Below is rather an informal definition of a *norm* on $\mathbb{R}$, a formal one will be given once one understands "conversation".

Definition 2.3.3 A **norm** on math words $\mathbb{R}$ is an assignment of emotion value in $\mathbb{R}$ for each word $x \in \mathbb{R}$, denoted $|x|$ such that it satisfies the following three properties:

- The emotion value of the word 0 is 0, i.e., $|0| = 0$. Also, the only word with emotion value of 0 is 0, i.e. $|x| = 0 \Rightarrow x = 0$ is True.
- The emotion value of a word produced by multiplication is the multiplication of the emotional values, i.e. $|ab| = |a||b|$ for all $a, b \in \mathbb{R}$.
- The emotion value of a word produced by addition is less than or equal to the addition of the emotional values, i.e. $|a+b| \leq |a| + |b|$ for all $a, b \in \mathbb{R}$.

R The second and third property of a norm can be understood in the following way: when multiplying two words, the emotions do not have a "cancel-out" effect but when adding two words together, the result word might have less emotion value than the emotional value of two words added together, which is a *cancel-out effect*.

Example 2.7 — Norms.

- The first, and the simplest norm is the following one, where:

$$|x| = \begin{cases} 0 & \text{if } x = 0, \\ 1 & \text{otherwise.} \end{cases}$$

This is indeed a norm, which is not hard to verify.

With the definition of order, The most important norm for $\mathbb{R}$ is the **Euclidean norm**, which can be defined in the following way with the use of $>$ and $<$. For all $x \in \mathbb{R}$,

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

It is simple to verify that the Euclidean norm satisfies the first property of norm, and in fact the proof of the second and the third property is not as straightforward. It is rather straight forward to see that the definition of Euclidean norm is in fact the definition of **absolute value**. From example 2.6, we can then verify the second property. We will prove the third property in the following Theorem.

Theorem 2.3.3 — Euclidean Norm. The Euclidean Norm satisfies three properties of norm.

Proof. The first property is very direct from the definition of Euclidean Norm, and will be omitted here as an exercise.
The proof of the second property is provided in example 2.6 via proof by cases.
While the third property can also be proved using proof by cases, we will rather prove it in a simpler way. By the property of Euclidean norm, we know that $|x|^2 = x^2$ for all $x \in \mathbb{R}$. Thus, for any $a, b \in \mathbb{R}$, $|a+b|^2 = (a+b)^2 = a^2 + 2ab + b^2$. Also, we have:

$$(|a| + |b|)^2 = |a|^2 + 2|a||b| + |b|^2 = a^2 + 2|a||b| + b^2.$$

Next, it is easy to verify based on the definition of Euclidean norm that $|a||b| \geq ab$ (can use proof by cases, left as an exercise). We then know that, by properties of inequalities:

$$(|a|+|b|)^2 = a^2 + 2|a||b| + b^2 \geq a^2 + 2ab + b^2 = (|a+b|)^2.$$

Next, since $|a|+|b| \geq 0$, $|a+b| \geq 0$, we can conclude that $|a|+|b| \geq |a+b|$, which is the statement we would like to prove. □

R In fact, the notion of "Euclidean norm" (i.e. "absolute value") can be defined for all ordered fields. This is the emotional measurement that we use for our words $\mathbb{R}$, representing the "emotions" of $\mathbb{R}$.

With the definition of Euclidean norm, we have actually another more intuitive way to define types of emotions, such as *positive* and *negative* emotions for $\mathbb{R}$.

Definition 2.3.4 We call **positive real numbers**, denoted $\mathbb{R}^+$ to be the set of nonzero real numbers (words) with emotion value equal to their own value under Euclidean norm, i.e. $\{x \in \mathbb{R} \setminus \{0\} : |x| = x\}$, and **negative real numbers** denoted $\mathbb{R}^-$ to be the set of nonzero real numbers (words) with own value being the additive inverse of emotion value under Euclidean norm, i.e. $\{x \in \mathbb{R} \setminus \{0\} : |x| = -x\}$.

R We can easily see that based on the definition of Euclidean norm, $\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\}$ and $\mathbb{R}^- = \{x \in \mathbb{R} : x < 0\}$. With this definition, we can see that the set $\mathbb{N}$ is the set of *positive integers*, which intuitively can be seen as the "*proper nouns with good (positive) emotions*".

There are some other properties of Euclidean norm which can be easily verified with the definition of ordered fields, or norms. We will end this subsection by listing some of them below (proofs omitted as exercises):

Proposition 2.3.4 In $\mathbb{R}$ with Euclidean norm $|\cdot|$, we have:
m$a \neq 0 \iff |a| > 0$. $a^2 = |a|^2$. $|a| = |-a|$. $|a^{-1}| = |a|^{-1}$.

2.4 Well-ordering and Mathematical Induction

In previous subsections, we have described various categories of nouns, and with the help of emotions and the notion of "norm" as the measurement of emotions, we gain intuitive understanding of some more categories of nouns. In the end of this chapter, we will focus on $\mathbb{N}$, since there are very important properties of $\mathbb{N}$ and also since it gives rise to another important proof technique called *mathematical induction*.

The other reason for studying $\mathbb{N}$ is that natural numbers appear naturally in our daily lives as "counting", and appears naturally in mathematical language as "sequences" and "series", which we will study later on.

Let us first begin by talking about some elementary properties of the natural numbers:

Theorem 2.4.1 Properties of $\mathbb{N}$:
- For all $m,n \in \mathbb{N}$, $m < n \iff n-m \in \mathbb{N}$.
- For all $n \in \mathbb{N}$, there are no natural numbers m such that $n < m < n+1$.

Proof. We will prove these properties one by one.
- Consider any $m,n \in \mathbb{N}$. We know by definition that $m,n \in \mathbb{Z}$. Hence, since $\mathbb{Z}$ is a group under addition, $n-m \in \mathbb{Z}$. Since $m < n$, $n-m > 0$, thus by definition, $n-m \in \mathbb{N}$.
- Suppose by contradiction that there exists $m \in \mathbb{N}$ such that for some $n \in \mathbb{N}$, $n < m < n+1$. Then, we know that $0 < m < 1$ by properties of inequalities, however there are no integers between 0 and 1 by definition, hence no natural numbers, which gives us a contradiction. □

We will prove other stronger properties of $\mathbb{N}$ by first develop a very important proof technique specifically used for $\mathbb{N}$ called **proof by mathematical induction**. Before we go into details about this proof technique, we first need to understand what *"induction"* means. We will begin by defining inductive subsets of $\mathbb{R}$.

Definition 2.4.1 An **inductive subset** $S \subset \mathbb{R}$ is a set satisfying the property that:
- $1 \in S$.
- $x \in S \Rightarrow x+1 \in S$ is True for all $x \in \mathbb{R}$.

Theorem 2.4.2 $\mathbb{N}$ is the smallet inductive subset of $\mathbb{R}$, i.e. for all inductive subsets $S \subset \mathbb{R}$, $\mathbb{N} \subset S$.

Proof. Let S be any inductive subset of $\mathbb{R}$. Then, we know that $1 \in S$, and $x \in S \Rightarrow x+1 \in S$ is True for all $x \in \mathbb{R}$. Since $1 \in S$, we know that $2 \in S$, then $3 \in S$, and thus all additive multiples of 1 is in S. Since all additive multiples of 1 are positive integers, by definition of $\mathbb{N}$, we know that $\mathbb{N} \subset S$. □

From above, we not only know that $\mathbb{N}$ is inductive, but also $\mathbb{N}$ is the smallest subset that is inductive. With this property of $\mathbb{N}$, we can actually mimick the definition of inductive subsets to develop a way to *prove* properties of $\mathbb{N}$.

Definition 2.4.2 A **proof by mathematical induction** is a proof technique specifically used to prove properties of $\mathbb{N}$. The proof first defines a statement $P(n)$ that we would like to prove for all $n \in \mathbb{N}$, then show $P(1)$ is true, and the statement "$P(n) \Rightarrow P(n+1)$" is True for all $n \in \mathbb{N}$.

The above definition being a *valid proof technique* can be shown by the inductive property of $\mathbb{N}$. Let $T = \{n \in \mathbb{N} : P(n) \text{ is True.}\}$. Then by definition of T, we know that $T \subset \mathbb{N}$.
Then, if we have successfully done a proof by mathematical induction, we know that $1 \in T$, $x \in T \Rightarrow x+1 \in T$ is True for all $x \in \mathbb{N}$. Since for all $x \in \mathbb{R} \setminus \mathbb{N}, x \notin T$, thus the statement $x \in T \Rightarrow x+1 \in T$ is True. Hence, by definition, T is inductive, and thus $\mathbb{N} \subset T$, hence we know that $T = \mathbb{N}$, i.e. $P(n)$ is True for all $n \in \mathbb{N}$.

R The proof by induction structure sometimes can be used to prove properties for some other sets as well, for example to prove something for all even numbers, we can start by proving $P(0)$, and then show $P(n) \Rightarrow P(n+2)$ is True for all n even. To prove something for all odd numbers, we can start by proving $P(1)$ and then show $P(n) \Rightarrow P(n+2)$ is True for all n odd. We can also prove statements for $\mathbb{Z}$ by proving $P(0)$, then show $P(n) \Rightarrow P(n+1)$ is True for all $n \in \mathbb{Z}, n \geq 0$ and show $P(n) \Rightarrow P(-n)$ is True for all $n \in \mathbb{Z}, n \geq 0$.

■ Example 2.8 — Mathematical Induction. ■

Prove that for all $n \in \mathbb{N}$, $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

Proof. It is very straight-forward for us to notice that the statement we are trying to prove here is about a property for $\mathbb{N}$. In this case, $P(n)$ corresponds to the statement "$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$".

Our first step is to show $P(1)$. For $n = 1$, we know that $\sum_{i=1}^{1} i = 1 = \frac{1(1+1)}{2}$, thus $P(1)$ is True by simple calculation.

Our next step is to show that $P(n) \Rightarrow P(n+1)$ is True for all $n \in \mathbb{N}$, which requires us to assume $P(n)$ being True and prove $P(n+1)$. Let n be any natural number, and assume $P(n)$ is True. Then, we know that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

We also know that $\sum_{i=1}^{n+1} i = \sum_{i=1}^{n} i + n + 1 = \frac{n(n+1)}{2} + n + 1 = \frac{(n+2)(n+1)}{2} = \frac{(n+1)(n+1+1)}{2}$.

Thus, we have verified that "$P(n) Rightarrow P(n+1)$" is True. Therefore, by proof of mathematical induction, $P(n)$ is True for all $n \in \mathbb{N}$. □

> **R** In the proof above, and in any proof by mathematical induction, we call the assumption that "$P(n)$ is correct" **the inductive hypothesis**.

There are a lot of amazing properties about natural numbers, and later on about polynomial rings (MAT347Y1) that can be proved with mathematical induction. I will try to provide an exercise sheet regarding properties of a very interesting polynomial called binomial, and their coefficients.

With the tool of mathematical induction, we can go on to prove an extremely useful and important result for $\mathbb{N}$ called the **well-ordering principle**.

Theorem 2.4.3 The **well ordering principle** states that for every nonempty subset $S \subset \mathbb{N}$, S contains a minimum element, i.e. there exists $s \in S$ such that for all $x \in S$, $s \leq x$.

Actually, the well-ordering principle is very intuitive, since the natural number represents proper nouns with good/positive emotions, for any nonempty subset of proper nouns with good/positive emotions, we might not have a noun with the *best* positive emotion, but we definitely have one with the *worst* positive emotion, that is the "minimum" in this subset.

Proof. This elegant proof combines both proof by contradiction and proof by mathematical induction. We first suppose we can find an nonempty $S \subset \mathbb{N}$ such that S does not have a smallest element.

The beautiful part is where induction comes in. We ask the question "What element is not in S?" Let $P(n)$ be the statement that for all $k \in \mathbb{N}, k \leq n$, $k \notin S$. Then, we know that $P(1)$ is True since if $1 \in S$, and $S \subset \mathbb{N}$, S has a smallest element since 1 is indeed the smallest element of $\mathbb{N}$ by definition.

Next, let us suppose $P(n)$ is true for any $n \in \mathbb{N}$. Then we know that for all $k \in \mathbb{N}, k \leq n$, $k \notin S$. Now, can $n+1 \in S$ be True? No, because if so, since $n+1$ is the next natural number after n, and we know that all natural numbers from 1 up to n are not in S, $n+1$ would then be the smallest natural number

in S, which is impossible. Therefore, $n+1 \notin S$, hence for all $k \in \mathbb{N}, k \leq n+1$, $k \notin S$, $P(n+1)$ is True.
Hence, by mathematical induction, we know that $P(n)$ is True for all $n \in \mathbb{N}$, i.e., for all $n \in \mathbb{N}, n \notin S$. However, since $S \subset \mathbb{N}$, this imples $S = \emptyset$, which forms the contradiction. Therefore, S has a smallest element. □

R We can see that the above induction hypothesis is carefully cooked up in the way that instead of assuming for a single $n \in \mathbb{N}$, we in fact assumed that $P(k)$ is True for all $k \in \mathbb{N}, k \leq n$. This technique is a sub-technique of proof by mathematical induction, often called **proof by strong mathematical induction**, where we first prove $P(1)$, then prove that $P(1)$, AND$\ldots$, AND $P(n) \Rightarrow P(n+1)$ is True. This is a valid proof technique to use since it is in fact using mathematical induction, with a little cleverness in the inductive hypothesis. This proof technique is sometimes useful and easier to prove properties of $\mathbb{N}$.

2.5 The Final Axiom

In this chapter, we aim to fully study the basics of the words $\mathbb{R}$. From Chapter 1, we know that $\mathbb{R}$ is **defined to be a field** with addition and multiplication. In this chapter, we also know that $\mathbb{R}$ is **defined to be an ordered field**. In fact, to make it easier to talk about and understand conversations, we need to define another intuitive property of emotions called **completeness**.
To make it simple, *completeness* simply means that for every set of words from $\mathbb{R}$ that is "emotionally bounded", their emotions approaches both upwards and downwards to some emotion values that are unique. This is just very intuitive, since we can just "order" the words and "figure out" the approaching emotional value. However, this cannot be formulated as a Theorem, because all previous definitions and properties of $\mathbb{R}$ do not deal with the idea of "approaching", hence it is impossible to prove this idea (Feel free to try it out, but you will get into trouble talking about "approaching").

First of all, let us make clear of the concept of *"approaching"* by defining "*emotional bounds*" for any ordered field $\mathbb{F}$.

Definition 2.5.1 For any ordered field $\mathbb{F}$, A set $S \subset \mathbb{F}$ is called **bounded** (above) if there exists an $a \in \mathbb{F}$ such that for all $s \in S$, $s \leq a$. **Bounded below** is defined similarly. In this case, a is called an **upper bound** of S. **Lower bound** can be defined similarly as well.

■ **Example 2.9 — Bounds.** ■

Let the ordered field be $\mathbb{R}$.

- Let $S = \{1,2,3\} \subset \mathbb{R}$. It is very easy to see that $3, 3.1, 4, 50$ are all upper bounds of S, $0, 0.2, -5, 1$ are all lower bounds of S. This tells us that upper and lower bounds are far from being *unique*.
- Let $S = \mathbb{R} \subset \mathbb{R}$. We can actually see that S is not bounded above or below. This can be proved using proof by contradiction. If S is bounded above by some upper bound x, then since $x \in \mathbb{R}$, $x+1 \in \mathbb{R}$, which means that x is not an upper bound of S, which gives us the desired contradiction. Lower bounds can be proved similarly. Hence, not all subsets of $\mathbb{R}$ have upper bounds or lower bounds.

Definition 2.5.2 An ordered field $\mathbb{F}$ is called **complete** if every non-empty bounded above $S \subset \mathbb{F}$ has a **least upper bound** $x \in \mathbb{F}$, which means for all upper bounds $x' \in \mathbb{F}$ of S, $x' \geq x$. x is denoted as $\sup(S)$.

Theorem 2.5.1 $\mathbb{R}$ is complete.

This is purely another **definition**, and we assume without proof that this is True.

R It is actually equivalent to denote the completeness definition for $\mathbb{R}$ (or for any ordered $\mathbb{F}$) to be every non-empty bounded below $S \subset \mathbb{R}$ has a **greatest lower bound** $x \in \mathbb{R}$, which means for all lower bounds $x' \in \mathbb{R}$, $x' \leq x$. x is denoted as $\inf(S)$. The equivalence can be seen by forming $S' = \{x : -x \in S\}$ for any bounded above nonempty $S \subset \mathbb{R}$.

Now, we have defined and everything we need to understand our words $\mathbb{R}$. However, I do want to remind you of a major problem: we took the definition of the set $\mathbb{R}$ for granted!!! We simply never defined what $\mathbb{R}$ is, just made a couple of property definitions (axioms) without proof, and defined $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{N}$ using $\mathbb{R}$. In fact, there is the beautiful definition for $\mathbb{R}$ which summarizes our Math Language 101: Learning the words. After this, we can go on the journey to explore and have interesting conversations.

Definition 2.5.3 We define $\mathbb{R}$, the **real numbers** to be the **unique ordered and complete field**, i.e. the **unique** set satisfying Theorem 1.2.1, Theorem 2.3.2, and Theorem 2.5.1.

3. Conversations with $\mathbb{R}$

In previous two chapters, we learned about the alphabet and words used in mathematical language, which is $\mathbb{R}$. Now we are ready to have conversations with these words. Since math words $\mathbb{R}$ is closed under the operations addition and multiplication, although alphabets generate words and words generate sentences, they mean the same in math.
In this chapter, we will mainly be talking about 1 input 1 response conversations, meaning that for every spoken input, there will be only one response. With this idea, we can start to analyze lots of interesting conversations, their emotions and information presented in these conversations.

3.1 Functions

3.1.1 The Basics

Let us start by defining general conversation between two person using two alphabets (which are sets). Let us keep in mind that when we are dealing with specific mathematical structures that are closed under some operation, alphabets have the same meaning as words and sentences.

Definition 3.1.1 A **function** f between sets X and Y (conversation between two person using two alphabets) is a communication log that corresponds each $x \in X$ with an element $y \in Y$. We say that "x maps to y by f", and we often denote $f : X \to Y$, $x \mapsto y$ and $y = f(x)$. One calls y the **image** of x and x the **preimage** of y. We call X the **domain** of f and Y the **codomain** of f. We call $f(X) = \{f(x) : x \in X\}$ to be the **image** of f.

■ **Example 3.1 — Functions.** ■

- In the set $X = \{1,2,3\} = Y$, we can have a function $f : X \to Y$ such that $f(1) = 2$, $f(2) = 3$ and $f(3) = 1$. We can also have a function $g : X \to Y$ such that $g(1) = g(2) = g(3) = 1$. However, if we associate $1 \in X$ to both $2 \in Y$ and $3 \in Y$, then we will not get a function. An important thing to note here is that for every $x \in X$ and any function $f : X \to Y$, $f(x)$ is **unique**.
- When $X = Y = \mathbb{R}$, we can get a lot of familiar functions, such as $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$ for all $x \in \mathbb{R}$. We can see that $1 \mapsto 1$, $2 \mapsto 4$ and $-3 \mapsto 9$. We can see that the *image* of f is

$\mathbb{R}^+ \cup \{0\}$. Another example can be $g : \mathbb{R} \to \mathbb{R}$ where $g(x) = \sin(x)$ for all $x \in \mathbb{R}$. In this case, $0 \mapsto 0$, $\frac{\pi}{2} \mapsto 1$ and $\frac{3\pi}{2} \mapsto -1$. In this case, the image of g is $[-1, 1]$.
- When $X = \mathbb{N}$, $Y = \mathbb{R}$, and if we have any function $f : X \to Y$, then we call the set of images of f, namely $\{f(1), f(2), \ldots\}$ a (real) **sequence**. As an example, if we have $f : \mathbb{N} \to \mathbb{R}$ such that $f(x) = x + 2$, then $f(1) = 3, f(2) = 4, f(3) = 5, \ldots$ is a sequence.
- In previous sections, we haven't defined the notion of "norm" rigorously. In fact, a norm N on $\mathbb{R}$ is a function $N : \mathbb{R} \to \mathbb{R}$ such that $x \mapsto |x|$ and satisfying the norm defining properties.

R A very important thing to notice is that in the definition of functions, the person with alphabet X uses *all* letters in the alphabet, while the person with alphabet Y does not necessarily use up all letters in Y, i.e. every $x \in X$ is mapped by f to some $y \in Y$, but not every $y \in Y$ is mapped from some $x \in X$.

From above remark and the definition of functions, we understand that not necessarily all letters in Y are mapped from some letter in X ($f(X)$ may not be equal to Y) and not necessarily each letter in X is mapped differently to some letter in Y. Functions with these type of properties are more desirable and very useful for analysis, so we would like to give them specific terminologies.

Definition 3.1.2 We say a function $f : X \to Y$ is **injective** if for all $x, x' \in X$, $f(x) \neq f(x')$. We say a function $f : X \to Y$ is **surjective** if for all $y \in Y$, there exists $x \in X$ such that $x \mapsto y$, i.e. $f(X) = Y$. We say f is **bijective** if it is both injective and surjective.

■ Example 3.2 — Injectivity, Surjectivity and Bijectivity. ■

- Same as in example 3.1.1, if $X = \{1, 2, 3\} = Y$, f defined in 3.1.1 is both *injective* and *surjective*, hence *bijective*. However, the g defined in 3.1.1 is not *injective* nor *surjective*. For any set $X = Y = \{1, 2, 3, \ldots, n\} \subset \mathbb{N}$, the set of bijective functions $f : X \to Y$ is called the set of **permutations** on X, since each function is essentially "permuting" the elements $1, 2, \ldots, n$.
- We call a function $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = a_0 + a_1 x + \ldots + a_n x^n$ for all $x \in \mathbb{R}$ where $a_i \in \mathbb{R}$ for all i and $n \in \mathbb{N}$ a (real) **polynomial** of degree n. An example would be $f(x) = x^2$, this is a real polynomial of degree 2, we can see that in this case f is not surjective or injective, since $f(1) = f(-1)$ and $f(x) \geq 0$ for all $x \in \mathbb{R}$. However, one can show that $f(x) = x^3$ is *bijective*. (*Hint:* Showing it is surjective is simple, since we can solve $x^3 = a$ for all $a \in \mathbb{R}$. To show it is injective, consider $x^3 - y^3 = 0$ and factor.)

It is a common scenario for us to consider only the conversation involving 1 person saying *specific* "words" and the other person's responses. This natural language intuition gives rise to the "restriction" of a function.

Definition 3.1.3 Given $f : X \to Y$, and $A \subset X$, we define the **restriction of f to A** to be the function $f|_A : A \to Y$ such that $f|_A(a) = f(a)$ for all $a \in A$. It is easy to see that since f is a function, $f|_A$ is also a function.

R $f|_A$ can be seen as the conversation log for usage of only words in A and their responses.

■ Example 3.3 — Function Restriction. ■

- Sometimes non-injective functions can restrict to injective ones. As an example, $f(x) = x^2$ on $\mathbb{R}$ restricts to an injective function $f|_{\mathbb{R}^+} : \mathbb{R}^+ \to \mathbb{R}$.

- By changing the codomain to be the function's image, we get a surjective function. With same example as above, we know that $f(x) = x^2$ is surjective if we consider $f : \mathbb{R} \to \mathbb{R}^+ \cup \{0\}$.

3.1.2 Functions with Group Structure

Let us consider two functions $f : X \to Y$ and $g : X \to Y$ between set X and group Y closed under $\star$. Then, we can define $f \star g$ such that $f \star g : X \to Y$ where $x \mapsto f(x) \star g(x)$. It is easy to see that $f \star g$ is indeed a function.

Intuitively, this means that if we have two separate conversations between two people using the same two alphabets, we can consider the response emotion together for each input. This is what we do all the time, consider everybody's opinion on the same input we provide. Everybody responding needs to use the same alphabet otherwise the combination won't make any sense. In fact, this gives the group operation for functions from X to Y, since multiplying two such functions still gives us a function from X to Y, provided that Y has a group structure.

Similarly, if Y is a (ring) field with another operation $\cdot$, we can also define $f \cdot g : x \mapsto f(x) \cdot g(x)$, and $-f : x \mapsto -f(x)$ and possibly also $f^{-1} : x \mapsto f(x)^{-1}$. We can see that the set of functions from given set X to given set Y has the exact same structure that Y has, with the operations defined as above.

What is truly powerful is that the set of bijective functions $f : X \to X$ for any set X with or without any structure is already a group, equipped with the following multiplication method:

Definition 3.1.4 Given any sets X, Y, Z, and functions $f : X \to Y$, $g : Y \to Z$, we define the **composition of** f **and** g, read as "g compose f" or "f pull back g" to be $f^*g : X \to Z$ such that $x \mapsto (f^*g)(x) = g(f(x))$.

R The above definition is a bit general, but we can see that for two functions $f, g : X \to X$, we can compose them together using composition defined above. It is important that we keep the codomain of f and domain of g to be the same for composition to make sense. Hence for functions from X to X, we assume them to be bijective for composition to make sense. Many sources use $g \circ f$ to represent g compose f, however it is applying f first, then g. The backward order in the terminology here is why I do not like it. I will use f^*g in all later occurrences.

In fact, with above remark we can see that given any set X, the set of functions from X to itself is **closed** under composition. It is easy to see that this operation is associative by definition.

Theorem 3.1.1

$$(f^*g)^*h = f^*(g^*h)$$

Proof. On the left hand side, we get $x \mapsto g(f(x))$ by f^*g pull back, then $g(f(x)) \mapsto h(g(f(x)))$ by h. On the right hand side, we get $x \mapsto h(g(x))$ by g^*h, then mapsto $h(g(f(x)))$ by f pullback. They are equal. □

Then another question before we conclude that given any set X, the set of bijective functions from X to itself is a group under composition: What is the identity element? In fact, it is easy to see that $f : X \to X$ such that $x \mapsto x$ is the identity element under composition (Try it!). This function is called the **identity function**, denoted id.

In fact, we also need to define the inverse operation of composition.

Definition 3.1.5 Given any bijective function $f : X \to Y$, we define the **inverse function** of f to be $f^{-1} : Y \to X$ such that $f^*(f^{-1}) = id = (f^{-1})^*(f)$, i.e. $f^{-1} : y = f(x) \mapsto x$.

R The reason for this definition requiring f being bijective is that we need every $y \in Y$ to be the image of some $x \in X$, suggesting that f must be *surjective*. Besides, in order for f^{-1} to be a function, we need each $y \in Y$ only has one $x \in X$ such that $f(x) = y$, meaning that f must also be *injective*.
Another thing is that we use f^{-1} both for multiplicative inverse at the beginning of this chapter and for composition inverse in the above definition. To ressolve ambiguity, we *always* use f^{-1} for composition inverse, and $\frac{1}{f}$ for multiplicative inverse, when Y is a group under multiplication.

Theorem 3.1.2 Given any set X, the set of bijective functions from X to itself forms a group under operation composition.

When you have a story, you tell person A, person A retells it to person B. The emotion that person B gets will be the emotion getting from person A telling person A's emotion getting from your version of the story. This is the composition operation! If the response of person A uses all letters in person A's alphabet and you have 1-1 correspondance between your input and person A's response, it is possible to based on person A's response to figure out what you said to person A, we usually use this technique when we are trying to remember what we have said. This is inversion.

Example 3.4 — Composition and Inversion of Functions.

- The set of polynomials on $\mathbb{R}$ of any finite degree (as defined before) is a **ring**. Why? First, it is closed under addition with additive identity being 0 (treated as a 0 degree constant polynomial). The additive inverse is also defined, being negative of each polynomial. Then, it is closed under multiplication, since we get a polynomial when two polynomials multiply, the identity for this operation is the constant polynomial 1. However, it is important to note that polynomials with addition and function composition is *not* a ring. This is because not all polynomials are bijective, i.e. $f(x) = x^2$.
- Given set $X = \{1, 2, 3, \ldots, n\}$, we know that the set of *permutations* is a group from Theorem 3.1.2, since all permutations are bijections. In fact, we call this group S_n for each $n \in \mathbb{N}$. It is a very important group in abstract algebra!

3.1.3 Graphs of Functions

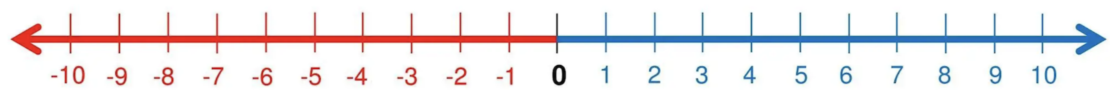
For functions from $\mathbb{R}$ to $\mathbb{R}$, we can actually plot them on a piece of paper. This involves the idea of defining the *graph* of functions, which also leads to one definition of the information of each conversation, which will be covered in future chapters. Let us first define abstractly what it means for the graph of functions.
Before we define the graphs of functions, we will first define a new operation on sets, called the Cartesian Product.

Definition 3.1.6 Given two sets X, Y, we define the **Cartesian Product** of X and Y to be the set $X \times Y = \{(x, y) : x \in X, y \in Y\}$ which is the set of ordered pairs of elements in X and Y.

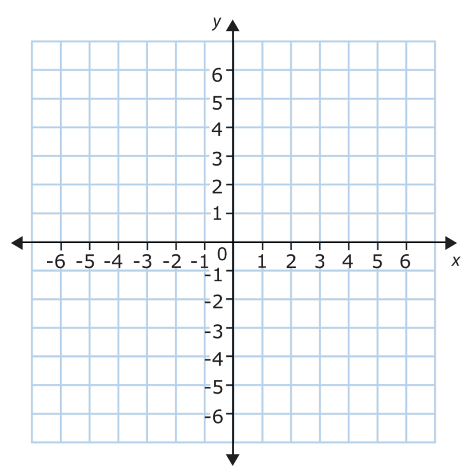
R By *ordered pairs*, we mean that even if $X = Y$, for $a \neq a' \in X = Y$, (a, a') and (a', a) are *two distinct elements* in $X \times Y$.

Definition 3.1.7 Given a function $f : X \to Y$, we define the graph of f, $\text{Graph}(f) : X \to X \times Y$, such that $x \mapsto (x, f(x))$. This is actually a function from X to the Cartesian Product of X and Y. We call the image of this function **the graph of f**.

The intuition behind "graph" is that it is a conversation logging between two person using alphabets X and Y, but instead of logging only the 2nd person's response, we also log the 1st person's input.
After defining abstractly what a *graph* of function means, we need to know how to draw the graph. As a starting point, we will define the notion of a "number line." A **number line** is a visual representation of $\mathbb{R}$, where numbers in $\mathbb{R}$ are ordered from left to right, as shown below:

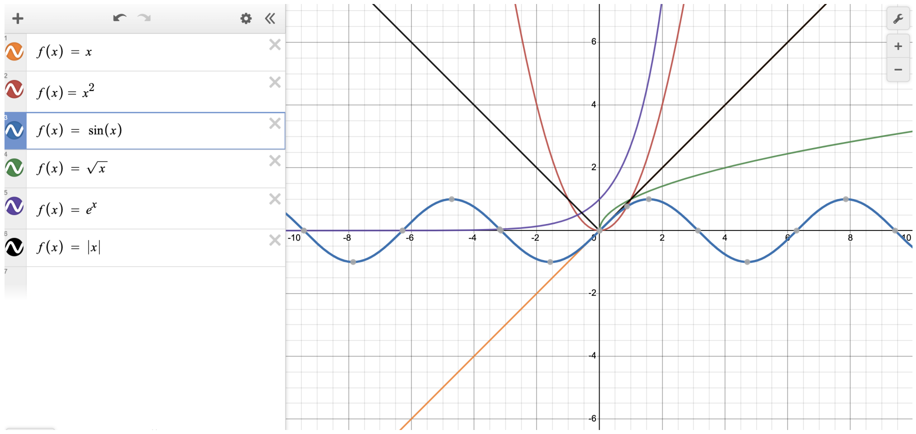


Moving from right to left, the number gets smaller. Moving from left to right the number gets bigger. The red half of the number line above represents *negative* emotions, while the blue half represents *positive* emotions. In order to graph a function $f : \mathbb{R} \to \mathbb{R}$, or even a function $f : X \to Y$ where $X, Y \subset \mathbb{R}$, we first combine two number lines together to make a number plane, like below:

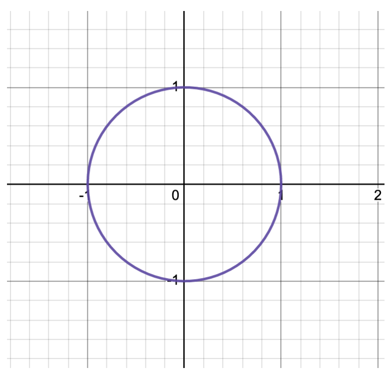


Then, we will label the horizontal axis the x-axis, and the vertical axis y-axis or $f(x)$-axis. Each point on the plane can be labelled by (x, y) where x represents its horizontal position and y represents its vertical position. We call the point $(0, 0)$ the **origin**.
In fact, given any function $f : \mathbb{R} \to \mathbb{R}$, each element in the graph of f is a point on the number plane. After we find all the points on the number plane, we connect them together to form the "graph" of function f.

■ **Example 3.5 — Graphs of Functions.** ■



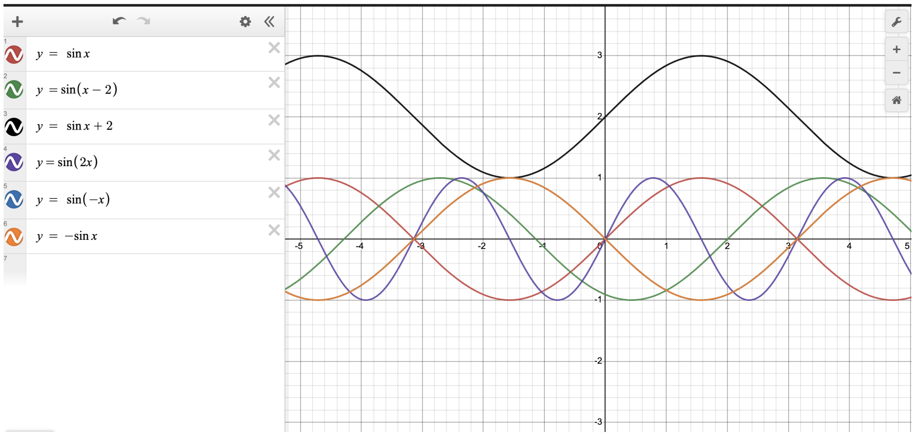
Above is the graphs of several important functions, some of which we haven't seen/defined, but we will in the future. A great tool to use is Desmos, where you can see the graph of all kinds of functions from $\mathbb{R}$ to $\mathbb{R}$. In fact, we can graph all kinds of shapes on the number plane, but is everything representing a function? Not necessarily, let us see an example.



If this is a function, the point 0.5 has two points being its image, which is impossible. It is not hard to realize that for every graph of functions, for every $x \in \mathbb{R}$, if we draw a vertical line through x on the x-axis, this line can only meet the graph at *one point*. In fact, this is a way to tell whether a graph represents a valid function, called **the Vertical Line Test.**

Let us consider the $f(x) = x^2$ graph included above, we know that this function is not *injective*, but can we tell from the graph? Yes, in fact, if we draw a horizontal line through 2 on the y-axis, we can see that it meets the graph at 2 points, meaning that two x values are mapped to the same y value, hence $f(x) = x^2$ is not injective. We can check every horizontal line from each value on the y-axis with the graph to see whether a function is injective or not, this is called the **Horizontal Line Test**.

There are several things you can do to a graph, including translation ($y = \sin(x) + 2$, $y = \sin(x - 2)$), stretching ($y = \sin(3x)$), and reflection $y = \sin(-x)$, $y = -sin(x)$. See below for the effects of these operations:



3.1.4 Countability

In this subsection, we will describe a very important idea called *countability*, meaning whether a set S can be *counted* with natural numbers. By definition of natural numbers, we know that S can be finite or infinite, but it should be "as many as" the natural numbers.

Definition 3.1.8 A set S is **countable** if we can find a surjective function $f : \mathbb{N} \to S$, i.e. we have a way to count *every* element of S using natural numbers.

R It can be seen from the above definition that we do not require f to be *injective*, meaning that we can count S's elements *repetitively*.

Example 3.6 — Countable Sets.

- Finite sets are countable. Why? Consider any finite set $S = \{s_1, \dots, s_n\}$ for some $n \in \mathbb{N}$. Simply take $f : \mathbb{N} \to S$ such that $f(x) = s_x$ if $1 \le x \le n$, and $f(x) = s_1$ otherwise. It is easy to see that f is surjective.
- $\mathbb{N}$ is countable. [Left as an exercise.]
- $\mathbb{Z}$ is countable. Why? Let us define $f : \mathbb{N} \to \mathbb{Z}$ such that $f(x) = -\frac{x-1}{2}$ if x is odd, and $f(x) = \frac{x}{2}$ if x is even. It can be shown that this function is surjective. This actually shows that on the set level, a set that is *bigger* than $\mathbb{N}$ *may still be countable*.
- Given a countable set S, every $A \subset S$ is countable. [By definition, left as an exercise.]

We do have some important results with regards to countability of sets, one of which is that countability holds with $\mathbb{N}$-unions.

Theorem 3.1.3 If $\{S_i\}_{i\in\mathbb{N}}$ are a collection of countable sets, $S = \bigcup_{i=1}^{\infty} S_i$ is also countable.

Proof. This proof is a very clever trick to count the elements in the union. We can observe that each element $a \in S$ can be characterized uniquely by (i, x) where i represents which S_i that a is in, and x represents the natural number we give $a \in S_i$ since S_i is countable. Given this $a \in S$ and (i, x) representing a, we consider $m = i + x$, and we define the function $f : \mathbb{N} \to S$ such that

$$f(\frac{(m-1)(m-2)}{2} + i) = a$$

for all $a \in S$, $m \in \mathbb{N}$. Basically we are counting elements in S in the following order: $s_{11}, s_{12}, s_{21}, s_{13}, s_{22}, s_{31}, \dots$, where s_{11} is the first element we count in S_1 and so on.

Since each $a \in S$ gets a unique (i,x) pair, and each pair corresponds to a unique $\frac{(m-1)(m-2)}{2}+i$ value, we know that f is surjective. Hence, S is countable. $\square$

> **R** In fact, we can try to do the above proof using mathematical induction, but even if we show that $S_n = \bigcup_{i=1}^{n} S_i$ is countable for all $n \in \mathbb{N}$, it is still different (requires more theory) to show that S is countable since S is an infinite union of sets.

Above proof for the theorem also can be used to prove another important theorem, whose proof will left as an exercise:

Theorem 3.1.4 The rational numbers $\mathbb{Q}$ is countable.

Next, what about $\mathbb{R}$? In fact, $\mathbb{R}$ is *not* countable. Let us see why.

Theorem 3.1.5 $\mathbb{R}$ is not countable.

Proof. Suppose by contradiction that $\mathbb{R}$ is countable. Then, we can define a function $f:\mathbb{N}\to\mathbb{R}$ such that f is surjective. Since f is surjective, we can actually define an *injective* inverse function $f^{-1}:\mathbb{R}\to\mathbb{N}$, mapping each number in $\mathbb{R}$ to their counting order, and in fact $f^{-1}(\mathbb{R})\subset\mathbb{N}$.
Let us consider the numbers $0.2111111\ldots, 0.22111111\ldots, 0.22211111\ldots, 0.222211111\ldots, \cdots$
Each of the numbers are *different* since they has different number of 2's. Let S be the set of these numbers. We know that they all have different counting orders. Also, we can define $g:S\to\mathbb{N}$ such that $g(s) =$"number of 2's in s.", and in fact g is a bijection onto $\mathbb{N}$.
Since we can find a bijective mapping from S to $\mathbb{N}$, it is not possible that $f^{-1}(S)\neq\mathbb{N}$ since we already know f^{-1} is injective on all of $\mathbb{R}$. Why is this? Consider any $h:\mathbb{N}\to\mathbb{N}$ such that $h(n) = f^{-1}(g^{-1}(n))$. Then, if h is not onto, we know that there are $n_1,n_2\in\mathbb{N}$ such that $h(n_1)=h(n_2)$, however $g^{-1}(n_1)\neq g^{-1}(n_2)$, thus by definition of h, f^{-1} cannot be injective. Therefore, h must be onto $\mathbb{N}$. It is easy to see that $g^{*}h = f^{-1}|_S$ by definition of h, thus f^{-1} is bijective.
This means $f^{-1}(S)=\mathbb{N}$, then S must be equal to $\mathbb{R}$, by definition of f^{-1}. However, simply consider the number 0.3, which is different from all numbers in S but is still a real number in $\mathbb{R}$. Thus, $S\neq\mathbb{R}$, hence we reached a contradiction. $\mathbb{R}$ is uncountable. $\square$

3.2 Limits

In the last subsection, we understand what is a conversation with $\mathbb{R}$ (functions) and different types of functions, understands the operations on these conversations and how to log them (graphs of functions). In the end, we talked about the concept of countability defined using functions.
In this subsection, we will try do make sense of the idea that "where the conversation is going". In fact, this is the idea similar to what we have for defining sup and inf for $\mathbb{R}$: the idea of "approaching". We will give more concrete definitions of this using *limits*.

3.2.1 Limit Definition and Metric

Definition 3.2.1 Given a function $f:A\to\mathbb{R}$, $A\subset\mathbb{R}$, and given $a\in\mathbb{R}$ we write $\lim_{x\to a} f(x) = l$ (called as x goes to a, $f(x)$ *approaches* l or the **limit** of f as x goes to a is l) if for all $\varepsilon>0$, we can find a $\delta>0$ such that for all $x\in A$, "$0<|x-a|<\delta \Rightarrow |f(x)-l|<\varepsilon$" is True. We say the limit of f as x goes to a **exists** if we can find such an $l\in\mathbb{R}$, if not, we say that the limit **does not**

exist (DNE).

> **R** The formulation of above definition suggests that depending on what we choose for ε, δ can change dependently. However, for every value of ε, finding at least one δ that works in the definition is required.

It is in fact the most complicated definition so far, because it involves a lot of logical statements and math symbols. I will try to explain the intuition behind the definition of limits.
As we mentioned above, we would like to know "where the conversation was going", in fact "If I am going to say something with emotion a, what would the other person's response emotion look like?". It would be a *successful and accurate* prediction of the other person's response emotion l if I am able to say something close to a that can trigger this person's response to be as close as possible to l. This is exactly the logic behind predicting the emotional flow of the conversation. That's why the limit definition precisely reflects the idea that, if: no matter how close (any $\varepsilon > 0$) I would want, I can say something close to what I am predicting that I would say ($|x-a| < \delta$), and making the other person's response emotion $f(x)$, ε-close to my prediction l. Then: this prediction l must be where this conversation is going (emotion-wise) if I'm going to say a word with emotion a.
With the above intuition, it actually subtly presents the idea that we need to have a sense of "closeness" between two words with some emotion values. In fact, having the idea of closeness of emotions between words is *much more important* than the idea of "what the emotion of a *single* word is", and the idea of "approaching" is very fundamental in calculus and analysis. In many real analysis sources, authors define *closeness (emotional distance)* at the start, and define *norms* based on the idea of closeness. I will do the other way around, and I believe this way is more intuitive with our natural language analogy.

Definition 3.2.2 We can define a map $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that $d(x,y) = |x-y|$ where $|\cdot|$ is a norm defined on $\mathbb{R}$. We call the map d a **metric** on $\mathbb{R}$.

Example 3.7 — Metrics.

- In example 2.7, we described several examples for norms. The first one gives rise to **discrete metric**, which is defined as:

$$d(x,y) = \begin{cases} 0 \text{ if } x = y \\ 1 \text{ otherwise.} \end{cases}$$

- The second example is the Euclidean Norm. With the definition of Euclidean norm, we can define **Euclidean metric**, which is defined as:

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

Intuitively, as soon as we have a metric, we have a sense of emotional closeness.

> **R** Without specification, we always assume the metric we are working with is the Euclidean metric.

The idea of changing closeness with response value (ε) resulted by changing closeness of input (δ) is present in various proofs involving limits. We will begin by some interesting results that can be proved directly from limit definition.

Theorem 3.2.1 Given the function $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = ax + b$, where $a, b \in \mathbb{R}$. Given any $y \in \mathbb{R}$, $\lim_{x \to y} f(x) = f(y) = ay + b$.

Proof. By the definition of limits, we first let $\varepsilon > 0$ be given. We would like to show that we can achieve this ε-distance between $f(x)$ and $f(y)$ by choosing a δ-distance between x and y. The remaining work is about finding this δ. But how?
Let us suppose that we have already found a δ, then for any $|x - y| < \delta$, $|f(x) - f(y)| = |ax + b - ay + b| = |a||x - y| < |a|\delta$. In order to make sure that $|f(x) - f(y)| < \varepsilon$, we get the inequality that $|a|\delta \leq \varepsilon$, $\delta \leq \frac{\varepsilon}{|a|}$. This shows that any δ less than or equal to $\frac{\varepsilon}{|a|}$ works, then we need to specify a single value of δ, for example $\frac{\varepsilon}{|a|}$. Then we are done. □

R The above proof is more of a thought process than a proof. A proof will go in the opposite direction as the above proof, namely starting by defining $\delta = \frac{\varepsilon}{|a|}$ and in the end show that $|f(x) - f(y)| < \varepsilon$ for this particular ε.

Sometimes more modifications on these inequalities before choosing a δ is required, for example in the below exercise:

Exercise: Show that $\lim_{x \to y} x^2 = y^2$. ■

Hint: We need $|x^2 - y^2| = |x - y||x + y| < \varepsilon$, $|x - y|$ needs to be smaller than δ. The problem is with $|x + y|$, consider triangle inequality $|x + y| = |x - y + 2y|$, and then solve for δ.

3.2.2 Limit Properties

Theorem 3.2.2 Suppose $f : A \to \mathbb{R}$, $A \subset \mathbb{R}$, and $\lim_{x \to a} f(x)$ exists for some $a \in \mathbb{R}$. Then, we know that the limit is unique.

Proof. We will prove by contradiction. Suppose $\lim_{x \to a} f(x) = l_1$ and $\lim_{x \to a} f(x) = l_2$, where $l_1 \neq l_2$. Then, we know that given any $\varepsilon > 0$, we can find $\delta_1, \delta_2 > 0$ such that for all $x \in A$, $|x - a| < \delta_1, |f(x) - l_1| < \varepsilon$, and for all $x \in A$, $|x - a| < \delta_2, |f(x) - l_2| < \varepsilon$. Let $\delta = \min\{\delta_1, \delta_2\}$, then we know that for all $x \in A, x < \delta$, both conditions satisfy. Hence, $|f(x) - l_1| = |l_1 - f(x)| < \varepsilon$ and $|f(x) - l_2| < \varepsilon$. Thus by triangle inequality:

$$|l_1 - f(x) + f(x) - l_2| = |l_1 - l_2| \leq |l_1 - f(x)| + |f(x) - l_2| < 2\varepsilon.$$

Therefore, we have shown that $|l_1 - l_2| < 2\varepsilon$ for any $\varepsilon > 0$, which means $l_1 = l_2$, and we get a contradiction. □

The intuition for above proof is very straightfoward, we are basically saying if we have two correct predictions l_1, l_2 of the response when I say y in a conversation, they must be the same. Why? Because I can say something x that is very close to input y to make the response as close to l_1 as I want, and similarly, I can also say something x' that make the response as close to l_2 as I want.
Therefore, I can say something that makes l_1 and l_2 as close to *each other* as I want, meaning that they must be the same prediction.

In fact, just like response emotions add, subtract (add by inverse), multiply and divide (multiply by inverse), the predictions of response emotions should also add, subtract, multiply and divide. This result is summarized below:

Theorem 3.2.3 **Properties of Limits:** Given functions $f, g : A \to \mathbb{R}$ for some $A \subset \mathbb{R}$:
1. $\lim_{x\to a}(f \pm g)(x) = \lim_{x\to a} f(x) \pm \lim_{x\to a} g(x)$.
2. $\lim_{x\to a}(fg)(x) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x)$
3. If we can find a m such that for all $|x-a| < q$, $g(x) \neq 0$, then we know that $\lim_{x\to a} \frac{f}{g}(x) = \lim_{x\to a} f(x)/\lim_{x\to a} g(x)$.

Proof. For property 1., the proof is extremely similar to the proof used for Theorem 3.2.2, so is left as an exercise.
To make it simpler for property 2 and 3, we suppose $\lim_{x\to a} f(x) = l, \lim_{x\to a} g(x) = m$. Let $\varepsilon > 0$ be given.
For property 2, our ultimate goal is to make $|f(x)g(x) - lm| < \varepsilon$ for $|x-a| < \delta$ for some δ. What we know is that we can make $|f(x)-l|, |g(x)-m|$ as close as possible. We can observe that:

$$\begin{aligned} |f(x)g(x) - lm| &= |(f(x)-l)(g(x)-m) + mf(x) + lg(x) - 2ml| \\ &= |(f(x)-l)(g(x)-m) + m(f(x)-l) + l(g(x)-m)| \\ &\leq |f(x)-l||g(x)-m| + |m||f(x)-l| + |l||g(x)-m| \\ &< \varepsilon \end{aligned} \tag{3.1}$$

It is great that we get $|f(x)-l|, |g(x)-m|$ to appear in each term of addition. Thus, the most straight forward thing to do is to make sure $|f(x)-l||g(x)-m| < \frac{\varepsilon}{3}$, $|m||f(x)-l| < \frac{\varepsilon}{3}$ and $|l||g(x)-m| < \frac{\varepsilon}{3}$. Furthermore, to make $|f(x)-l||g(x)-m| < \frac{\varepsilon}{3}$, we can make $|f(x)-l| < \frac{\sqrt{\varepsilon}}{\sqrt{3}}$ and $|g(x)-m| < \frac{\sqrt{\varepsilon}}{\sqrt{3}}$. We know that we can do these things, it is left to make it formal.
Suppose $|x-a| < \delta_1$ makes $|f(x)-l| < \frac{\sqrt{\varepsilon}}{\sqrt{3}}$, $|x-a| < \delta_2$ makes $|g(x)-m| < \frac{\sqrt{\varepsilon}}{\sqrt{3}}$, $|x-a| < \delta_3$ makes $|f(x)-l| < \frac{\varepsilon}{3|m|}$, and $|x-a| < \delta_4$ makes $|g(x)-m| < \frac{\varepsilon}{3|l|}$. Finally, we set our $\delta = \min\{\delta_1, \delta_2, \delta_3, \delta_4\}$ and we are done.
For property 3., our ultimate goal is to prove that $\lim_{x\to a} \frac{f}{g}(x) = \lim_{x\to a} f(x)/\lim_{x\to a} g(x)$, however with property 2. already proved, we only need to show that $\lim_{x\to a} \frac{1}{g}(x) = \frac{1}{\lim_{x\to a} g(x)}$. Hence, we need that $|\frac{1}{g(x)} - \frac{1}{m}| < \varepsilon$, thus we know:

$$\begin{aligned} |\frac{1}{g(x)} - \frac{1}{m}| &= |\frac{m-g(x)}{g(x)m}| \\ &= \frac{|g(x)-m|}{|g(x)| \cdot |m|} \\ &= |g(x)-m| \cdot \frac{1}{(|g(x)| \cdot |m|)}. \end{aligned} \tag{3.2}$$

Let us first bound $|g(x)|$ by $|m|$ using the triangle inequality that $|m| - |g(x)| \leq |g(x)-m|$. Choose $\delta_1 > 0$ such that for all $0 < |x-a| < \delta_1$, we have $|g(x)-m| < \frac{|m|}{2}$, thus $|g(x)| > \frac{|m|}{2}$. Hence, we know that $\frac{1}{|g(x)|\cdot|m|} \leq \frac{2}{|m^2|}$. Pick $\delta_2 > 0$ such that for all $0 < |x-a| < \delta_2$, we have $|g(x)-m| <$

$\min\{\frac{|m^2|\varepsilon}{2}, \frac{|m|}{2}\}$, in this case, we know that:

$$|g(x)-m|\cdot\frac{1}{(|g(x)|\cdot|m|)} < \frac{|m^2|\varepsilon}{2}\cdot\frac{2}{|m^2|} = \varepsilon. \tag{3.3}$$

Then, we can let $\delta = \min\{\delta_1, \delta_2, q\}$ since we can't have $g(x)$ to be 0 anywhere, and then we have proved the claim. $\square$

Theorem 3.2.4 Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ with $A, B \subset \mathbb{R}$ and $f(A) \subset B$. Suppose we have $\lim_{x\to a} f(x) = l$ and $\lim_{y\to l} g(y) = g(l)$, then we have $\lim_{x\to a} g(f(x)) = g(l)$.

Proof. Let $\varepsilon > 0$ be given. Then we know that we can find δ_1 such that $|y-l| < \delta_1$, we have $|g(y)-g(l)| < \varepsilon$. Next, we also know that we can find δ such that $0 < |x-a| < \delta$, $|f(x)-l| < \delta_1$, and in this case we know $|g(f(x))-g(l)| < \varepsilon$. $\square$

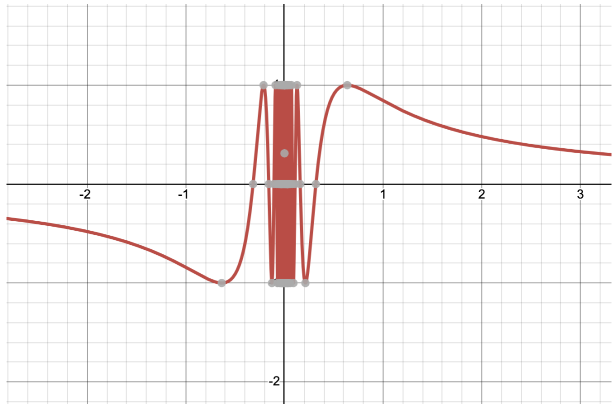
The above theorem and proof follow from our natural language analogy. The theorem is dealing with the conversation that invovles "retelling", meaning I tell you, you retell person A. The theorem suggests that if I can predict that your response is l if I say a, and you can predict that person A's response is $g(l)$ when you say l, then we know that we can predict when I say a, person A's response is $g(l)$, which is very intuitive.
The way we try to prove this theorem, we use the idea that since you can predict person A's response $g(l)$, you can say something close to l and prompt person A to say something as close to $g(l)$ as we want. Then, base on how close you need to be to l, I can say something close to a to achieve this, since I can predict that your response is l if I say a, which is also very intuitive.

3.2.3 Limits DNE, one-sided limits and limits of infinity

In the above paragraphs, we have showed several examples of limits and proved properties of limits. There are also scenarios where the limit does not exist, meaning that we cannot find a prediction of where the response emotions are going. With mathematics symbols, it means that for every l, we can find an ε such that no δ works as in the limit definition, i.e. $\forall l \in \mathbb{R}.\exists\varepsilon > 0.\forall\delta > 0.\exists x \in A : 0 < |x-a| < \delta$ AND $|f(x)-l| \geq \varepsilon$. Let us consider some examples.

■ **Example 3.8** ■

Let us consider the function $f : \mathbb{R}\setminus\{0\} \to \mathbb{R}$ such that $f(x) = \sin(\frac{1}{x})$. Our claim is $\lim_{x\to 0} f(x)$ does not exist.



With the above graph of f, we can already see the problem. The response emotions oscillates too fast close to 0, making it impossible to predict. We are going to rigorously illustrate this idea.
Firstly, we know that $\sin(\frac{1}{x}) = 1$ if $\frac{1}{x} = \frac{\pi}{2} + 2\pi k = \frac{\pi+4\pi k}{2}$ for some $k \in \mathbb{Z}$, and $\sin(\frac{1}{x}) = -1$ if $\frac{1}{x} = \frac{3\pi}{2} + 2\pi k = \frac{3\pi+4\pi k}{2}$ for some $k \in \mathbb{Z}$. Let $l \in \mathbb{R}$ be given. Let $\varepsilon = 1$. We will consider two cases.
Firstly, $l > 0$. Pick any $\delta \in \mathbb{R}^+$, and choose k large enough so that $0 < \frac{2}{3\pi+4\pi k} < \delta$, then we know that if $x = \frac{2}{3\pi+4\pi k}$, $\sin(\frac{1}{x}) = -1$. In this case, $|f(x) - l| = |-1-l| = |l+1| > 1 = \varepsilon$.
If $l < 0$, for any $\delta \in \mathbb{R}^+$, we can choose k large enough so that $0 < \frac{2}{\pi+4\pi k} < \delta$, then we know that if $x = \frac{2}{\pi+4\pi k}$, $\sin(\frac{1}{x}) = 1$. In this case, $|f(x) - l| = |1-l| => 1 = \varepsilon$. Thus, we know that there is no δ that works for $\varepsilon = 1$, and the limit does not exist.
What we did above is simply use the notion of "closeness" and the fact that $\sin(\frac{1}{x})$'s response emotion oscillates very fast between 1 and -1 near 0 to show that the limit does not exist.

The definition of the limit and the definition of emotional closeness suggests that as long as I am close enough to a, no matter whether my input emotion is more negative/positive than a, I can prompt you to respond something close to l. However, there exists situations where whether my emotion is more negative/positive than a directly effects your response, and you respond differently depending on whether my emotion is more negative/positive than a. In this case, your emotional response when I say a is also not predictable since on each side I am having a *different* prediction. This motivates the concept of **one-sided limits**.

Definition 3.2.3 Given function $f : A \to \mathbb{R}$ where $A \subset \mathbb{R}$ and $a \in \mathbb{R}$, we write $\lim\limits_{x\to a^+} f(x) = l$ if $\forall \varepsilon > 0.\exists \delta > 0.\forall x \in A : x - a < \delta \Rightarrow |f(x) - l| < \varepsilon$. $\lim\limits_{x\to a^-} f(x) = l$ is defined similarly by replacing $x - a$ by $a - x$.

■ **Example 3.9** Let us consider the function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ such that $f(x) = \frac{x}{|x|}$, then we know that $f(x)$ on $\mathbb{R}^+$ is constant 1 and $f(x)$ on $\mathbb{R}^-$ is constant -1. In this case, we can easily show that $\lim\limits_{x\to 0^+} = 1$, and $\lim\limits_{x\to 0^-} = -1$. However, $\lim_{x\to 0} f(x)$ does not exist, since the function reaches two "different limits". We can show this by analyzing different cases of l. If $l \leq -1$ or $l \geq 1$, we can let $\varepsilon = 1$ then we know that no matter what δ we choose, there is one side of f that has distance greater than 1 from l. If $-1 < l < 1$, pick $\varepsilon = \frac{\min\{|1-l|,|l+1|\}}{2}$, i.e. half of the shorter distance between 1 and l and -1 and l, it is not possible to reach closer than ε for any δ. (Try drawing it out for better understanding).

■

From above example, we can actually summarize to another important result, which is the following:

Theorem 3.2.5 Given $f : A \to \mathbb{R}$ where $A \subset \mathbb{R}$, given $a \in \mathbb{R}$, $\lim_{x\to a} f(x)$ exists if and only if $\lim_{x\to a^+} f(x)$ and $\lim_{x\to a^-} f(x)$ both exists and are equal.

Proof. The proof of this theorem follows the same logic as Example 3.9, hence will be left as an exercise.

3.3 Sequences

3.4 Continuity

3.5 Differentiability