MATH 1210

Mathematical Discovery 1

current editor
Dr Paul Vrbik

last edit
March 28, 2017

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Contents

0 Basics

"Begin at the beginning," the King said, gravely, "and go on till you come to an end; then stop."

– Lewis Carroll, *Alice in Wonderland*

Note to students: The following is assumed background material presented in a more precise mathematical way. Do not stress too much about it as the 'real' material begins in Chapter 1. (That's why this is Chapter *Zero*).

0.1 Sets and Logic

We start with a set.

Definition 0.1 (set). A *Set*, in the mathematical sense, is a *finite* or *infinite* collection of *unordered* and *distinct* objects.

Anything surrounded by curly braces '{ }' is a set.

Example 0.2. A set of integers.

$$A = \{3, 8, 9, 10, 42, -3\}.$$

Definition 0.3. A set's *cardinality* (denoted by '| |' or '#') is the number of elements the set contains (finite or otherwise).

Example 0.4. A has cardinality 6.

$$\begin{aligned} |A| &= |\{3, 8, 9, 10, 42, -3\}| \\ &= 6 \end{aligned} \qquad \begin{aligned} \#A &= \#\{3, 8, 9, 10, 42, -3\} \\ &= 6 \end{aligned}$$

The most concise, or perhaps only, way to work with sets is to express everything with a *formal language* of logical symbols. Let us quickly review these symbols. taking for granted the definitions and notions of implication ($\implies$ and $\iff$) as well as the meaning of 'or' and 'and'. (Those who need a refresher can refer to the Lecture 1 slides.)

Definition 0.5 (Mapping). A mapping 'connects' elements of one set with another. Writing

$$\begin{aligned} M : A &\to B \\ a &\mapsto b. \end{aligned}$$

expresses: M maps $a \in A$ to $b \in B$.

Since functions are also maps—for instance, $f(x) = x^2$ can be given as the mapping:

$$\begin{aligned} f : \mathbb{Z} &\to \mathbb{Z} \\ x &\mapsto x^2 \end{aligned}$$

—the term 'map' and 'function' are sometimes used interchangeably.

When $B = \{\text{true}, \text{false}\} = \{\mathsf{T}, \mathsf{F}\}$ in Definition 0.5 the map is called a *predicate*.

Definition 0.6 (Predicate). An operator of logic that returns true (T) or false (F) over some universe (e.g. $\mathbb{Z}$, days of the week, or reserved words).

Example 0.7. A predicate given by

$$\begin{aligned} P : \mathbb{Z} &\to \{\mathsf{T}, \mathsf{F}\} \\ P : x &\mapsto \begin{cases} \mathsf{T} & \text{if } x \text{ is prime} \\ \mathsf{F} & \text{otherwise} \end{cases} \end{aligned}$$

evaluates to true only when x is a prime number:

$$P(7) = \mathsf{T} \qquad P(8) = \mathsf{F} \qquad P(101) = \mathsf{T}.$$

Sometimes a predicate is used so often as to merit a special symbol (mathematicians are, by necessity, inherently lazy when writing things down). There are many of these in set theory.

Definition 0.8 (Element)**.** The symbol $\in$, read 'is an element of', is a predicate with definition

$$\begin{aligned} \in &: (\text{element}, \text{Universe}) \to \{\mathsf{T}, \mathsf{F}\} \\ \in &: (e, U) \mapsto e \text{ is an element of } U \end{aligned}$$

(Note: The symbol ε, which will later denote arbitrarily small quantities, should *not* be used for set inclusion.)

Example 0.9. Over the universe of *even* numbers $E = \{2, 4, 6, \ldots\}$ we deduce

$$\begin{aligned} 2 \in E \iff \in (2, E) &\iff 2 \text{ is an element of } \{2, 4, 6, \ldots\} \\ &\iff \mathsf{T} \end{aligned}$$

and

$$\begin{aligned} 17 \in E \iff \in (17, E) &\iff 17 \text{ is an element of } \{2, 4, 6, \ldots\} \\ &\iff \mathsf{F} \end{aligned}$$

which can be abbreviated as $2 \in E$ and $17 \notin E$ (both evaluate to true).

Example 0.9 utilises a widely employed short form for invoking binary mappings (those mappings taking two inputs to one). Consider the addition function which takes two integers and maps them to their sum

$$\begin{aligned} + &: \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} \\ + &: (x, y) \mapsto x + y. \end{aligned}$$

It is universally understood that $2 + 5$ is a short form for $+(2, 5)$. When functions and mappings on sets are defined, keep it in the back of your mind that we are applying binary mappings in this way.

Definition 0.10 (Existence)**.** The logical statement '*there exists*' (alternatively '*there is*') is denoted by $\exists$. It is used to express that there is some element of the predicate's universe for which the predicate is true:

$$\exists x \in \{x_0, x_1, \ldots\}\,;\, P(x) \iff P(x_0) \text{ or } P(x_1) \text{ or } \cdots.$$

Example 0.11. Using the prime test predicate of Example 0.7:

$$\exists x P(x) = \mathsf{T}$$

when P's universe is $\mathbb{Z}$ (in fact $\mathbb{Z}$ contains *all* the primes). *But* when P's universe is $\{4, 6, 8, \ldots\}$

$$\exists x P(x) = \mathsf{F} \iff \neg \exists x P(x)$$

as no even number—excluding 2—is prime! (The later statement reads 'there is no $x \in \{4, 6, 8, \ldots\}$ for which x is a prime'.) To address this subtlety it is common to make the universe explicit,

$$\exists x \in \mathbb{Z};\, P(X).$$

Definition 0.12 (Every). The symbol $\forall$ denotes *'for all'* (alternatively *'for every'*, *'for each'*) and is called the *universal quantifier*. $\forall$ is used to make assertions 'universally' over an entire set:

$$\forall x \in \{x_0, x_1, \ldots\};\, P(x) \iff P(x_0) \text{ and } P(x_1) \text{ and } \cdots$$

Proposition 0.13. For any predicate $P : U \to \{\mathsf{T}, \mathsf{F}\}$

$$\neg \forall x \in U;\, P(x) \equiv \exists x \in U;\, \neg P(x).$$

In prose: $P(x)$ is *not* true for every $x \in U$ only when there is $x \in U$ for which $P(x)$ is false (and vice versa)

Example 0.14. Let $Q(x)$ given over $\mathbb{Z}$ be true only when x is divisible by two:

$$Q(x) \iff 2 \,\big|\, x \iff \exists y \in \mathbb{Z};\, 2y = x.$$

It is *not* true that

$$\forall x \in \mathbb{Z};\, Q(x)$$

because

$$\exists x \in \mathbb{Z};\, \neg Q(x).$$

That is to say, there is some $x \in \mathbb{Z}$ (7 for instance) which is *not* divisible by 2.

If we provide another predicate $R(x) \iff 2 \,\big|\, x - 1$ then

$$\forall x \in \mathbb{Z};\, Q(x) \text{ or } R(x) = \mathsf{T}$$

since it *is* true that every integer is either even or one more than an even (i.e. all integers can be expressed as $2x$ or $2x + 1$).

Our first axiom (a fundamental assumption that cannot be proved) states the condition for two sets, A and B, to be identical.

Axiom 1 (Axiom of Extensionality). If every element of A is also in B (and vice versa) then A and B are equal:

$$\left[\forall x;\, x \in A \iff x \in B\right] \overset{\text{def.}}{\iff} A = B.$$

The notion of set equality can be weakened,

Definition 0.15 (Subset). A is a subset of B when each element of A is also an element of B,

$$\left[\forall x;\, x \in A \implies x \in B\right] \overset{\text{def.}}{\iff} A \subseteq B$$

and weakened again,

Definition 0.16 (Proper/Strict subset).

$$A \subset B \overset{\text{def.}}{\iff} \left[[A \subseteq B] \text{ and } \neg\,[B \subseteq A]\right].$$

> Unfortunately there is some differences (mostly cross-culturally and cross-disciplinary) regarding the symbols used to distinguish subsets from proper subsets. Although we use $\subseteq$ and $\subset$, it is handy to know that other people/texts instead use $\subset$ and $\subsetneq$ to distinguish subset and proper subset (i.e. subset but not equal).

Example 0.17. For $A = \{1, 2, 3\}$, $B = \{3, 1, 2\}$ and $C = \{1, 2, 3, 4\}$

$$A \subseteq A, \qquad A \subseteq B, \qquad A \subseteq C,$$

and

$$\neg\,[A \subset A], \qquad \neg\,[A \subset B] \qquad [A \subset C].$$

(In logic everything written should be true. To express something false, state the negation as true.)

Proposition 0.18. $A \subseteq B$ and $B \subseteq A \implies A = B$.

PROOF. Simply utilizing the definitions we derive

$$\begin{aligned}
&A \subseteq B \text{ and } B \subseteq A \\
&\iff \left[\forall x;\, x \in A \implies x \in B\right] \text{ and } \left[\forall x;\, x \in B \implies x \in A\right] \quad \text{Defn. 0.15} \\
&\iff \forall x;\, x \in A \implies x \in B \text{ and } x \in B \implies x \in A \\
&\iff \forall x;\, x \in A \iff x \in B
\end{aligned}$$

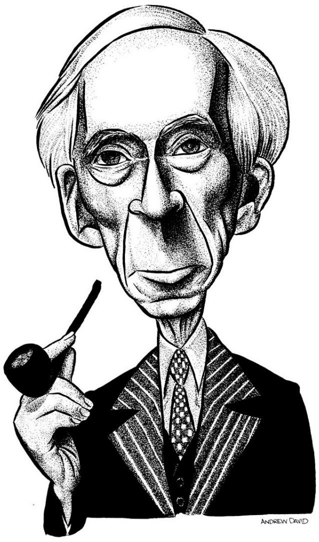


Figure 1: Bertrand Russel—winner of the Nobel prize in literature, religious and political philosopher, and renowned logician and mathematician—is depicted with his iconic pipe by the artist Andrew David.

$\iff A = B$ Axiom 1

(Only statements that are logical consequences of the line before can forgo justification.) ■

Axiom 2. There is a *unique* set (say, $\varnothing$) with no members,

$$\exists! \varnothing \; \forall x; \; x \notin \varnothing.$$

('!' is a short form for 'unique'.)

Definition 0.19 (Empty set). The empty set

$$\varnothing$$

is the unique set with no members.

Distinguish carefully between $\varnothing$ (the empty set) and φ/φ (the greek letter 'phi')!

Proposition 0.20. The empty set satisfies:

1. $\forall x; \; x \notin \varnothing$,
2. $\forall A; \; \varnothing \subseteq A$, and
3. $\forall A; \; A \subseteq \varnothing \implies A = \varnothing$

A natural question to raise is that of the *universal set*—the set containing everything (the complement of the empty set). We will prove there is no such set soon.

0.1.1 Class Abstraction

There are several ways to express a set. We could specify the members outright:

$$A = \{1, 2, 4, 9, 16, \ldots\};$$

use *set builder notation* or *class abstraction* by providing a predicate (with implicit universe):

$$\begin{aligned}
B &= \{x \text{ `such that' } x \text{ is a prime number}\} \\
&= \{x : x \text{ is a prime number}\} \\
&= \{2, 3, 5, 7, 11, \ldots\}, \\
C &= \{x : x \text{ is an english word and also a palindrome}\} \\
&= \{\text{a, dad, mom}, \ldots\}, \\
D &= \{x : x \text{ is a positive integer divisible by 3 and less than 10}\} \\
&= \{3, 6, 9\};
\end{aligned}$$

(notice the '...' of C are somewhat meaningless and that D is a finite set); or recursively.

1. $2 \in E$,
2. $a, b \in E \iff a \cdot b \in E$.

Exercise 0.1. What is the definition of E using set building notation?

0.1.2 Operations on Sets

We have already discussed the subset operation on sets and, expectedly, there are more.

Definition 0.21 (Power set). The power set of A, denoted $\mathcal{P}(A)$, is the set of all subsets of A.

$$\mathcal{P}(A) = \{B : B \subseteq A\}.$$

Example 0.22. If $A = \{1, 2, 3\}$ then the power set of A is

$$\mathcal{P}(A) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

> As $\varnothing \subset A$ for any set A, notice $\varnothing$ is automatically a member of *every* power set—including the power set of itself!

Exercise 0.2. What is $\mathcal{P}(\varnothing)$? In particular, is $\mathcal{P}(\varnothing) = \varnothing$?

> The notion of $\varnothing$ as distinct from $\{\varnothing\}$, is a subtle yet incredibly powerful idea. All objects of mathematics, at their very core, are sets of sets. For instance the natural numbers (as described by Italian mathematician Giuseppe Peano in 1890) are given in this way:
>
> $$\begin{aligned} 0 &= \varnothing \\ 1 &= 0 \cup \{0\} = \varnothing \cup \{\varnothing\} \\ 2 &= 1 \cup \{1\} = \varnothing \cup \{\varnothing\} \cup \{\varnothing \cup \{\varnothing\}\} \\ &\vdots \\ n+1 &= n \cup \{n\}\,. \end{aligned}$$

ANSWER. Placeholder. ◆

Some prefer to regard the power set from the perspective of Combinatorics (a branch of mathematics concerned with counting arrangements of discrete objects). Occupiers of this camp would characterize the power set as enumerating the ways one can include/exclude elements.

Example 0.23. Again let $A = \{1, 2, 3\}$ and consider this encoding of the subsets, where we will regard $\square$ as an element that has been removed.

$$\begin{aligned} \texttt{000} &= \{\square, \square, \square\} = \{\} = \varnothing \\ \texttt{001} &= \{\square, \square, 3\} = \{3\} \\ \texttt{010} &= \{\square, 2, \square\} = \{2\} \\ \texttt{011} &= \{\square, 2, 3\} = \{2, 3\} \\ \texttt{100} &= \{1, \square, \square\} = \{1\} \\ \texttt{101} &= \{1, \square, 3\} = \{1, 3\} \\ \texttt{110} &= \{1, 2, \square\} = \{1, 2\} \\ \texttt{111} &= \{1, 2, 3\} = \{1, 2, 3\}\,. \end{aligned}$$

Realizing the encoding is simply the 3-bit binary numbers (of which there are $2^3 = 8$) written in ascending order, and that there is a one-to-one correspondence between these binary encodings and the subsets they encode, we can easily write down an equation for the cardinality of a power set.

Proposition 0.24. The cardinality of A's power set is $2^{|A|}$,

$$|\mathcal{P}(A)| = 2^{|A|}.$$

(This is likely the motivation for the alternative notation 2^A for the power set of A.)

Example 0.25. The set $B = \{0, 1, \ldots, 299\}$ corresponds to a power set with cardinality 2^{300} (i.e. huge—for comparison, the number of atoms in the known universe is approximately 2^{265}).

Let us review the notions of Intersection, Set Difference, and Union. For the sake of brevity, but mostly because what follows is widely known, we eschew a lengthy discussion and assume that those who desire it will sample the literature[**?**].

Definition 0.26 (Set difference)**.** 'A without B' written $A \setminus B$ is given

$$A \setminus B = \{z : z \in A \text{ and } z \notin B\}\,.$$

Example 0.27. $\{1, 2, 3\} \setminus \{3, 4, 5\} = \{1, 2\}$.

Proposition 0.28. $z \in A \setminus B \iff \{z \in A \text{ and } z \notin B\}\,.$

Definition 0.29 (Intersection)**.** 'A intersect B' denoted $A \cap B$ is given by

$$A \cap B = \{z : z \in A \text{ and } z \in B\}\,.$$

Proposition 0.30. $z \in A \cap B \iff z \in A \text{ and } z \in B$.

Example 0.31. $\{1, 2, 3, 4, 5\} \cap \{2, 3, 4\} = \{2, 3, 4\}$.

Exercise 0.3. What is $C \cap \varnothing$?

Definition 0.32 (Disjoint)**.** We say that A and B are *disjoint* when

$$A \cap B = \varnothing.$$

Example 0.33. $\{1, 2\}$ and $\{3, 4\}$ are disjoint. $\{1, 2, 3\}$ and $\{3, 4\}$ are *not* disjoint as $\{1, 2, 3\} \cap \{3, 4\} = \{3\} \neq \varnothing$.

Definition 0.34 (Union)**.** A 'union' B denoted $A \cup B$ is given by

$$z \in A \cup B \iff (z \in A \text{ or } z \in B)\,.$$

Example 0.35. If $A = \{a, b, c\}$ and $B = \{b, c, d, e\}$ then $A \cup B = \{a, b, c, d, e\}$.

The various parts of the next theorem, which combines $\setminus$, $\cap$, and $\cup$, are the namesake of British mathematician Augustine De Morgan (1806-1871) who also formalized the 'Principal of Mathematical induction' (Theorem 0.46).

Figure 2: Indian born British logician Augustine De Morgan was teased as a child because a vision problem in his left eye prevented him from participating in sports. Later, a crater on the moon would be named in his honour.

Theorem 0.36 (De Morgan's Laws). For sets A, B, and C

1. $A \cap \{B \cup C\} = \{A \cap B\} \cup \{A \cap C\}$,
2. $A \cup \{B \cap C\} = \{A \cup B\} \cap \{A \cup C\}$,
3. $A \setminus \{B \cap C\} = \{A \setminus B\} \cap \{A \setminus C\}$, and
4. $A \setminus \{B \cap C\} = \{A \setminus B\} \cap \{A \setminus C\}$.

0.1.3 Set Constructs

Sets define ordered pairs.

Definition 0.37 (Ordered Pair). The *ordered pair* 'x followed by y' is denoted (x, y) and satisfies

$$(x, y) = \{\{x\}, \{x, y\}\}$$

Proposition 0.38.

$$(x, y) = (A, B) \iff x = A \text{ and } y = B$$

Example 0.39. $(2, 3)$ and $(3, 2)$ are (both) ordered pairs. These two ordered pairs are distinct despite despite having identical elements.

In general, an ordered n-tuple is written

$$(x_0, \ldots, x_{n-1})$$

and is the natural extension of Definition 0.37. (As a matter of convention we call say a: 2-tuple is a 'tuple'; 3-tuple is a 'triple'; 4-tuple is a 'quadruple'; and so on.)

The cartesian product quickly builds sets of ordered pairs:

Definition 0.40 (Cartesian product)**.** The Cartesian product of sets A and B is denoted $A \times B$ and given by

$$A \times B = \{(a, b) \ : \ a \in A \text{ and } b \in B\} .$$

Example 0.41.

$$\{2, 3\} \times \{x, y, z\} = \{(2, x), (2, y), (2, z), (3, x), (3, y), (3, z)\}$$

Notation. When $A = B$ in Definition 0.40 we may write A^2 for $A \times A$.

In contrast to the definition of a set, a *multiset* is a set where duplicates *are* counted (ordering is still ignored).

Definition 0.42 (Multiset)**.** A *multiset* is a set where *duplicate elements are counted*.

Example 0.43. $A = \{1, 2, 2, 3, 3, 3\}$ and $B = \{2, 1, 3, 3, 3, 2\}$ are multisets satisfying

$$A = B$$

and $|A| = |B| = 6$.

Unfortunately we have no way of distinguishing sets from multisets in writing (both use $\{\}$). We assume that sets omit duplicates and sets with duplicates present are mutisets.

Definition 0.44 (Sequence)**.** A *sequence* is an *ordered*-multiset. Sequences are denoted

$$S = (x_0, x_1, \dots, x_{n-1}) \tag{1}$$

and have this short form: $S_i = x_i$ (like array indexing). The *length* of a sequence is the cardinality of the sequence when viewed as a multiset.

Two sequences, A and B are equal when $\forall i \ A_i = B_i$.

Infinite sequences are given explicitly, recursively, or using a type of class abstraction called the 'closed form'.

Example 0.45 (Fibonacci sequence)**.** The *Fibonacci sequence* is given/generated by

Explicitly $F = (0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots)$
(explicit is a misnomer as '. . .' means the reader is left to determine the pattern.)

Recursively $F_{n-2} = F_n - F_{n-1}$ with $F_0 = 0$, $F_1 = 1$.

Closed Form Let $\varphi = \frac{1+\sqrt{5}}{2}$ and $\psi = \frac{1-\sqrt{5}}{2}$

$$F_n = \frac{\varphi^n - \psi^n}{\varphi - \psi} = \frac{\varphi^n - \psi^n}{\sqrt{5}}. \tag{2}$$

> $\varphi = \frac{1+\sqrt{5}}{2}$ or the *golden ratio* is pervasive in industrial and everyday design. For instance the pantheon and any credit card share the same relative dimensions because of the golden ratio.

0.2 *Proof Methods*

Much to the dismay of students, there is no 'recipe' for doing proofs (or solving problems in general). However, there are some established strategies and tools that often do the trick. We review them now.

0.2.1 *Direct Proof*

Let us substitute a formal description *of* a direct proof with a story *about* a direct proof.

The misbehaving (and still unrealized genius of number theory) Gauss (1777-1855), was exiled to the corner of his classroom and told not to return until he had calculated the sum of the first hundred numbers.

To the astonishment of the teacher Gauss returned with the correct answer in a matter of minutes—the pupil had deduced what the teacher could not:

$$\begin{aligned} &1 + 2 + 3 + \cdots + 50 + 51 + \cdots + 98 + 99 + 100 \\ &= (1 + 100) + (2 + 99) + (3 + 98) + \cdots + (50 + 51) \\ &= (101) + (101) + (101) + \cdots + (101) \\ &= \left(\frac{100}{2}\right) \cdot (100 + 1) \\ &= 5\,050. \end{aligned}$$

Figure 3: Gauss as pictured on the German 10-Deutsche Mark banknote (1993; discontinued). The Normal, or 'Gaussian' distribution is depicted to his left.

0.2.2 *The Principle of Mathematical Induction*

The *principle of mathematical induction* (or 'PMI' or just 'induction'), states simply that a proposition $P : m \to \{\mathsf{T}, \mathsf{F}\}$ is true for all $m \in \mathbb{N}$ if

1. $P(m) \implies P(m+1)$ for any $m \in \mathbb{N}$; and
2. $P(0)$.

(It is implicit that $P(0)$ means $P(0) \equiv \mathsf{T}$.)

To elaborate, the first point is a *weakening* of the intended conclusion and is equivalent to validating the following chain of implications:

$$P(0) \implies P(1) \implies \cdots \implies P(n) \implies P(n+1) \implies \cdots$$

Provided that $P(0) \equiv \mathsf{T}$ (the second point) *each* proposition in the chain is shown true and thus it is proved that $\forall n P(n)$.

Remark. *How* one proves $\forall n; P(n) \implies P(n+1)$ is often the source of confusion. The direct way of demonstrating $\varphi \implies \psi$ is to assume φ and show ψ is a consequence. The case of $\varphi \equiv \mathsf{F}$ is not ignored, but is considered too trivial to write as $\mathsf{F} \implies \varphi \equiv \mathsf{T}$ (such statements are called 'vacuous').

As is the case, induction proofs contain the bizarre statement "assume for any $n \in \mathbb{N}$ that $P(n) \equiv \mathsf{T}$" which is seemingly what requires proof.

However, there is gulf of difference between

$$(\forall n P(n)) \implies P(n+1)$$

and

$$\forall n\,(P(n) \implies P(n+1))\,. \tag{3}$$

We are *not* assuming $\forall n P(n)$, rather we are assuming, for any n, the premise of (3). In the same spirit, assuming 'there is' some n for which $P(n) \equiv \mathsf{T}$ is also wrong. Demonstrating $\exists n;\, P(n) \implies P(n+1)$ is insufficient.

Formally, the PMI is given this way.

Theorem 0.46 (The Principle of Mathematical Induction)**.** For *every* predicate $P : \mathbb{N} \to \{\mathsf{T}, \mathsf{F}\}$

$$\{P(0) \text{ and } \forall m \in \mathbb{N}\,[P(m) \implies P(m+1)]\} \implies \{\forall m \in \mathbb{N}\,[P(m)]\}$$

(we use superfluous bracketing to express something more meaningful).

To apply this theorem, let us prove 'Gauss' formula' using induction rather than deduction.

Proposition 0.47. For any $n \in \mathbb{N}$

$$0 + 1 + 2 + \cdots n = \sum_{i=0}^{n} i = \frac{n \cdot (n+1)}{2}.$$

PROOF. Proceeding with induction it is clear that $n = 0$ is satisfied as

$$\left[\sum_{i=0}^{n} i\right]_{n=0} = \sum_{i=0}^{0} i = 0 \tag{4}$$

and

$$\left[\frac{n \cdot (n+1)}{2}\right]_{n=0} = \frac{0 \cdot (0+1)}{2} = 0.$$

(the 'Base case').

Assume for arbitrary $m \in \mathbb{N}$ the validity of

$$0 + 1 + 2 \cdots + m = \sum_{i=0}^{m} i = \frac{m \cdot (m+1)}{2}$$

(the 'Induction hypothesis').

Using this deduce

$$\begin{aligned} 0+1+2\cdots+m+(m+1) &\overset{\text{IH}}{=} \frac{m\cdot(m+1)}{2}+(m+1) \\ &\text{(by Induction Hypothesis)} \\ &= \frac{m\cdot(m+1)+2\,(m+1)}{2} \\ &= \frac{m^2+3m+2}{2} \\ &= \frac{(m+1)(m+2)}{2} \end{aligned}$$

which shows (4) holds for $n = m + 1$ provided it holds for $n = m$.

By the PMI (4) is valid $\forall n \in \mathbb{N}$. ■

0.2.3 *Contradiction*

Contradiction is a proof technique where, in order to show some predicate P true, we assume $\neg P$ and deduce F. (That is, we show invalid P has an absurd consequence).

Logically, proof by contradiction can be expressed as

Theorem 0.48 (Proof by Contradiction)**.** For any predicate P

$$(\neg P \implies \mathsf{F}) \implies P.$$

We give two standard examples (in ascending difficulty) for which contradiction is applicable.

Proposition 0.49. $\sqrt{2}$ can not be expressed as a fraction (i.e. $\sqrt{2}$ is an irrational number)

(Note: Read $a \mid b$ as 'a divides b' which means $\exists c : ac = b$. E.g. $2 \mid 6$.)

PROOF. Towards a contradiction (TAC for short), suppose $\sqrt{2}$ *can* be expressed as

$$\sqrt{2} = \frac{a}{b} \tag{5}$$

with $a, b \in \mathbb{N}$. Assume further that $\frac{a}{b}$ is a *reduced fraction* so that $\gcd(a, b) = 1$.

Squaring (5) yields

$$2 = \frac{a^2}{b^2} \implies 2b^2 = a^2.$$

Trivially $2 \mid 2b^2$ and so $2 \mid a^2$.

But 2 cannot be decomposed (it is prime) so it must be that $2 \mid a$ and thus $4 \mid a^2$. Similarly (applying this argument in the reverse direction) $4 \mid 2b^2 \implies 2 \mid b^2$ and thus $2 \mid b$.

However, if *both* a and b are divisible by 2 it must be the case that $\frac{a}{b}$ is *not* reduced, i.e. $\gcd(a, b) = 2 \neq 1$. ↯[1] ■

Proposition 0.50. There are infinitely many prime numbers.

PROOF. TAC suppose there are *finitely* many prime numbers $\mathbb{P} = \{p_0, p_1, p_2, \dots, p_\ell\}$ and consider $n \in \mathbb{N}$ given by

$$n = p_0 \cdot p_1 \cdot p_2 \cdots p_\ell + 1.$$

As every integer has a prime divisor (this is the fundamental theorem of algebra) there must be some $p_i \in \mathbb{P} : p_i \mid n$. Clearly $p_i \mid p_0 \cdot p_1 \cdots p_i \cdots p_\ell$ so it follows

$$p_i \mid (n - p_0 \cdot p_1 \cdots p_n)$$

(It is readily shown that $a \mid b$ and $a \mid c \implies a \mid (b - c)$.)

Consequently, $p_i \mid 1 \implies p_i = 1$ (by definition 1 is *not* a prime) and thus $p_i \notin \mathbb{P}$. ↯ ■

Theorem 0.51 (Russel's Paradox). There is no *Universal Set*. That is, there is no set containing all sets:

$$\neg \exists A \forall x \, (x \in A).$$

PROOF. TAC suppose there is *universal set* U containing all sets

$$\exists U \, \forall x \, (x \in U).$$

Notice $\forall x \, (x \notin x) \equiv \mathsf{T}$ and thus we can write U with *class abstraction* as

$$U = \{x : x \notin x\}.$$

This derives contradiction because

$$U \in U \iff U \notin U. \text{ ↯}$$

[1]There are many symbols for contradiction, among them: $\Rightarrow\Leftarrow$, ↯, and $\nLeftrightarrow$. Feel free to use whichever one you like!

Thus there is no universal set. ■

0.2.4 The Pigeonhole Principle

The *pigeonhole principle* is the mathematical formalization of the statement:

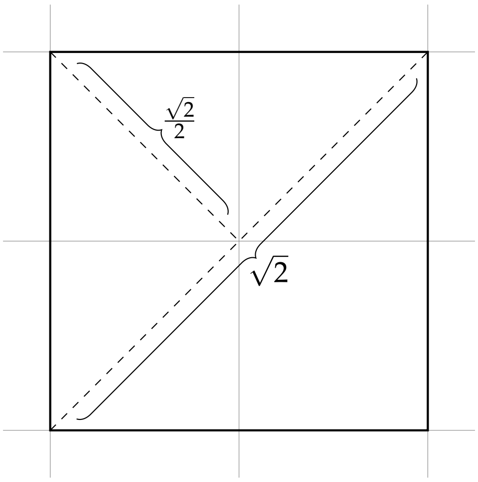
> If you put $n+1$ pigeons into n holes then there is (at least) one hole with two pigeons.[2]

Recall the Euclidean distance between two points in the plane:

$$\left|\overline{(x_1,y_1)(x_2,y_2)}\right| = \sqrt{(x_1-x_2)^2+(y_1-y_2)^2}.$$

Proposition 0.52. If 5 points are drawn within the interior (i.e. *not* on the edge) of unit square then there are two points that have Euclidean distance $< \frac{1}{\sqrt{2}}$.

PROOF. The diagonal of a unit square has length $\sqrt{2}$ (Pythagoras' theorem) and a subsquare $\frac{1}{4}$ the area has diagonal length $\frac{\sqrt{2}}{2}$,



Therefore, any points within the same subsquare can be at *most* $\frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}}$ units apart.

By PHP, if five points are drawn within the interior of a square, then two points must be in the same subsquare (there are only four such subsquares). Thus there are two points with Euclidean distance less than $\frac{1}{\sqrt{2}}$. ■

[2] The motivation for placing pigeons into holes eludes the Author (who is also bothered by the notion of cramming two pigeons into a hole meant for one).

0.3 Number Systems

There are standard sets, with fixed names, that mostly all math students have seen. Among them are the

1. *natural numbers*: $\mathbb{N} = \{0, 1, 2, \ldots\}$,
2. *integers*: $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$,[3]
3. *rational numbers*: $\mathbb{Q} = \left\{\frac{a}{b} : a, b \in \mathbb{Z} \text{ and } b \neq 0\right\}$,
4. *irrational numbers*: $\overline{\mathbb{Q}} = \{$ numbers, like π, which cannot be represented by fractions$\}$, and
5. *real numbers*: $\mathbb{R} = \mathbb{Q} \cup \overline{\mathbb{Q}}$.

They are related by

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$$

as (for example) every natural number is indeed and integer, rational, and real number, but (for example) not every real number is rational, integer, or natural.

But what operations do each number system admit? The simplest operation, addition, is defined for the naturals because the sum of any two is a natural number:

$$a, b \in \mathbb{N} \implies a + b \in \mathbb{N}$$

For this reason we say the naturals are *closed under addition*. Moreover, this addition operation is *commutative* and *associative*, that is, for $a, b, c \in \mathbb{N}$ (resp.)

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

There is also an *additive identity*, a number b satisfying $a + b = b + a = a$ which we know to be zero: $a + 0 = 0 + a = a$.

Natural numbers, however, do not have *additive inverses*. For any $a \in \mathbb{N}$, there is no $b \in \mathbb{N}$ satisfying $a + b = b + a = 0$. Inverse addition is more broadly known as *subtraction* (so perhaps $-a$ is a more appropriate notation than b) and it is easy to see the naturals are not closed over this operation. Consider that 5 and 7 are natural numbers but $5 - 7 = -2$

[3] $\mathbb{Z}$ because the German word for 'number' is 'Zahlen'

is not. The integers, which include $-1, -2, \ldots$, *do* have unique additive inverses. We can express this logically as:

$$\forall a \in \mathbb{Z};\ \exists! - a \in \mathbb{Z} \ :\ a + (-a) = 0$$

or in prose by: for any $a \in \mathbb{Z}$ there is a unique $b \in \mathbb{Z}$ such that $a + (-a) = 0$.

Definition 0.53 (Group)**.** A set G along with an addition operation $+$ is called a group, denoted $(G, +)$, when

$$\begin{aligned}
&a, b \in G \implies a + b \in G, && \text{closed under addition,}\\
&\forall a, b, c \in G;\ (a+b)+c = a+(b+c), && \text{associativity of addition,}\\
&\exists 0 \in G \ :\ \forall a \in G;\ a + 0 = 0 + a = a, && \text{additive identity,}\\
&\forall a \in G;\ \exists! b \in G \ :\ a + b = 0 && \text{additive inverse.}
\end{aligned}$$

Definition 0.54 (Albelian group)**.** The group $(G, +)$ is an *abelian group* when its addition operation is commutative. That is,

$$\forall a, b \in G;\ a + b = b + a \qquad \text{commutativity of addition.}$$

Example 0.55. $(\mathbb{Z}, +)$ is a group but $(\mathbb{N}, +)$ is not.

If we imbue $(G, +)$ with a multiplication operation $\cdot$ (read as "dot") that satisfies some properties the group becomes a *ring*.[4] Notice $\mathbb{N}$ and $\mathbb{Z}$ are closed over $\cdot$ (which is associative and commutative) and even have the *multiplicative identity* '1'. There is only one more property required which combines addition and multiplication, namely, *distributively*:

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

Definition 0.56 (Ring)**.** A group $(G, +)$ along with a multiplication operation $\cdot$ is a *ring*, denoted $(G, +, \cdot)$, when

$$\begin{aligned}
&a, b \in G \implies a \cdot b \in G, && \text{closed over multiplication,}\\
&\forall a, b, c \in G;\ (a \cdot b) \cdot c = a \cdot (b \cdot c), && \text{associativity of multiplication,}\\
&\exists 1 \neq 0 \in G \ :\ \forall a \in G;\ a \cdot 1 = 1 \cdot a = a, && \text{multiplicative identity,}\\
&\forall a, b, c \in G;\ a \cdot (b + c) = a \cdot b + a \cdot c && \text{distributivity.}
\end{aligned}$$

[4] We use $\cdot$ instead of $\times$ for multiplication because the later will eventually be used for the "cross-product." Moreover, we sometimes forgo writing $\cdot$ and say $a \cdot b = ab = (a)(b)$.

Moreover $(G, +, \cdot)$ is a *commutative ring* when we also have

$$\forall a, b \in G;\ a \cdot b = b \cdot a \qquad \text{commutativity of multiplication.}$$

Example 0.57. $(\mathbb{Z}, +, \cdot)$ is a commutative ring.

Elements from $\mathbb{N}$ and $\mathbb{Z}$, however, do *not* have multiplicative inverses. Consider that the multiplicative inverse of 2 must be a unique number a such that $2a = 1$. That is, the multiplicative inverse of 2 is $a^{-1} = \frac{1}{2}$. (Read a^{-1} as "a inverse".) What is required for inversion of integers are fractions because any integer a inverts to $\frac{1}{a}$. Rings $(G, +, \cdot)$ that also have multiplicative inverses are called *fields*.

Definition 0.58 (Field). A ring $(G, +, \cdot)$ is a *field* when every nonzero element from G has a *multiplicative* inverse. That is,

$$\forall a \neq 0 \in G;\ \exists a^{-1} \in G : a \cdot a^{-1} = 1 \qquad \text{multiplicative inverse,}$$

where $1 \in G$ is the multiplicative identity.

Example 0.59. $(\mathbb{Q}, +, \cdot)$ is a field.

0.4 What is to come

In this course we will, in part, investigate number systems which fall into these categories and an additional one called a *vector space*. In particular *complex numbers* and *matrices*, along with corresponding operations, form a (resp.) a field and a non-commutative ring.

1 Complex Numbers

"Anyone who is capable of getting themselves made President should on no account be allowed to do the job."
– Douglas Adams, *The Hitchhiker's Guide to the Galaxy*

1.1 Number Systems

The need for more numbers is apparent simply by considering roots of polynomials. For instance, consider the quadratic equation

$$ax^2 + bx + c = 0$$

with its roots

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{1.1}$$

for $a \neq 0$. The term

$$\Delta := b^2 - 4ac$$

is called the *discriminant*, because it *discriminates* between three different kinds of behaviour:

$\Delta = 0$ implies

$$ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2$$

and so there is *only* a *single repeated root* $\frac{-b}{2a}$. (See Figure 1.1.)

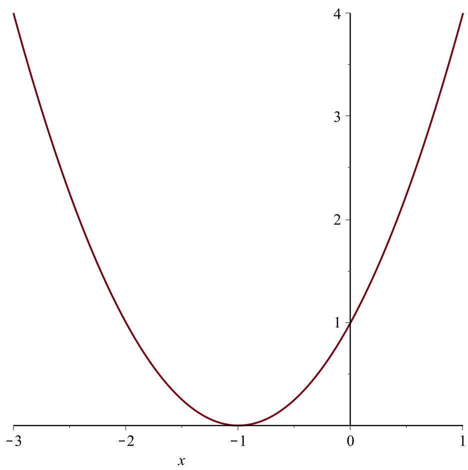


Figure 1.1: $x^2 + 2x + 1$ with $\Delta = 0$ and a single repeated root at $(-1, 0)$.

$\Delta > 0$ implies

$$ax^2 + bx + c = a\left(x + \frac{b - \sqrt{\Delta}}{2a}\right)\left(x + \frac{b + \sqrt{\Delta}}{2a}\right)$$

which correspond to the *two distinct real roots* given by (1.1). (See Figure 1.2.)

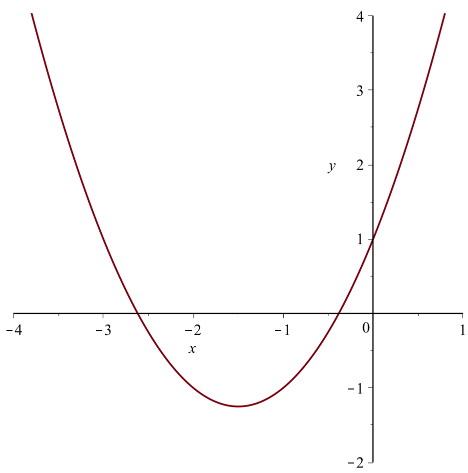


Figure 1.2: $x^2 + 3x + 1$ with $\Delta = 5$ and irrational roots $(-\frac{3}{2} \pm \frac{1}{2}\sqrt{5}, 0)$.

$\Delta < 0$ implies that there are *no real roots*. (See Figure 1.3.)

Here we see an increasing need for more numbers. Single repeated roots, by their definition, will always be fractions (i.e. rational numbers), whereas when the discriminant is strictly greater than zero we may require irrational numbers (like $\sqrt{2}$) which fractions cannot express. In the last case, when $\Delta < 0$, we would have roots of negative numbers — seemingly impossible as no number can square to a negative.

This demonstrates that we clearly need more numbers, which is not

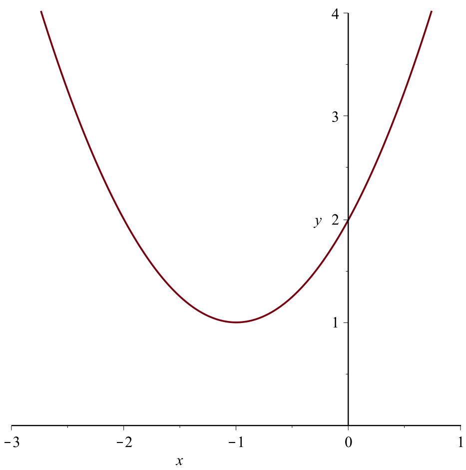


Figure 1.3: $x^2 + 2x + 2$ with $\Delta = -4$ and no real roots.

unprecedented: $x - 1 = 0$ needed negative numbers and $x^2 - 2 = 0$ needed irrationals. (Considering how perplexing it must have been to not have a number which could describe the length of the diagonal of a unit square!)

1.2 Complex Numbers

Definition 1.1. Let the *imaginary number* $\sqrt{-1}$ be denoted by:

$$\mathrm{i} := \sqrt{-1}.$$

Now let us combine these new imaginary numbers with the real numbers to make a new set of numbers called the *complex numbers* and denote them by $\mathbb{C}$.

Definition 1.2 (Complex Number). Let the set of complex numbers be given by

$$\mathbb{C} = \{x + \mathrm{i}\,y \,:\, x, y \in \mathbb{R}\}.$$

Further let, when $x, y \in \mathbb{R}$ and $z = x + \mathrm{i}\,y$, the *real part* of z be given by

$$\mathrm{Re}(z) = x$$

and the *imaginary part* of z be given by

$$\mathrm{Im}(z) = y.$$

1.2.1 Complex Arithmetic

Now we need to define arithmetic (i.e. $+$ and $\cdot$) on $\mathbb{C}$ in such a way to make $(\mathbb{C}, +, \cdot)$ a field. Complex multiplication needs to be *commutative* and *distributive* so our intuition tells us that

$$\begin{aligned}
&(x_1 + \mathrm{i}\, y_1) \cdot (x_2 + \mathrm{i}\, y_2) \\
&\quad = (x_1 + \mathrm{i}\, y_1) \cdot x_2 + (x_1 + \mathrm{i}\, y_1) \cdot \mathrm{i}\, y_2 && \text{Distributivity} \\
&\quad = x_2 \cdot (x_1 + \mathrm{i}\, y_1) + \mathrm{i}\, y_2 \cdot (x_1 + \mathrm{i}\, y_1) && \text{Commutativity} \\
&\quad = x_1 x_2 + \mathrm{i}\, x_2 y_1 + \mathrm{i}\, x_1 y_2 - y_1 y_2 && \text{Commutativity} \\
&\quad = x_1 x_2 + \mathrm{i}\, x_2 y_1 + \mathrm{i}\, x_1 y_2 + (-1) y_1 y_2 && \text{Definition} \\
&\quad = x_1 x_2 + \mathrm{i}(x_2 y_1 + x_1 y_2) - y_1 y_2 && \text{Distributivity} \\
&\quad = (x_1 x_2 - y_1 y_2) + \mathrm{i}(x_1 y_2 + x_2 y_1) && \text{Commutativity.}
\end{aligned}$$

(This deduction is deliberately pedantic — in general this level of detail is not necessary.) We define multiplication to be consistent with the above.

Definition 1.3 (Complex Arithmetic)**.** For $z_1 := x_1 + \mathrm{i}\, y_1$ and $z_2 := x_2 + \mathrm{i}\, y_2$ complex numbers let

$$\begin{aligned}
z_1 + z_2 &:= (x_1 + x_2) + i(y_1 + y_2) && \text{Addition,} \\
z_1 \cdot z_2 &:= (x_1 x_2 - y_1 y_2) + \mathrm{i}(x_1 y_2 + x_2 y_1) && \text{Multiplication,}
\end{aligned}$$

where the sums and products among the real and complex parts are done in $\mathbb{R}$.

Proposition 1.4. $(\mathbb{C}, +, \cdot)$ is a commutative ring.

PROOF. Let $x_1, x_2, y_1, y_2 \in \mathbb{R}$ and $z_1 := x_1 + \mathrm{i}\, y_1$ and $z_2 := x_2 + \mathrm{i}\, y_2$.

Clearly, when $z_1, z_2 \in \mathbb{C}$ then $z_1 + z_2 \in \mathbb{C}$ and $z_1 \cdot z_2 \in \mathbb{C}$ as we need only appeal to the definition of $+$ and $\cdot$. Also $1 \in \mathbb{C}$ and $0 \in \mathbb{C}$ satisfy

$$\begin{aligned}
1 \cdot z_1 &= (1 + \mathrm{i}\, 0) \cdot (x_1 + \mathrm{i}\, y_1) && \text{Definition of complex number} \\
&= (1 \cdot x_1 - 0 \cdot y_1) + \mathrm{i}\,(0 \cdot x_1 + 1 \cdot y_1) && \text{Definition of multiplication} \\
&= x_1 + \mathrm{i}\, y_1 \\
&= z_1
\end{aligned}$$

and (for the same reason) $0 \cdot z_1 = 0$.

It only remains to show that $\mathbb{C}$ has the distributivity property.

Exercise 1.1. Demonstrate with sufficient rigour (i.e. be sufficiently pedantic) that $(\mathbb{C}, +, \cdot)$ has the distributivity property. Namely, for any $z_1, z_1,$

$z_3 \in \mathbb{C}$, that

$$z_1 \cdot (z_2 + z_3) = z_1 \cdot z_2 + z_1 + z_3.$$

■

1.2.2 Complex Conjugation and Inversion

In order for $(\mathbb{C}, +, \cdot)$ to be a field we must demonstrate that every nonzero $z \in \mathbb{C}$ has a multiplicative inverse (the *additive* inverse of z_1 is just $-z_1$).

The inverse of $z = x + iy$, say z^{-1}, must satisfy $z \cdot z^{-1} = 1$. Let us write this as $(x + iy) \cdot \frac{1}{(x+iy)} = 1$ with the understanding that $z^{-1} = \frac{1}{(x+iy)}$. Better yet, let us write,

$$\frac{(x+iy)}{(x+iy)} = 1. \tag{1.2}$$

Notice now that if we multiply (1.2) by $\frac{x-iy}{x-iy}$ (which is just one) we get

$$\frac{(x+iy)}{(x+iy)}\frac{(x-iy)}{(x-iy)} = 1 \implies \frac{1}{(x+iy)}\frac{(x-iy)}{(x-iy)} = \frac{1}{(x+iy)}$$

and, because $(x + iy)(x - iy) = x^2 + y^2$, can conclude

$$z^{-1} = (x - iy) \cdot \frac{1}{x^2 + y^2}.$$

Observe—crucially—that $x^2 + y^2 \in \mathbb{R}$ and therefore we know how to invert it.

The fact $(x + iy)(x - iy) = x^2 + y^2$ is quite important. So much so that we give $(x - iy)$ a special name.

Definition 1.5 (Complex conjugate). Let $z = x + iy \in \mathbb{C}$. The *complex conjugate* of z, denoted $\overline{z}$, is given by

$$\overline{z} = \overline{(x+iy)} = (x - iy).$$

Proposition 1.6. Let $z = x + iy$ and $w = s + it$ be complex numbers, then the following hold:

1. $z \cdot \overline{z} = x^2 + y^2$,
2. $z + \overline{z} = 2\mathrm{Re}(z)$,
3. $\overline{z + w} = \overline{z} + \overline{w}$,
4. $w \neq 0 \implies \overline{z/w} = \overline{z}/\overline{w}$, and
5. $\overline{z} = z \iff z \in \mathbb{R}$.

Exercise 1.2. Prove Proposition 1.6.

Definition 1.5 enables us to write a more compact expression for the inversion of a complex number.

Definition 1.7 (Complex inversion). Let $z = x + iy \in \mathbb{C}$ and z^{-1} be the inverse of z, then

$$z^{-1} := \frac{\overline{z}}{z \cdot \overline{z}} = \frac{\overline{z}}{x^2 + y^2}.$$

Exercise 1.3. Take for granted that 0 has no inverse in $\mathbb{R}$. Demonstrate why this implies 0 has no inverse in $\mathbb{C}$.

Finally, we can conclude that $(\mathbb{C}, +, \cdot)$ is a field.

Proposition 1.8. $(\mathbb{C}, +, \cdot)$ is a field.

PROOF. We have by Proposition 1.4 that $(\mathbb{C}, +, \cdot)$ is a commutative ring. It suffices to show that any nonzero complex number has an inverse. So, let $z = x + iy$ be an arbitrary non-zero complex number and notice

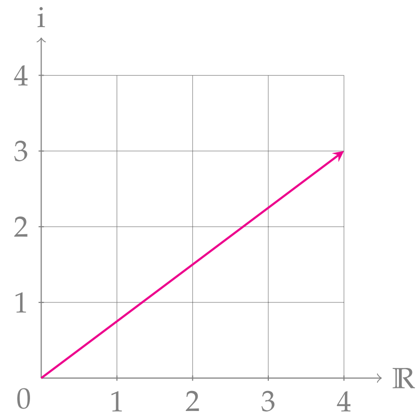
$$\begin{aligned} z \cdot z^{-1} &= \frac{z \cdot \overline{z}}{z \cdot \overline{z}} && \text{Definition} \\ &= \frac{x^2 + y^2}{x^2 + y^2} && \text{Proposition 1.6} \\ &= 1 && \text{Property of } \mathbb{R}. \end{aligned}$$

■

1.3 Cartesian Complex Numbers

Complex number can be visualized by drawing them on the *complex plane*, sometimes called the *Argand plane*, by the mapping

$$x + iy \mapsto (x, y).$$



$4 + 3i$ drawn on the complex plane.

This is called the *cartesian form* of a complex number.

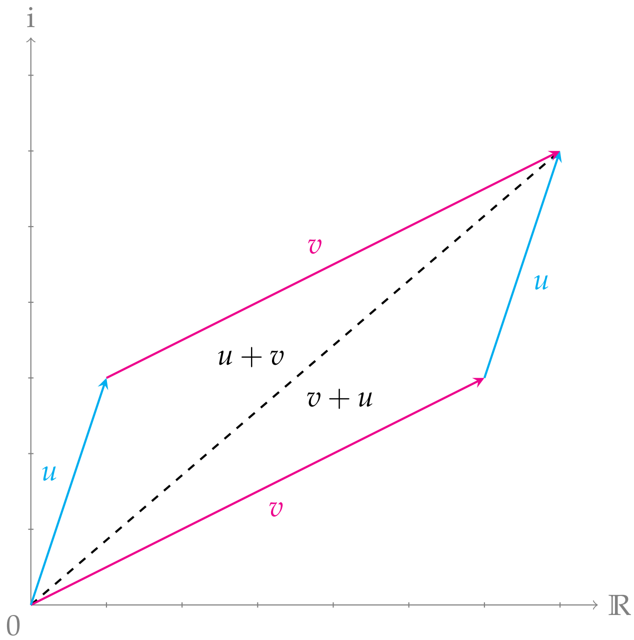
> Jean-Robert Argand (July 18, 1768 – August 13, 1822) was an amateur mathematician. In 1806, while managing a bookstore in Paris, he published the idea of geometrical interpretation of complex numbers known as the Argand diagram and is known for the first rigorous proof of the Fundamental Theorem of Algebra. [Wikipedia]

Definition 1.9 (Cartesian form of a Complex Number)**.** The *cartesian form* of a complex number $z \in \mathbb{C}$ is written with a real part $x \in \mathbb{R}$ and imaginary part $y \in \mathbb{R}$ as

$$z = x + iy.$$

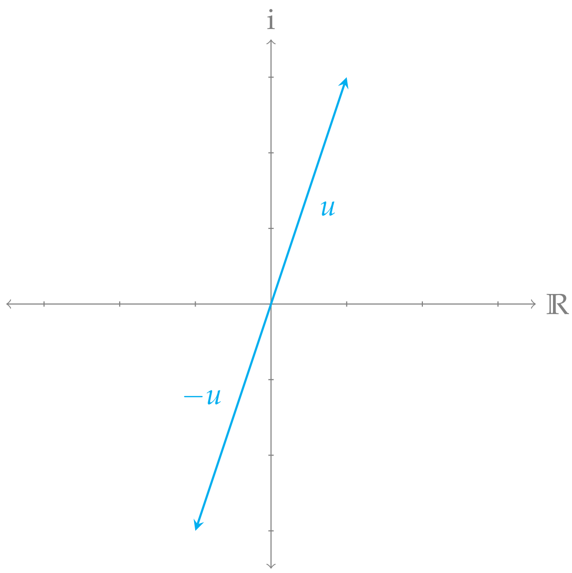
The geometric interpretation of complex numbers means that we can *illustrate* all the operations we defined in the last section to see how they act.

Example 1.10. Complex addition is indeed commutative.



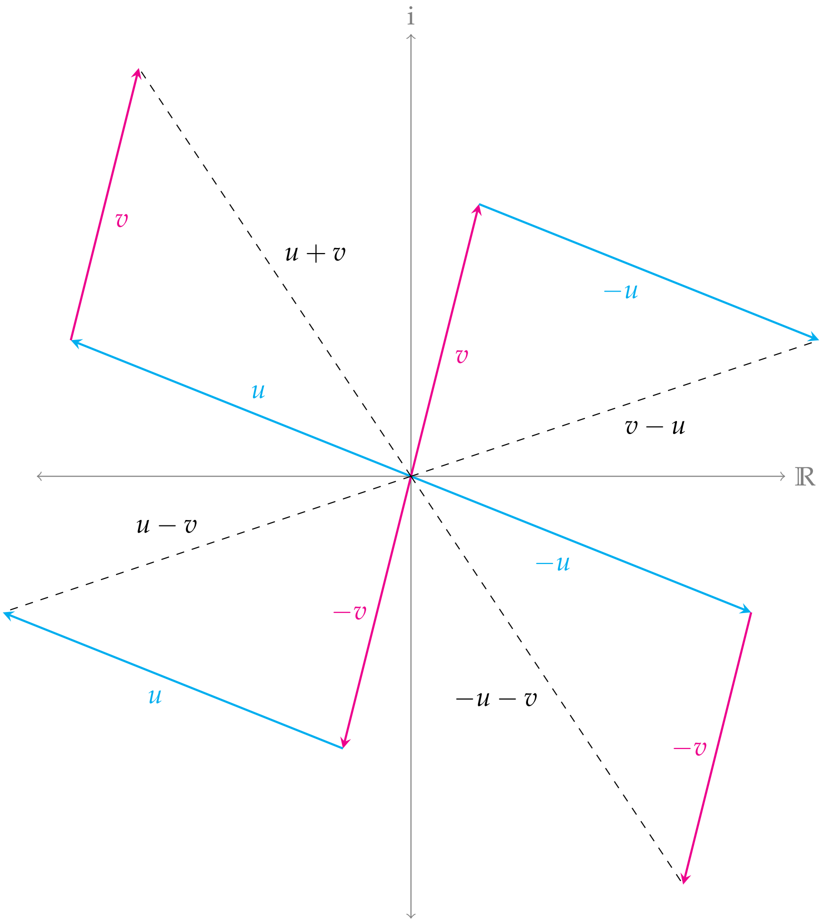
Commutativity of complex addition.

Example 1.11. Unlike addition, subtraction is *not* commutative. To understand why, first notice that the negation of a complex number, geometrically, means we reflect in both the $\mathbb{C}$ and $\mathbb{R}$ axis.



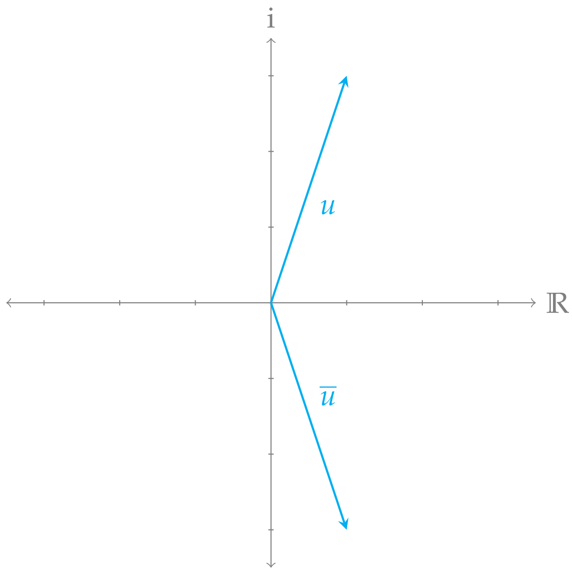
The complex number u and its negation $-u$.

So, we can subtract v from u by adding $-v$ to u.



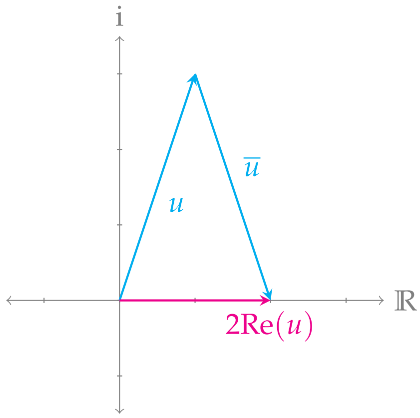
The different ways u and v can be subtracted and added.

Example 1.12. Since conjugation merely means negating the imaginary part of a complex number, geometrically this means we just reflect in the horizontal axis.



The complex number u and its conjugate $\overline{u}$.

Which means we can verify that $u + \overline{u} = 2\text{Re}(u)$



and that $\overline{u+v} = \overline{u} + \overline{v}$ (see Figure 1.4).

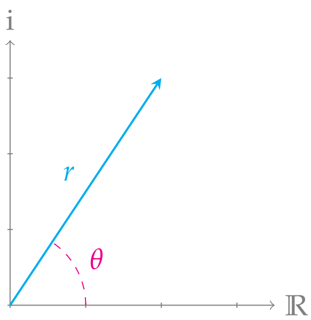
But what is the geometric interpretation of multiplication and inversion (i.e. division)? It turns out the way we are plotting numbers, while helpful for visualising summations and conjugates, does not provide much insight for multiplication (see Figure 1.5). Switching to Polar coordinates will solve this mystery.

1.4 *Polar Complex Numbers*

Instead of using cartesian co-ordinates, we can identify a complex number in the complex plane using length and angle. That is, by

$$(r, \theta)\,.$$

(Note, as a matter of convention we assume angles are always given in radians.)



The complex number (r, θ).

Definition 1.13 (Polar Form of Complex Number)**.** The *polar form* of a complex number is written using a length $r \in \mathbb{R}$ and angle $\theta \in \mathbb{R}$ as

$$z = r(\cos(\theta) + \mathrm{i}\sin(\theta)).$$

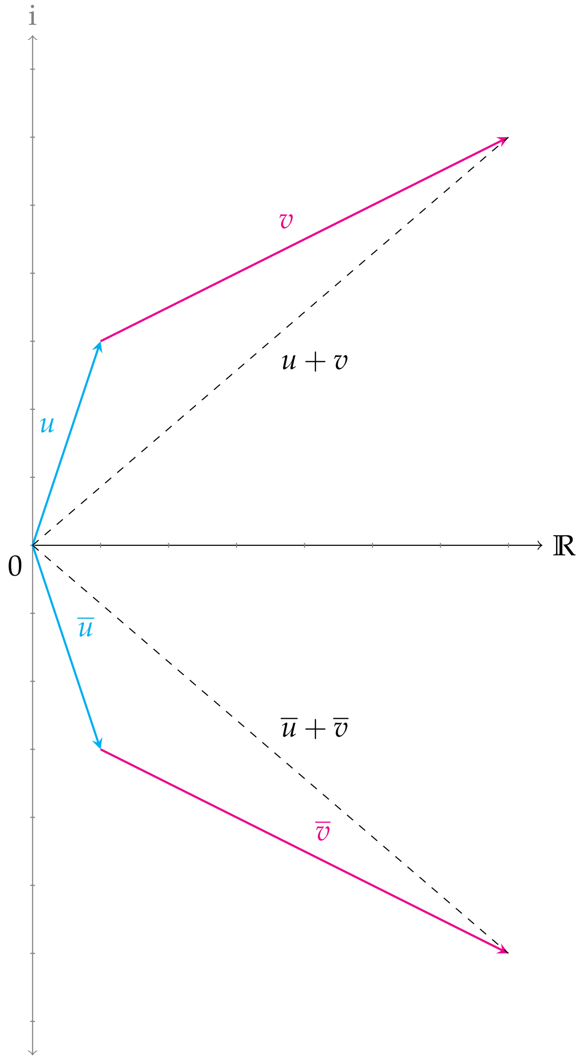


Figure 1.4: $\overline{u+v} = \overline{u} + \overline{v}$.

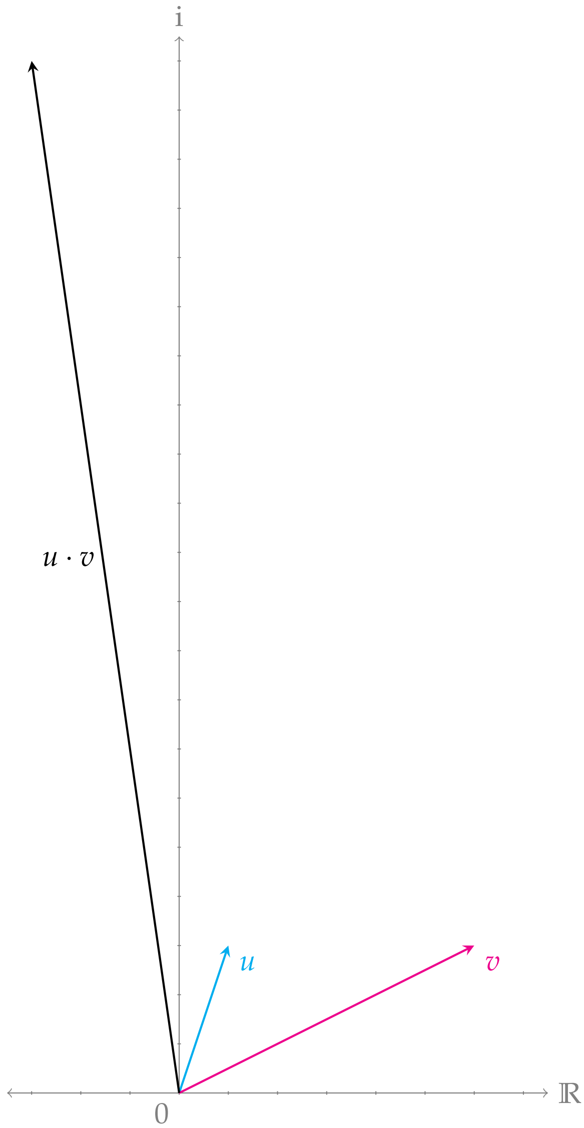


Figure 1.5: Let $u = (1 + 3\mathrm{i})$ and $v = (6 + 3\mathrm{i})$ and notice $u \cdot v = (-3 + 21\mathrm{i})$.

Definition 1.14 (Modulus)**.** For $z = r(\cos(\theta) + \mathrm{i}\sin(\theta))$, a complex number in polar form, the *modulus of z*, denoted $|z|$, is given by

$$|z| := r.$$

Definition 1.15 (Argument)**.** For $z = r(\cos(\theta) + \mathrm{i}\sin(\theta))$, a complex number in polar form, the *argument of z*, denoted $arg(z)$, is given by

$$arg(z) := \theta$$

and moreover, if

$$0 \leq arg(z) \leq 2\pi$$

then $arg(z)$ is called the *principle argument of z*.

Since the value of θ is not unique, as there are many (in fact infinitely many) values of θ representing the same position, the same complex number has many different writings in polar form, namely

$$r(\cos(\theta) + \mathrm{i}\sin(\theta)) = r(\cos(\theta + 2k\pi) + \mathrm{i}\sin(\theta + 2k\pi))$$

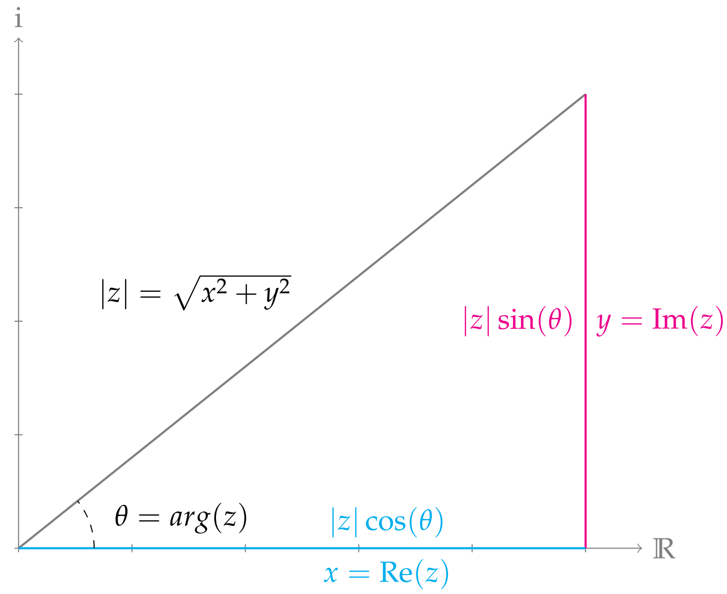
for any integer k. To mitigate this we usually insist that the angle θ lies in $[0, 2\pi]$.

Exercise 1.4. What is the polar coordinate of the complex number $0 \in \mathbb{C}$? In particular, what is $arg(0)$?

The relationship between the cartesian and polar form of a complex number $z \in \mathbb{C}$ is given by

$$\mathrm{Re}(z) + \mathrm{i}\,\mathrm{Im}(z) = |z|\left[\cos(arg(z)) + \mathrm{i}\sin(arg(z))\right] \tag{1.3}$$

as illustrated below.



The relationship between polar and cartesian forms.

Example 1.16. The complex number in Cartesian form $z = 1 + \text{i}$ has polar form

$$\begin{aligned} z &= \sqrt{1^2+1^2}\left[\cos(\arctan(1)) + \text{i}\sin(\arctan(1))\right] \\ &= \sqrt{2}\left(\cos\left(\frac{\pi}{4}\right) + \text{i}\sin\left(\frac{\pi}{4}\right)\right). \end{aligned}$$

Proposition 1.17. The complex numbers satisfying $|z| = r \in \mathbb{R}$ form a circle of radius r in the complex plane.

PROOF. Let $z = x + \text{i}y$. The points satisfying $|z| = r$ must be solutions to

$$\sqrt{x^2+y^2} = r \implies x^2+y^2 = r^2$$

which we recognize as the equation for a circle. ■

Proposition 1.18. For $z \in \mathbb{C}$, $|z| = 0$ if and only if $z = 0$.

PROOF. Let $z = x + \text{i}y$ then $|z| = 0 \iff \sqrt{x^2+y^2} = 0 \iff x = y = 0$. Thus $z = 0$ only when $z = 0$. ■

1.4.1 Multiplication of Polar Complex Numbers

Multiplication has an interesting geometric interpretation when computed in polar form. If $z = r(\cos(\theta) + i\sin(\theta))$ and $w = s(\cos(\varphi) + i\sin(\varphi))$, then

$$\begin{aligned} zw &= r(\cos(\theta) + \mathrm{i}\sin(\theta)) \cdot s(\cos(\varphi) + \mathrm{i}\sin(\varphi)) \\ &= rs\left((\cos(\theta)\cos(\varphi) - \sin(\theta)\sin(\varphi)) + \mathrm{i}\left(\cos(\theta)\sin(\varphi) + \sin(\theta)\cos(\varphi)\right)\right) \\ &= rs(\cos(\theta + \varphi) + \mathrm{i}\sin(\theta + \varphi)) \end{aligned}$$

Thus we have proven that.

Proposition 1.19. For two complex numbers $z, w \in \mathbb{C}$

1. $|zw| = |z||w|$, and
2. $arg(zw) = arg(z) + arg(w)$.

PROOF. See above. ■

1.4.2 Division of Polar Complex Numbers

Division also has an interesting geometric interpretation. An argument like that of Proposition 1.19 proves a similar result for division.

Proposition 1.20. For complex numbers $z, w \in \mathbb{C}$,

1. $\left|\frac{z}{w}\right| = \frac{|z|}{|w|}$, and
2. $arg(\frac{z}{w}) = arg(z) - arg(w)$.

PROOF. ■

Exercise 1.5. Find complex numbers z and w such that

$$\mathrm{Im}(zw) \neq \mathrm{Im}(z)\mathrm{Im}(w).$$

(Thus the above cannot hold in general.)

1.5 Euler's Formula

One of the most famous theorems in complex numbers, and indeed in mathematics, is Euler's formula. It specifies a connection between the

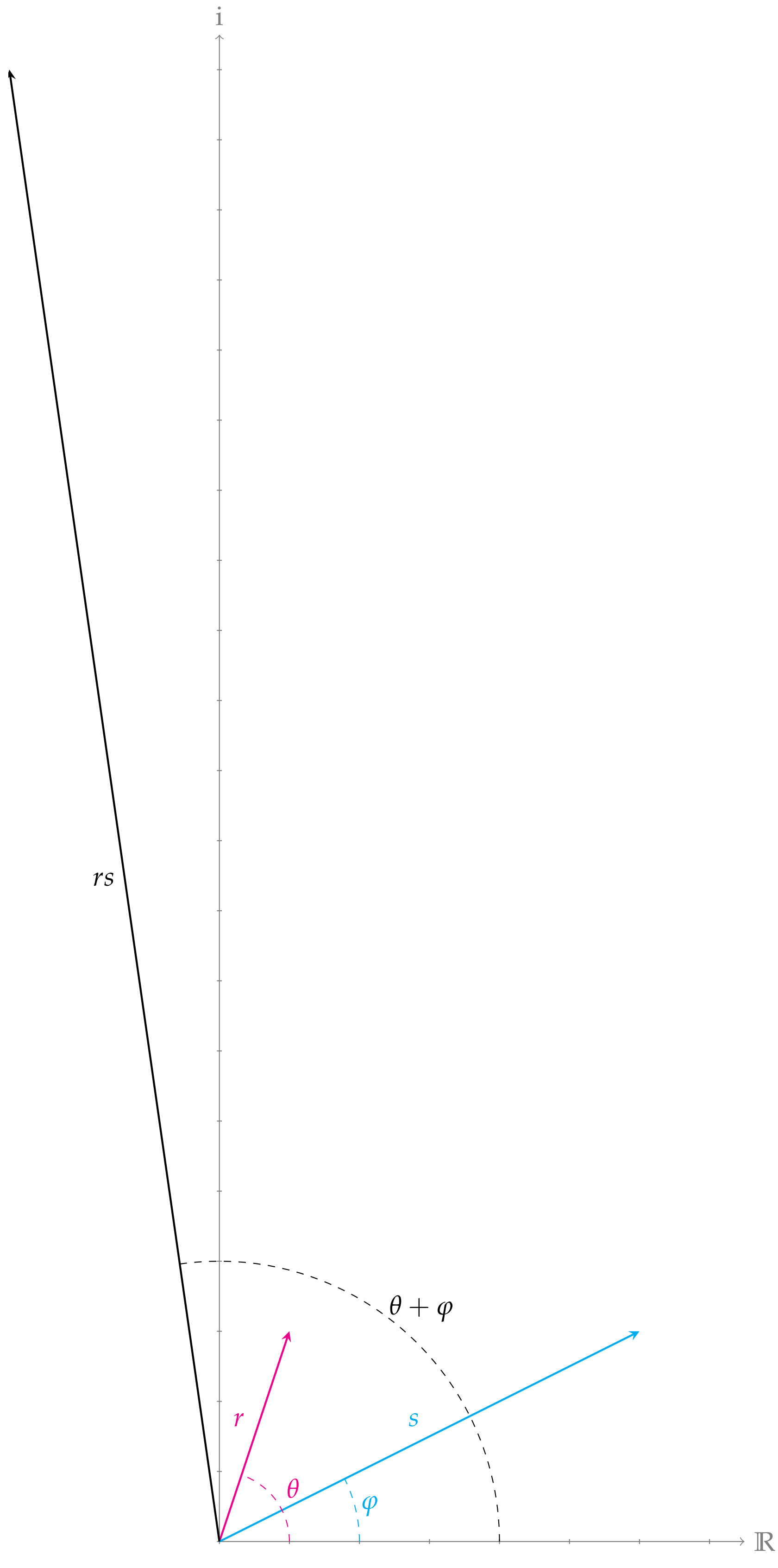
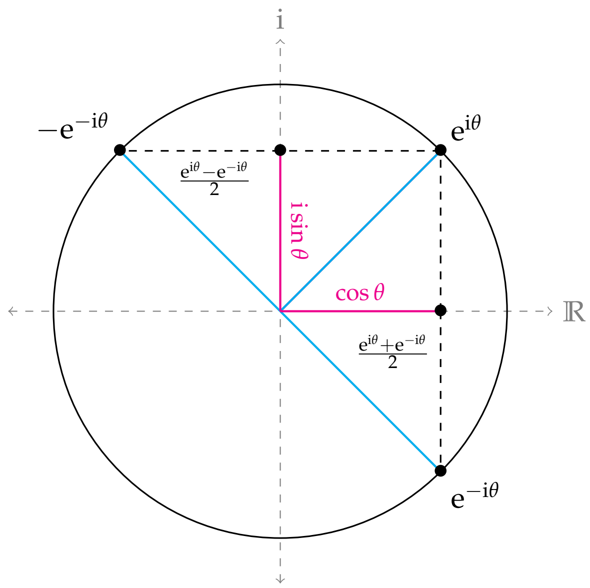


Figure 1.6: Multiplication in polar.

exponential function applied to complex numbers, and the trigonometric functions.



Theorem 1.21 (Euler's Formula)**.** Let $r \geq 0$ and θ be real numbers. Then

$$e^{i\theta} = \cos(\theta) + i\sin(\theta). \tag{1.4}$$

PROOF. We know that e^{ix} should correspond to *some* complex number $r[\cos(\theta) + i\sin(\theta)]$, that is, there must be some θ and r such that

$$e^{ix} = r\left[\cos(\theta) + i\sin(\theta)\right]. \tag{1.5}$$

So, to prove Euler's formula, let us find solutions to (1.5). (We expect in the end to find $x = \theta$ and $r = 1$).

RHS and LHS denotes right-hand-side of and left-hand-side of (some equation).

Differentiating both sides of (1.5) with respect to x gives[1]

$$\frac{d\text{LHS}(1.5)}{dx} = \frac{de^{ix}}{dx} = ie^{ix}$$

and

$$\frac{d\text{RHS}(1.5)}{dx} = \frac{d\,r\left[\cos(\theta) + i\sin(\theta)\right]}{dx}$$

[1]Here we have assumed that differentiation on complex numbers behave the same way as real numbers. This is true, but you will not learn why until your course in complex analysis!

$$= \frac{\mathrm{d}r}{\mathrm{d}x}\left[\cos(\theta) + \mathrm{i}\sin(\theta)\right] + r\left[-\sin(\theta) + \mathrm{i}\cos(\theta)\right]\frac{\mathrm{d}\theta}{\mathrm{d}x}.$$

This implies, using (1.5) to substitute for $\mathrm{e}^{\mathrm{i}x}$, that

$$\begin{aligned} &\mathrm{i}\, r\left[\cos(\theta) + \mathrm{i}\sin(\theta)\right] \\ &\quad = \frac{\mathrm{d}r}{\mathrm{d}x}\left[\cos(\theta) + \mathrm{i}\sin(\theta)\right] + r\left[-\sin(\theta) + \mathrm{i}\cos(\theta)\right]\frac{\mathrm{d}\theta}{\mathrm{d}x}. \end{aligned} \tag{1.6}$$

Notice $\mathrm{Re}(\text{LHS}(1.6)) = -\sin(\theta)$ and $\mathrm{Im}(\text{LHS}(1.6)) = \cos(\theta)$ and, by equating these with $\mathrm{Re}(\text{RHS}(1.6))$ and $\mathrm{Im}(\text{RHS}(1.6))$, we are left with the system of equations

$$\begin{aligned} -r\sin(\theta) &= \frac{\mathrm{d}r}{\mathrm{d}x}\cos(\theta) - r\sin(\theta)\frac{\mathrm{d}\theta}{\mathrm{d}x}, \\ r\cos(\theta) &= \frac{\mathrm{d}r}{\mathrm{d}x}\sin(\theta) + r\cos(\theta)\frac{\mathrm{d}\theta}{\mathrm{d}x}. \end{aligned}$$

and thereby $\frac{\mathrm{d}r}{\mathrm{d}x} = 0$, $\frac{\mathrm{d}\theta}{\mathrm{d}x} = 1$. Moreover,

$$\theta = \int \frac{\mathrm{d}\theta}{\mathrm{d}x}\mathrm{d}x = \int 1\,\mathrm{d}x = x + C_0$$

and

$$r = \int \frac{\mathrm{d}r}{\mathrm{d}x}\mathrm{d}x = \int 0\,\mathrm{d}x = C_1.$$

To complete our proof remember that θ and r functions because their values depend on the value of x. Coupling this observation with

$$\mathrm{e}^{\mathrm{i}0} = 1 = (1)(\cos(0) + \mathrm{i}\sin(0))$$

enables us deduce the boundary conditions

$$\theta(0) = 0 \qquad\qquad \text{and} \qquad\qquad r(0) = 1.$$

and thereby $C_0 = 0$ and $C_1 = 1$ which implies $r = 1$ and $\theta = x$.

We have from (1.5) that

$$\mathrm{i}\mathrm{e}^{\mathrm{i}x} = r[-\sin(\theta) + \mathrm{i}\cos(\theta)] \implies \mathrm{e}^{\mathrm{i}x} = r[\cos(\theta) + \mathrm{i}\sin(\theta)]$$

(we multiplied by $-\mathrm{i}$ on both sides). Combining this with $r = 1$, and $\theta = x$ yields

$$\mathrm{e}^{\mathrm{i}\theta} = \cos(\theta) + \mathrm{i}\sin(\theta).$$

■

1.5.1 The Most Beautiful formula

Setting $\theta = \pi$ in Euler's formula gives $e^{i\pi} = -1$ which we can write as

$$e^{i\pi} + 1 = 0$$

so as to include the five most well-known constants in all of mathematics. Namely

1. *Euler's number* "e" which is the base of the natural logarithm.
2. The *Complex number* "i" which we are studying.
3. *Pi* "π" which is the circumference of a unit circle.
4. *Zero* "0" whose discovery not only let us write 1999 instead of

   $$MDCCCCLXXXXVIIII$$

   but also made arithmetic on numbers much easier.

1.6 De Moivre's Theorem

We know how to multiply complex numbers and thereby also know how to take complex numbers to powers. However, consider calculating

$$(1 + i)^{100}$$

— this would require 99 complex multiplications! Surely there must be a better way.

There is. First notice a consequence of Euler's formula is that any complex number can be written in *complex exponential form*

Definition 1.22 (Complex Exponential Form)**.** Let $z = r[\cos(\theta) + i\sin(\theta)]$ be a complex number in polar form. The *complex exponential form* of z is

$$re^{i\theta}.$$

Proposition 1.23. Let $z = x + iy$ be a complex number written in cartesian form. The complex exponential form of z is

$$|z| = e^{i\arg(z)}.$$

PROOF. This follows immediately from the definition. ■

In this form, taking high powers of complex numbers becomes dramatically easier because

$$\left(re^{i\theta}\right)^n = r^n e^{in\theta}.$$

Example 1.24. Let $z := (1+i)$ and notice $|z| = \sqrt{2}$ and $arg(z) = \arctan(1) = \frac{\pi}{4}$.

$$\begin{aligned} z^{100} &= (1+i)^{100} \\ &= [\sqrt{2}e^{i\frac{\pi}{4}}]^{100} \\ &= 2^{50}e^{i\,25\,\pi} \\ &= 2^{50}e^{i\,12\,(2\pi)}e^{i\pi} \\ &= 2^{50}\cdot 1\cdot(-1) \\ &= -2^{50}. \end{aligned}$$

Generally, when $z = r[\cos(\theta) + i\sin(\theta)]$ is a complex number in polar form and $n \in \mathbb{N}$ is an integer we have

$$z^n = |r|^n e^{ni\theta}. \tag{1.7}$$

Proposition 1.25. Let $z \in \mathbb{C}$ be a complex number.

1. $|z^n| = |z|^n$, and
2. $arg(z^n) = n\,arg(z)$.

PROOF. Exercise. ■

This method of powering is actually due to De Moivre's who states it as below.[2]

Theorem 1.26 (De Moivre's formula). Let $x \in \mathbb{R}$ and $n \in \mathbb{N}$ an integer. Then

$$[\cos(x) + i\sin(x)]^n = \cos(nx) + i\sin(nx).$$

PROOF. We have by Euler's formula that $e^{ix} = \cos(x) + i\sin(x)$ and also that $\left(e^{ix}\right)^n = e^{inx}$ by the exponential law. Thereby

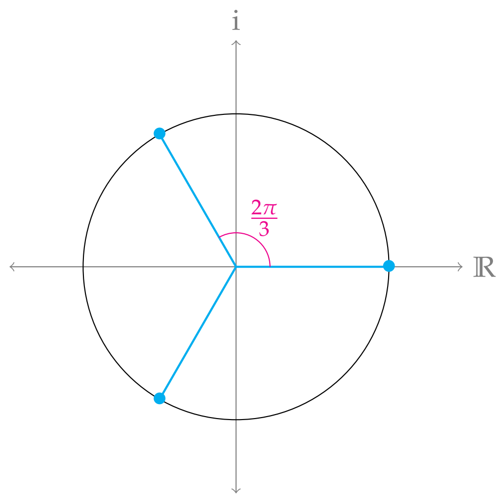
$$e^{i(nx)} = \cos(nx) + i\sin(nx).$$

[2] Abraham de Moivre (1667–1754) French mathematician known for de Moivre's formula, one of those that link complex numbers and trigonometry, and for his work on the normal distribution and probability theory. He was a friend of Isaac Newton, Edmond Halley, and James Stirlingand wrote e a book on probability theory that was prized by gamblers. [Wikipedia]

■

1.7 Roots of Complex Numbers

The ideas in De Moivre's Theorem can be inverted to find roots of complex numbers. For example we know that the cube-root of 8 is 2 (because $2^3 = 8$) but did you know $-1 + i\sqrt{3}$ and $-1 - \sqrt{3}$ are also cube roots of 8?



In fact it is no accident that there are three cube-roots of 8 — any complex number z has n distinct nth roots.

Definition 1.27 (nth root)**.** For z and ζ (zeta) complex numbers, we say ζ is a nth root of z when

$$\zeta^n = z.$$

Proposition 1.28. Let $z = re^{i\theta}$ and let $n \in \mathbb{N}$ then z has n distinct nth roots given by

$$\zeta_k = r^{\frac{1}{n}} e^{i\left(\frac{\theta+2\pi k}{n}\right)}$$

for $k = 0, \ldots, n-1$.

PROOF. Exercise. ■

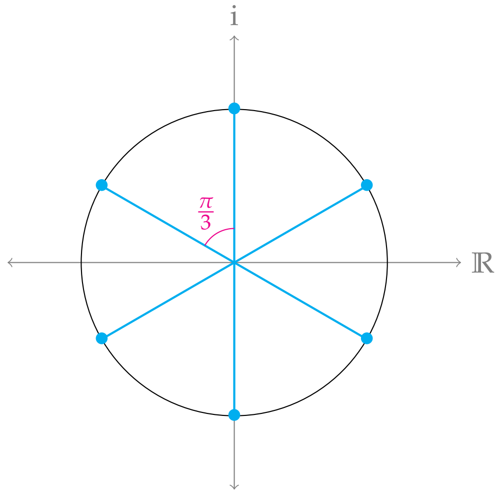
Example 1.29. To find the 6th roots of -8 first write -8 as $8e^{i\pi}$ and thus

$$\sqrt[6]{-8} = \sqrt[6]{8}\, e^{i\frac{\pi+2\pi k}{6}}$$

for $k = 1, \ldots, 6$. The six distinct roots are thereby

$$\left\{ \sqrt[6]{8}e^{i\frac{\pi}{6}},\ \sqrt[6]{8}e^{i\frac{\pi}{2}},\ \sqrt[6]{8}e^{i\frac{5\pi}{6}},\ \sqrt[6]{8}e^{i\frac{7\pi}{6}},\ \sqrt[6]{8}e^{i\frac{3\pi}{2}},\ \sqrt[6]{8}e^{i\frac{11\pi}{6}} \right\}$$

which, plotted on the plane, are



In cartesian form these numbers are

$$\left\{ \pm\frac{\sqrt[6]{8}\sqrt{3}}{2} \pm \frac{\sqrt[6]{8}}{2}i,\ \pm\sqrt[6]{8}i,\ \pm\frac{\sqrt[6]{8}}{2}i \mp \frac{\sqrt[6]{8}\sqrt{3}}{2} \right\}.$$

We have just found all solutions to $z^n = -8$ for a particular value of n. When we sove $z^n = 1$, the roots are called *roots of unity.*

Exercise 1.6. Check that the cube roots of 1 are

$$\left\{ e^0,\ e^{\frac{2\pi}{3}i},\ e^{\frac{4\pi}{3}i} \right\}$$

by cubing their equivalent cartesian forms, which are

$$\left\{ 1,\ \frac{-1}{2} + \frac{\sqrt{3}}{2}i,\ \frac{-1}{2} - \frac{\sqrt{3}}{2}i \right\}.$$

1.8 *The Fundamental Theorem of Algebra*

We have seen that the equation $z^n = 1$ has exactly n distinct solutions. This idea generalizes, as follows.

Theorem 1.30 (Fundamental Theorem of Algebra)**.** A degree n polynomial from $\mathbb{R}[z]$ has at least one complex root. Namely,

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

where $a_n \neq 0$ and $a_0, \ldots, a_n \in \mathbb{R}$ has at least one complex solution. Moreover, the polynomial factorizes, or "splits", completely over $\mathbb{C}$, in the form

$$a_n (z - \zeta_1)(z - \zeta_2) \cdots (\zeta - c_\ell)$$

for some (not necessarily distinct) complex numbers $\zeta_1, \zeta_2, \ldots, \zeta_\ell \in \mathbb{C}$ and $\ell \leq n$.

A way of restating the Fundamental Theorem is to say that every polynomial of degree n with real coefficients has n roots, 'counting multiplicities'.

Example 1.31. The polynomial $x^4 + 2x^2 - 1$ has four root counting multiplicities because

$$x^4 + 2x^2 - 1 = (x - \mathrm{i})^2 (x + \mathrm{i})^2.$$

Here we say i and $-\mathrm{i}$ have *multiplicity* 2; making a total of 2+2=4 roots.

We will not prove the Fundamental Theorem of Algebra here because it is far out of the scope of this course. You will likely see a proof in some subsequent Algebra or Complex Analysis course. We do note that it is quite a surprising result. Why should we not need more kinds of numbers in order to find the roots of high degree polynomials?

An even more surprising result also holds.

Theorem 1.32. The Fundamental Theorem of Algebra remains true when the coefficients $a_0, ..., a_n$ are taken to be elements of $\mathbb{C}$.

There is a name for this last result. We say that $\mathbb{C}$ is *algebraically closed* because any polynomial with complex coefficients will have all its solutions in the complex numbers.

1.8.1 Solving Polynomials

Theorem 1.33 (Cubic Equation). The solutions to the cubic equation $ax^3 + bx^2 + cx + d$ are given by

$$x = \sqrt[3]{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right) + \sqrt{\left(\frac{-b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a}\right)^2 + \left(\frac{c}{3a} - \frac{b^2}{9a^2}\right)^3}}$$

$$+\sqrt[3]{\left(\frac{-b^3}{27a^3}+\frac{bc}{6a^2}-\frac{d}{2a}\right)-\sqrt{\left(\frac{-b^3}{27a^3}+\frac{bc}{6a^2}-\frac{d}{2a}\right)^2+\left(\frac{c}{3a}-\frac{b^2}{9a^2}\right)^3}}$$
$$-\frac{b}{3a}.$$

Theorem 1.34 (Quartic Equation). Let f be monic degree four polynomial given by

$$f(x) = x^4 + a_3x^3 + a_2x^2 + a_1x + a_0.$$

One can find the roots of $f(x)$ by splitting the *cubic* equation

$$g(x) = x^3 + b_2x^2 + b_1x + b_0$$

where the coefficient b_i are given in terms of the a_i by

$$\begin{aligned} b_2 &= -a_2 \\ b_1 &= a_1a_3 - 4a_0 \\ b_0 &= 4a_0a_2 - a_1^2 - a_0a_3^2. \end{aligned}$$

Suppose $g(x) = (x - \beta_1)(x - \beta_2)(x - \beta_3)$, that is, the individual βs are the solutions to the cubic. Then the roots of the quartic f are given by

$$\begin{aligned} \beta_1 &= \alpha_1\alpha_2 + \alpha_3\alpha_4, \\ \beta_2 &= \alpha_1\alpha_3 + \alpha_2\alpha_4, \\ \beta_3 &= \alpha_1\alpha_4 + \alpha_2\alpha_3. \end{aligned}$$

PROOF. Beyond scope. ■

Theorem 1.35. There are no general forms for the roots of polynomials of degree five or higher.

PROOF. The purpose of Galois theory. But, for instance,

$$x^5 - x - 1$$

has no radical solutions. ■

1.9 *Complex Equalities and Inequalities*

We study the regions defined by inequalities on complex numbers. These are regions and not simply points because there is no ordering on the complex numbers.

Consider that for any two real numbers $a, b \in \mathbb{R}$ it is either the case that $a = b$ or $a \leq b$ or $b \geq a$. Generally, a set X has a *total order* when its ordering $\leq$ satisfies some constraints.

Definition 1.36 (Total order)**.** Let X be a set and $\leq$ a binary relation. X is a *totally ordered set* when for any $a, b \in X$

$$\begin{aligned} a \leq b \text{ and } b \leq a \implies a = b && \text{antisymmetry} \\ a \leq b \text{ and } b \leq c \implies a \leq c && \text{transitivity} \\ a \leq b \text{ or } b \leq a && \text{totality.} \end{aligned}$$

The complex numbers are *not* ordered. For instance, which of $2 + \mathrm{i}$ or $1 + 2\mathrm{i}$ is bigger?

1.9.1 Regions the Complex Plane

Consider the *inequality* $|z| < 3$. Intuitively these will be the collection of complex numbers whose modulus is strictly less than 3 (as opposed to less than or equal to). We can plot both of these *regions* as below, noting that dashed lines are used to denote open regions (e.g. the points on the dashed line are *not* included).

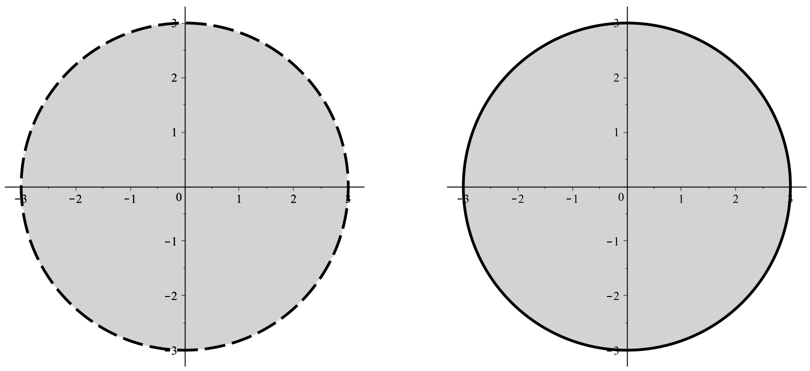


Figure 1.7: The complex numbers satisfying $|z| < 3$ (left) and $|z| \leq 3$ (right).

To give more detail, the complex numbers $z \in \mathbb{C}$ satisfying $|z| < 3$ are those points lying on the *inside* of the boundary $|z| = 3$. If we express z in its cartesian form $z = x + \mathrm{i}y$ then $|z| = 3$ becomes

$$\sqrt{x^2 + y^2} = 3 \implies x^2 + y^2 = 9.$$

We recognize this as a circle with radius three.

We may relate a circle of radius r centered at (a, b) in the cartesian plane, with a complex equation by

$$(x-a)^2 + (y-b)^2 = r^2 \iff |z - (a + bi)| = r.$$

Proposition 1.37 (Circle in the Complex Plane)**.** Let z be a complex variable, w be a complex number, and $r \in \mathbb{R}$, then

$$|z - w| = r$$

defines a circle centered at w with radius r in the complex plane.

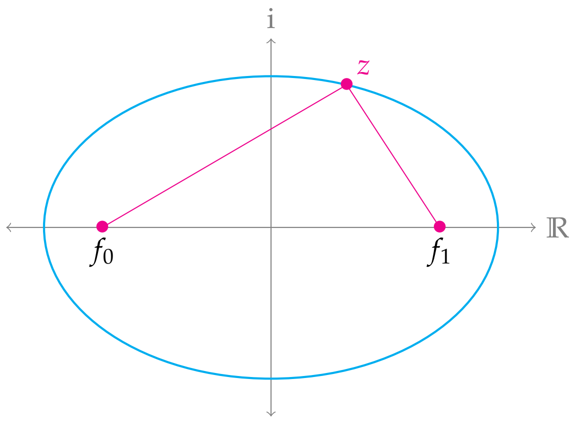
PROOF.

$$\begin{aligned} |z-w| = r &\iff |(x-s) + i(y-t)| < r \\ &\iff \sqrt{(x-s)^2 + (y-t)^2} = r \\ &\iff (x-s)^2 + (y-t)^2 = r^2. \end{aligned}$$

We recognize this as the equation of the circle. ■

1.9.2 Ellipses

An ellipse is a shape that is defined by two foci (f_0 and f_1) and looks like



When the two foci are distinct, the shape will be a general ellipse; when they overlap the shape becomes a circle.

An ellipse can be drawn by connecting the ends of a piece of string to the foci and using a pencil pushed up against the string to draw a line at the furthest distance possible from the two points. Keeping the string taut while drawing from 0 to 2π constructs an ellipse.

We can extract an equation for the ellipse by exploiting a geometric property it satisfies. On an ellipse the distances from both foci to the same point on the ellipse sum to a constant. That is,

$$|z - f_1| + |z - f_2| = c$$

when $f_0, f_1 \in \mathbb{C}$ and $c \in \mathbb{R}$ is a constant (c is the length of the string).

Proposition 1.38 (Ellipse in the Complex Plane)**.** Let z be a complex variable, f_0, f_1 be a complex numbers, and $r \in \mathbb{R}$ such that $r > |f_1 - f_2|$, then

$$|z - f_1| + |z - f_2| = r$$

defines an ellipse with foci at f_1 and f_2.

1.9.3 *Hyperbola*

Let $z = x + iy \in \mathbb{C}$ and consider the inequality given by

$$\operatorname{Re}(z^2) \leq 3. \tag{1.8}$$

The boundary for the plot of this inequality will be given by the associated equality $\operatorname{Re}(z^2) = 3$, which converts to a cartesian equation via

$$\begin{aligned} \operatorname{Re}(z^2) = 3 &\implies \operatorname{Re}((x + iy)^2) = 3 \\ &\implies x^2 - y^2 = 3. \end{aligned}$$

Thus, the points satisfying (1.8) will lie *inside* or *on* $h = x^2 - y^2 - 3$.

Notice h has x and y intercepts given by (resp.) $x^2 = 3$ and $y^2 = -3$. The equation also has asymptotes: when x and y are very large, their behaviour is mostly determined by the highest degree terms (in this case 2) and not the linear and constant terms. So we have asymptotes at

$$x^2 - y^2 = 0$$

which become

$$y = \pm x,$$

and, because the angles between the asymptotes are $\frac{\pi}{2}$, this is called a *rectangular hyperbola*.

The following sketch shows the hyperbola, the region associated with $x^2 - y^2 \leq 3$, and the asymptotes. We test which regions to shade by testing a point in each — the obvious choice is to choose $z = 5$ and $z = -5$ which lie *outside* the region, and 0 which lies $z =$ *inside*. (Because $5^2 - 0^2 \not\leq 3$ and so on.) A continuity argument (which we do not give here) determines that if a point in a region is shaded then the entire region is shaded.

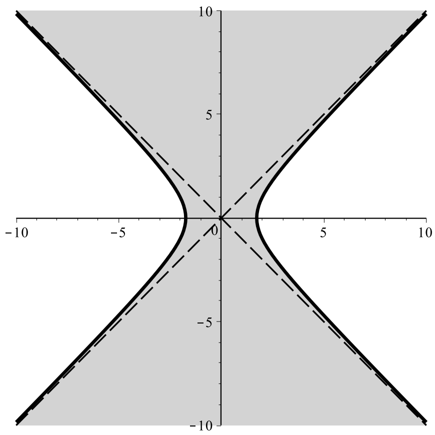


Figure 1.8: The points satisfying $\mathrm{Re}(z^2) \leq 3$.

Proposition 1.39. Let z be a complex variable, w be a complex number, and $c \in \mathbb{R}$, then

$$\mathrm{Re}((z-w)^2) = c$$

defines a rectangular hyperbola centered at w.

PROOF. Let $z = x + iy$ and $w = s + it$ then

$$\begin{aligned} \mathrm{Re}((z-w)^2) &= x^2 - 2s + s^2 - y^2 + 2yt - t^2 \\ &= (x-s)^2 - (y-t)^2 \end{aligned}$$

■

1.9.4 *Hyperbola with horizontal/vertical asymptotes*

The region defined by $\mathrm{Im}((x+iy)^2) \leq 3$ is similar to that of its real counterpart except now we have a boundary given by

$$\mathrm{Im}(z^2) = 3 \implies \mathrm{Im}((x+iy)^2) = 3 \implies 2xy = 3.$$

which defines the curve

$$y = \frac{3}{2x}.$$

This is still a hyperbola except the asymptotes are horizontal and vertical.

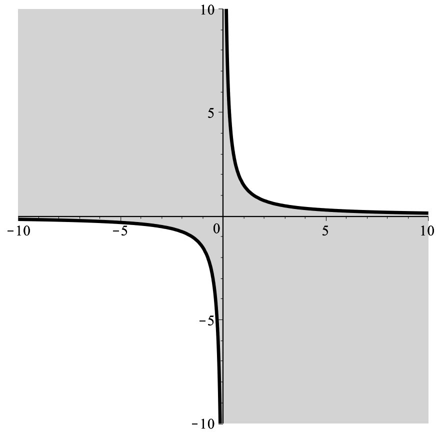


Figure 1.9: The points satisfying $\text{Im}(z^2) \leq 3$.

1.9.5 *Straight Lines*

The points satisfying the expression $|z - \text{i}| = |z - 3|$ are those $z \in \mathbb{C}$ equidistant from i and 3. This is a straight line which is the perpendicular bisector of the line segment joining i and 3.

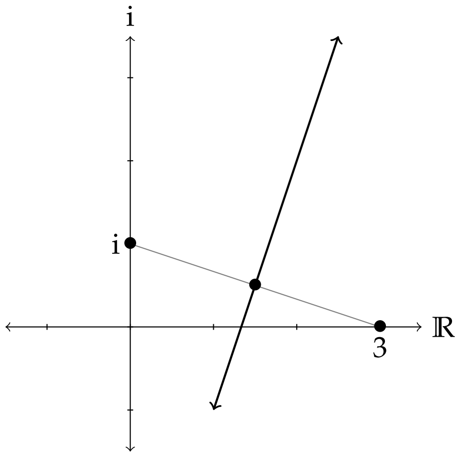


Figure 1.10: The points satisfying $|z - \text{i}| = |z - 3|$.

Proposition 1.40 (Line in the Complex Plane)**.** Let z be a complex variable, w_0 and w_1 be complex numbers such that $w_0 \neq w_1$, then

$$|z - w_0| = |z - w_1|$$

defines a line through $\frac{w_0+w_1}{2}$ with slope $-\frac{\text{Re}(w_1)-\text{Re}(w_0)}{\text{Im}(w_1)-\text{Im}(w_0)}$.

1.9.6 Wedges

The inequality $0 \leq arg(z) \leq \pi/4$ represents a wedge at z sweeping out the angle 0 through $\pi/4$.

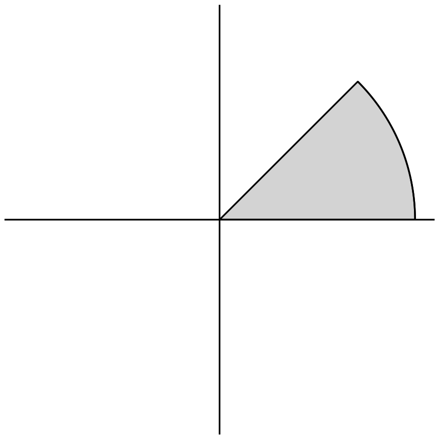


Figure 1.11: The points satisfying $0 \leq arg(z) \leq \pi/4$

Proposition 1.41 (Wedge in the Complex Plane)**.** Let θ_0 and θ_1 be angles in $[0, 2\pi]$ such that $\theta_0 < \theta_1$, z a complex variable, and $z_0 \in \mathbb{C}$ fixed. Then

$$\theta_0 \leq arg(z - z_0) \leq \theta_1$$

sweeps out a wedge at z_0 from θ_0 to θ_1.

2 Vectors

2.1 Why Vectors?

A natural question to ask is if the complex numbers be generalized. The answer is no...ish.

William Rowan Hamilton spent the years 1830–1843 searching (in vain) for rules that govern complex numbers with *two* imaginary numbers i and j. The property which proved unattainable was that complex numbers satisfy $|u||v| = |uv|$ for $u = x + \mathrm{i}y$ and $v = s + \mathrm{i}t$. In particular, we have

$$\begin{aligned} |x+\mathrm{i}y|^2|s+\mathrm{i}t|^2 &= (x^2+y^2)(s^2+t^2) \\ &= (xs)^2+(xt)^2+(ys)^2+(yt)^2 \\ &= (xs)^2+(xt)^2+(ys)^2+(yt)^2-2xyst+2xyst \\ &= (xs-yt)^2+(xt+ys)^2 \\ &= |(xs-yt)+\mathrm{i}(xt+ys)|^2 \end{aligned}$$

Definition 2.1 (Cartesian Product). Let A be a set. Then the *cartesian product* of A is given by

$$A^n = \underbrace{A \times \cdots \times A}_{n\text{-times}} := \{(a_0, \ldots, a_{n-1}) : a_i \in A\}$$

for n some non-negative integer.

Example 2.2 (Cartesian Product). The points of the real-Cartesian plane is given by

$$\mathbb{R} \times \mathbb{R} = \mathbb{R}^2 = \{(a, b) : a, b \in \mathbb{R}\}.$$

For example, $(1, 2) \in \mathbb{R}^2$. Moreover, let $\mathbf{0} = (0, \ldots, 0) \in \mathbb{R}^n$ denote the origin.

Proposition 2.3. $\mathbb{R}^2$ and $\mathbb{C}$ are isomorphic. That is, they are in a well-defined sense, identical to one-another $\mathbb{R}^2 \cong \mathbb{C}$.

PROOF. Take a group theory course. ■

Any complex number $x + iy$ can be identified by the ordered pair

$$(x, y) \in \mathbb{R} \times \mathbb{R}$$

and we can define arithmetic on this form by:

$$\begin{aligned}(x,\, y) + (s,\, t) &:= (x + y,\, s + t),\\ |(x, y)|^2 &:= x^2 + y^2,\\ (x,\, y)(s,\, t) &:= (xs - yt,\, xt + sy).\end{aligned}$$

Re-writing, we have, for complex numbers (x, y) and (s, t) that

$$|(x, y)|^2 |(s, t)|^2 = |(x, y)(s, t)|^2.$$

Thereby an extension to complex numbers in $\mathbb{R}^3$ must define $(x, y, z)(r, s, t)$ so that

$$|(x, y, z)|^2 |(r, s, t)|^2 = |(x, y, z)(r, s, t)|^2.$$

However, no matter how we define multiplication the product $(x, y, z)(r, s, t)$ must remain in $\mathbb{R}^3$ because of closure. Thus, there must be some $(\alpha, \beta, \gamma) \in \mathbb{R}^3$ such that

$$(x, y, z)(r, s, t) = (\alpha, \beta, \gamma)$$

which implies

$$|(x, y, z)(r, s, t)|^2 = |(\alpha, \beta, \gamma)|^2 = \alpha^2 + \beta^2 + \gamma^2.$$

It follows there must be some $(\alpha, \beta, \gamma) \in \mathbb{R}^3$ so that

$$\begin{aligned}|(x, y, z)|^2 |(r, s, t)|^2 &= (x^2 + y^2 + z^2)(r^2 + s^2 + t^2)\\ &= \alpha^2 + \beta^2 + \gamma^2. \qquad (2.1)\end{aligned}$$

Theorem 2.4 (Legendre). It is impossible to express 63 as a sum of three squares.

So consider $(1, 1, 1)$ and $(1, 2, 4)$. We have

$$|(1, 1, 1)|^2 |(1, 2, 4)|^2 = (3)(21) = 63.$$

Which means (2.1) cannot be true and we can conclude there is no extension of the complex numbers to $\mathbb{R}^3$.

2.1.1 Quaternions

There *does* exist an extension (ish — they are not commutative) to $\mathbb{R}^4$ of the complex numbers called *the quaternions*. These quaternions are denoted $\mathbb{H}$ in honour of Hamilton who discovered them and recorded their properties onto a bridge.

Definition 2.5 (Quaternions). Let i, $\mathbf{j}$, and $\mathbf{k}$ satisfy

$$\mathrm{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathrm{i}\mathbf{j}\mathbf{k} = -1$$

then

$$\mathbb{H} = \left\{ a + b\mathrm{i} + c\mathbf{j} + d\mathbf{k} \; : \; (a, b, c, d) \in \mathbb{R}^4 \right\}$$

defines the *quaternion* numbers.

Their arithmetic obeys the *distributive law* but *not* the *commutative law*:

$$\begin{aligned} \mathrm{i}\mathbf{j} &= \mathbf{k}, & \mathbf{j}\mathrm{i} &= -\mathbf{k}, \\ \mathbf{j}\mathbf{k} &= \mathrm{i}, & \mathbf{k}\mathbf{j} &= -\mathrm{i}, \\ \mathbf{k}\mathrm{i} &= \mathbf{j}, & \mathrm{i}\mathbf{k} &= -\mathbf{j}. \end{aligned}$$

Definition 2.6. Multiplication on Quaternions

$$\begin{aligned} &(a + \mathrm{i}b + \mathbf{j}c + \mathbf{k}d)(e + \mathrm{i}f + \mathbf{j}g + \mathbf{k}h) \\ &= (ae - bf - cg - dh) + \mathrm{i}(af + be + ch - dg) \\ &\quad + \mathbf{j}(ag - bh + ce + df) + \mathbf{k}(ah + bg - cf + de). \end{aligned}$$

Definition 2.7. Four-squares Identity

$$\begin{aligned} &(a^2 + b^2 + c^2 + d^2)(e^2 + f^2 + g^2 + h^2) \\ &= (ae - bf - cg - dh)^2 + (af + be + ch - dg)^2 \\ &\quad + (ag - bh + ce + df)^2 + (ah + bg - cf + de)^2. \end{aligned}$$

Notice this means, for two quaternions $h_0, h_1 \in \mathbb{H}$, we have

$$|h_0||h_1| = |h_0 h_1|.$$

2.2 Vector Spaces

What type of properties can we impose on A^n to get useful mathematics?

Definition 2.8 (Vector Space). V, a set, is a *vector space* when for any $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w} \in V$

$\mathbf{u}+\mathbf{v} \in V$	closed under addition,
$\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$	commutativity of addition,
$(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})$	associativity of addition,
$\exists \mathbf{0} : \mathbf{u}+0 = 0+\mathbf{u} = \mathbf{u}$	additive identity,
$\forall \mathbf{u}\ \exists -\mathbf{u} : \mathbf{u}+-\mathbf{u} = \mathbf{0}$	additive inverse,

Moreover let k, $\ell \in \mathbb{R}$ be scalars, then

$k\mathbf{u} \in V$	closed under scalar multiplication,
$k(\mathbf{u}+\mathbf{v}) = k\mathbf{u}+k\mathbf{v}$	distribution of scalar multiplication over vector addition,
$(k+\ell)\mathbf{u} = k\mathbf{u}+\ell\mathbf{u}$	distribution of scalar addition over vector scalar multiplication,
$(k\ell)\mathbf{u} = k(\ell\mathbf{u})$	associativity of scalar multiplication,
$1\mathbf{u} = \mathbf{u}$	scalar multiplicative inverse

(Notice there is no statement about moduli.)

Proposition 2.9. The complex numbers $\mathbb{C}$ form a vector space.

PROOF. We have already shown the complex numbers are closed over an associative and commutative addition operation which has 0 and 1. We need only show complex numbers satisfies the properties on scalars.

For any $k, \ell \in \mathbb{R}$, $z = x + iy \in \mathbb{C}$ and $w = s + it \in \mathbb{C}$ we have: closure under scalar multiplication

$$\begin{aligned} kz &= k(x+iy) \\ &= kx + iky \\ &\implies kz \in \mathbb{C}, \end{aligned}$$

distribution of scalar multiplication over vector addition

$$\begin{aligned} k(z+w) &= k(x+\mathrm{i}y+s+\mathrm{i}t) \\ &= kx+\mathrm{i}ky+ks+\mathrm{i}kt \\ &= k(x+\mathrm{i}y)+k(s+\mathrm{i}t) \\ &= kz+kw, \end{aligned}$$

distribution of scalar addition over vector scalar multiplication

$$\begin{aligned} (k+\ell)z &= (k+\ell)(x+\mathrm{i}y) \\ &= (k+\ell)x+\mathrm{i}(k+\ell)y \\ &= kx+\ell x+\mathrm{i}ky+\mathrm{i}\ell y \\ &= kx+\mathrm{i}ky+\ell x+\mathrm{i}\ell y \\ &= k(x+\mathrm{i}y)+\ell(x+\mathrm{i}y), \end{aligned}$$

and associativity of scalar multiplication

$$\begin{aligned} (k\ell)z &= (k\ell)(x+\mathrm{i}y) \\ &= (k\ell)x+\mathrm{i}(k\ell)y \\ &= k(\ell x)+\mathrm{i}k(\ell y) \\ &= k(\ell x+\mathrm{i}\ell y) \\ &= k(\ell z). \end{aligned}$$

■

2.3 Operations on Vectors

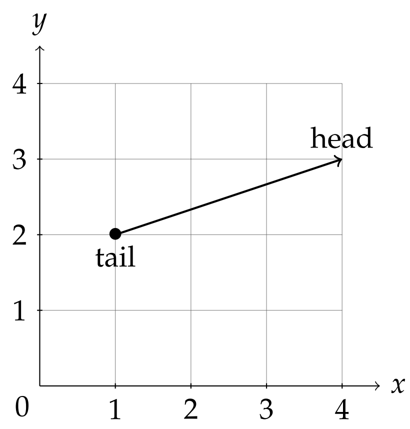
The essential ideas behind geometric/Euclidean vectors are that they have *length* (or magnitude) and *direction*.

Definition 2.10 (Euclidean Vector). A *directed line segment* with a *head* and a *tail*.

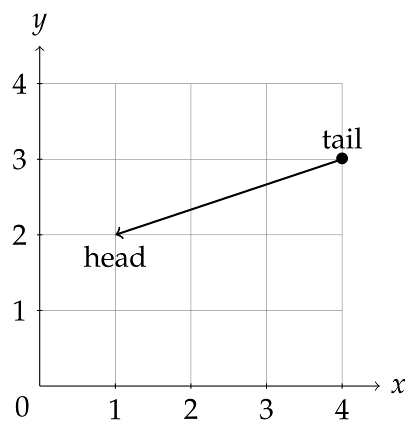
Example 2.11. In physics *velocity* is a vector (speed and direction).

Notation. An arbitrary vector from a vector space is typically denoted $\mathbf{u}$ (in books) and $\overrightarrow{u}$ (on paper). When points $A, B \in \mathbb{R}^n$ are specified then the vector is written $\overrightarrow{AB}$ although $\vec{AB}$ is sometimes preferred (if only because it is easier to write).

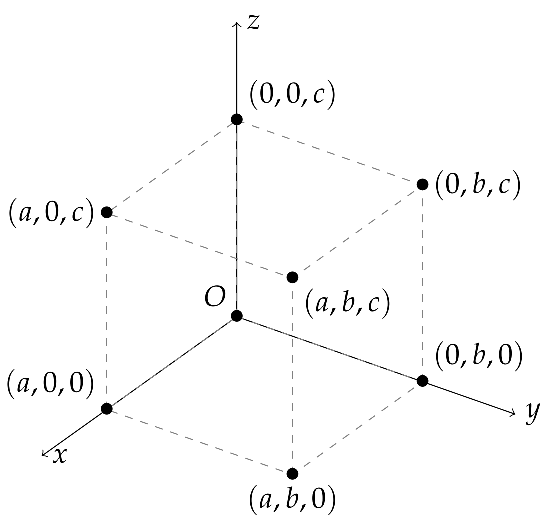
Example 2.12. The vector $\overrightarrow{(1,2)(4,3)}$ drawn on $\mathbb{R}^2$.



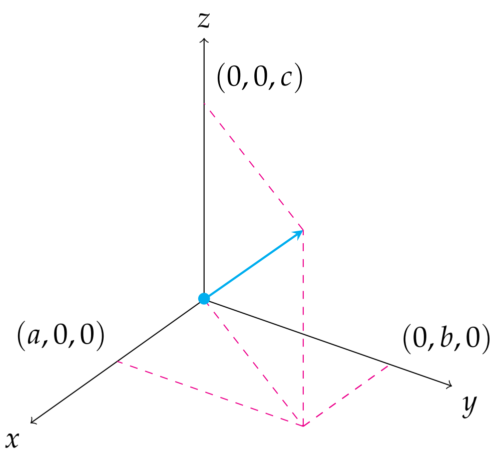
Example 2.13. The vector $\overrightarrow{(4,3)(1,2)}$ drawn on $\mathbb{R}^2$.



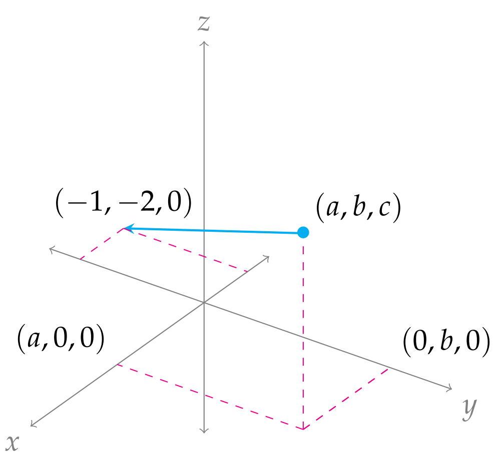
Example 2.14. Points in $\mathbb{R}^3$.



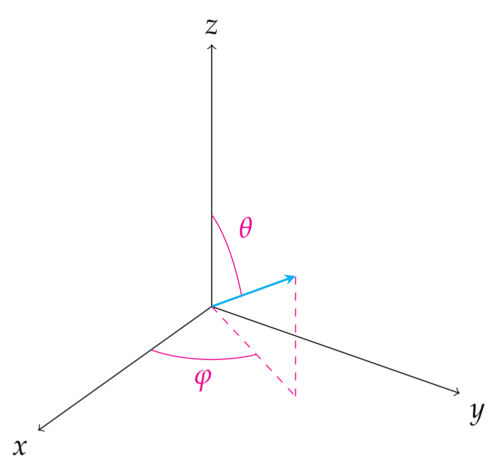
Example 2.15. The vector $\overrightarrow{(0,0,0)(a,b,c)}$ in $\mathbb{R}^3$.



Example 2.16. The vector $\overrightarrow{(a,b,c)(-1,-2,0)}$ in $\mathbb{R}^3$.

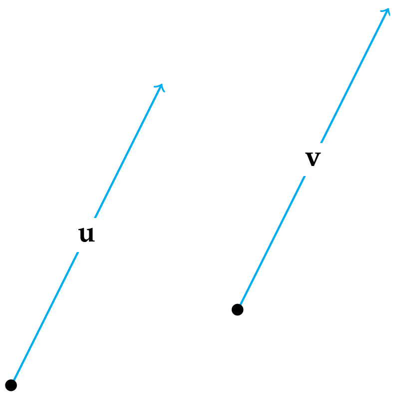


Example 2.17. The vector (r, θ, φ) in $\mathbb{R}^3$ given in *spherical coordinates*.



2.3.1 Vector Equivalence

We say the vectors $\mathbf{u}$ and $\mathbf{v}$ are equal even though they have different heads and tails.



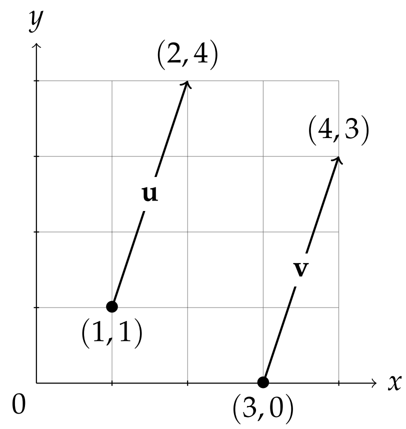
This is because they have equivalent *length* and *direction*.

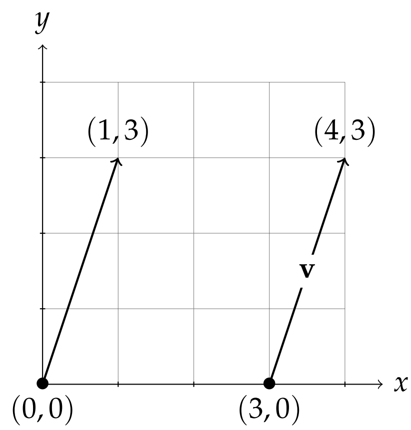
Definition 2.18 (Euclidean-vector equality). Two vectors $\mathbf{u}$ and $\mathbf{v}$ from $\mathbb{R}^n$ are *equal* when their direction and length are equal.

Proposition 2.19. Let $A, B, C, D \in \mathbb{R}^n$, $\mathbf{u} = \overrightarrow{AB}$, $\mathbf{v} = \overrightarrow{CD}$, then

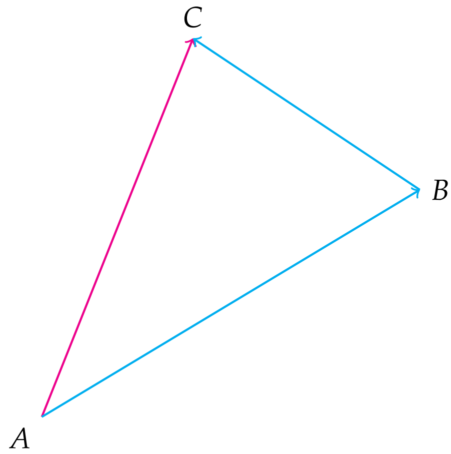
$$\mathbf{u} = \mathbf{v} \iff \overrightarrow{(A-A)(B-A)} - \overrightarrow{(C-C)(D-C)} = 0.$$

Example 2.20. Comparing the equivalent vectors $\mathbf{u}$ and $\mathbf{v}$ by shifting to the origin.

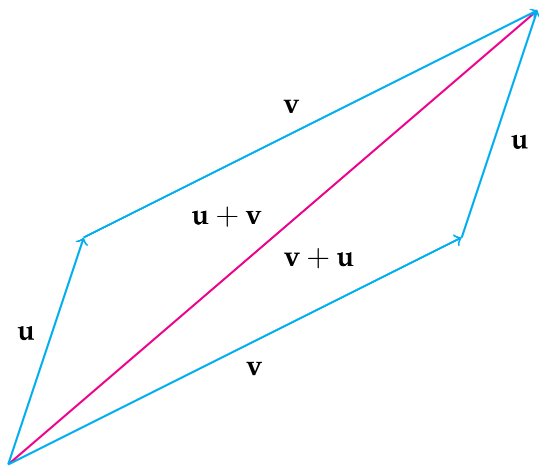




2.3.2 Combining Vectors

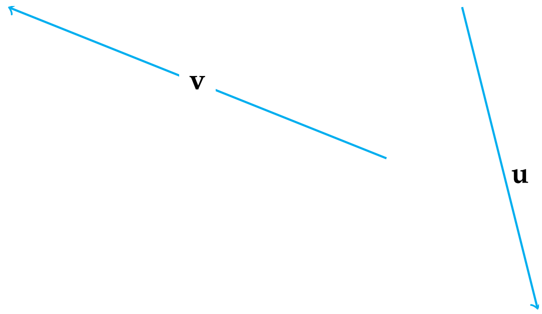


$$\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{BC}$$

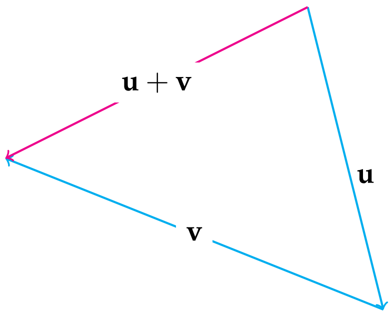


The parallelogram rule for adding vectors.

Example 2.21. The sum of



is



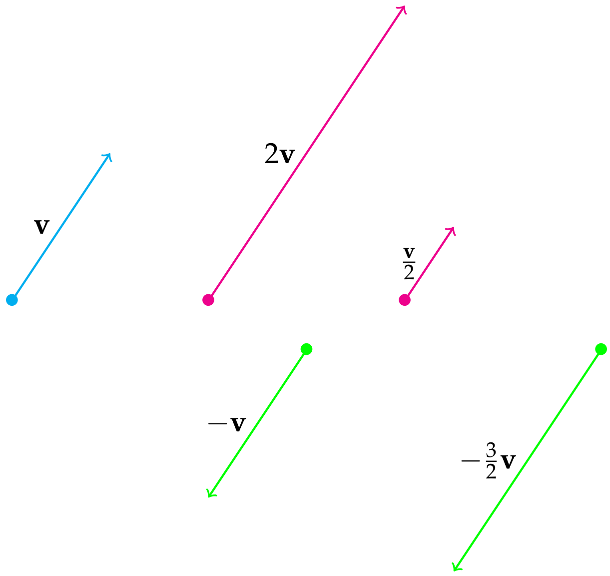
.

2.3.3 Scalars

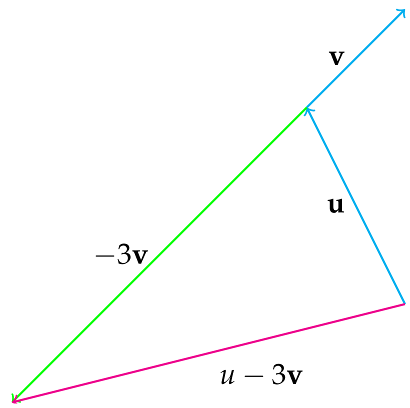
A *scalar* is not a vector, rather it is a *scaling* element which resides in $\mathbb{R}$ (or A when working in A^n).

Definition 2.22 (Scalar Multiplication)**.** If c is a scalar and $\mathbf{v}$ a vector, than the *scalar multiple* of $c\mathbf{v}$ is the vector whose length is $|c|$ times the length of $\mathbf{v}$ and whose direction is reversed if $c < 0$.

Example 2.23. Scalar multiples of the vector $\mathbf{v}$.

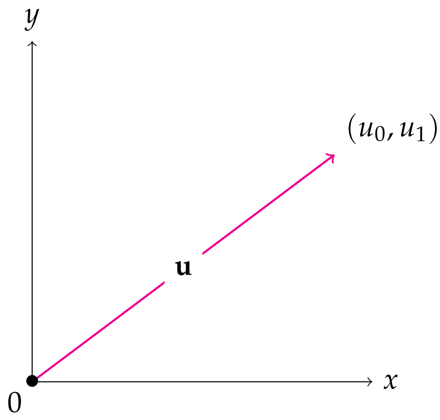


Example 2.24. Draw $\mathbf{u} - 3\mathbf{v}$ where

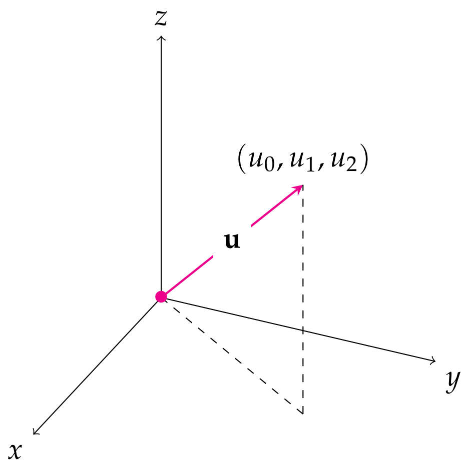


Sometimes it will be best to treat vectors algebraically. Since we saw that these vectors can always be moved to the origin while leaving their direction and length invariant, we can assume all vectors have head $\mathbf{0}$ and identify them only by their tail.

Example 2.25. The vector $\mathbf{u} = \langle u_0, u_1 \rangle$ in $\mathbb{R}^2$.



Example 2.26. The vector $\mathbf{u} = \langle u_0, u_1, u_2 \rangle$ in $\mathbb{R}^3$.



Definition 2.27. Given the two points $A = (a_0, \dots, a_{n-1})$ and $B = (b_0, \dots, b_{n-1})$ in $\mathbb{R}^n$ the vector $\mathbf{u}$ representing $\overrightarrow{AB}$ is

$$\mathbf{u} = \langle b_0 - a_0, \dots, b_{n-1} - a_{n-1} \rangle.$$

These coordinates are called the *components* of $\mathbf{u}$. (We are essentially just shifting to the origin.)

Example 2.28. The vector with head $(2, -3, 4)$ and tail $(-2, 1, 1)$ is given by

$$\langle -2 - 2, 1 - (-3), 1 - 4 \rangle = \langle -4, 4, -3 \rangle.$$

Now arithmetic becomes much easier.

Definition 2.29 (Arithmetic on Vectors)**.** If $\mathbf{u} = \langle 8_0, \dots, u_{n-1} \rangle$ and $\mathbf{v} = \langle v_0,$

$\ldots, v_{n-1}\rangle$ are vectors in $\mathbb{R}^n$ and $c \in \mathbb{R}$ a scalar then

$$\begin{aligned} \mathbf{u}+\mathbf{v} &= \langle u_0+v_0, \ldots, u_{n-1}+v_{n-1}\rangle, \\ \mathbf{u}-\mathbf{v} &= \langle u_0-v_0, \ldots, u_{n-1}-v_{n-1}\rangle, \text{ and} \\ c\mathbf{u} &= \langle cu_0, \ldots, cu_{n-1}\rangle. \end{aligned}$$

Example 2.30. Let $\mathbf{u} = \langle -2, 1, 4\rangle$ and $\mathbf{v} = \langle 7, -3, 0\rangle$ in $\mathbb{R}^3$ then

$$2\mathbf{u}-\mathbf{v} = \langle -4-7, 2-(-3), 8-0\rangle = \langle -11, 5, 8\rangle.$$

2.3.4 *Vector Norms*

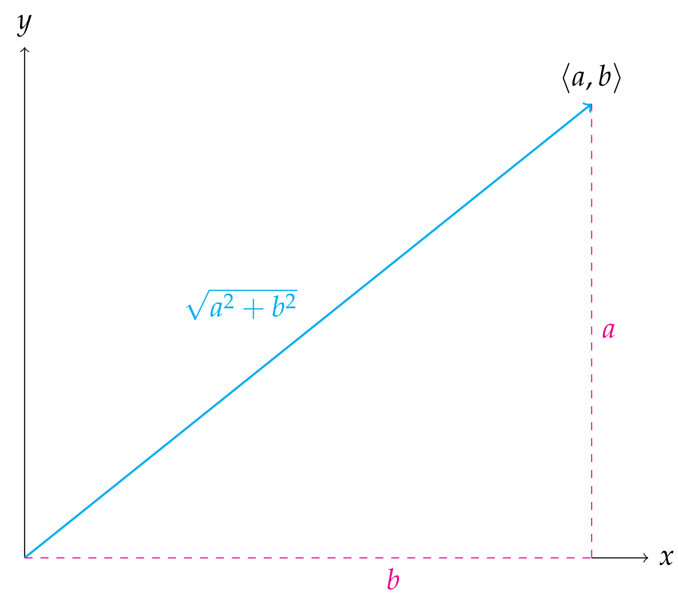
Definition 2.31 (Norm). The *magnitude* or *length* or *norm* of a vector $\mathbf{v}$ is denoted by

$$|\mathbf{v}| \quad \text{or} \quad ||\mathbf{v}||.$$

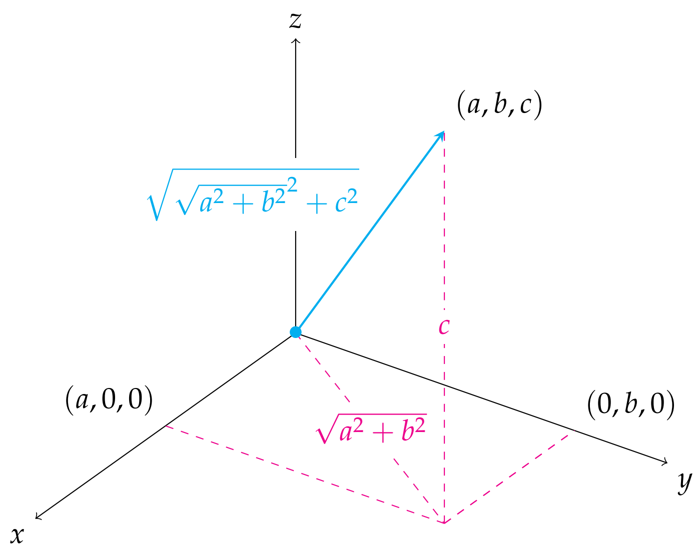
Proposition 2.32. For $\mathbf{v} = \langle v_0, \ldots, v_{n-1}\rangle \in \mathbb{R}^n$

$$||\mathbf{v}|| = \sqrt{v_0^2 + \cdots + v_{n-1}^2}.$$

Example 2.33. The norm of $\langle a, b\rangle$ in $\mathbb{R}^2$.



Example 2.34. The norm of $\langle a, b, c\rangle$ in $\mathbb{R}^3$.



Example 2.35. The norm of $\langle 4, 0, 2\rangle$ is

$$||\langle 4, 0, 3\rangle|| = \sqrt{16+0+9} = 5.$$

Proposition 2.36. The norm satisfies the following properties for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ vectors and $k \in \mathbb{R}$ a scalar:

1. $||\mathbf{v}|| = 0 \iff \mathbf{v} = 0$,
2. $\forall \mathbf{v} \in \mathbb{R}^n$; $||\mathbf{v}|| \geq 0$,
3. $||k\mathbf{v}|| = |k|||\mathbf{v}||$, and
4. $||\mathbf{u}+\mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$.

(The last one is the triangle inequality.)

2.4 Basis Vectors

There are special vectors which form a *basis* (in a strict sense) of $\mathbb{R}^n$. For $\mathbb{R}^2$ these vectors are

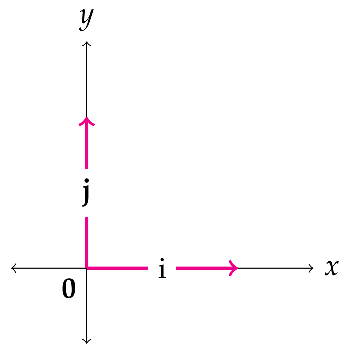
$$\mathbf{i} := \langle 1, 0\rangle \qquad \mathbf{j} := \langle 0, 1\rangle.$$

and for $\mathbb{R}^3$ they are

$$\mathbf{i} := \langle 1, 0, 0\rangle \qquad \mathbf{j} := \langle 0, 1, 0\rangle \qquad \mathbf{k} := \langle 0, 0, 1\rangle.$$

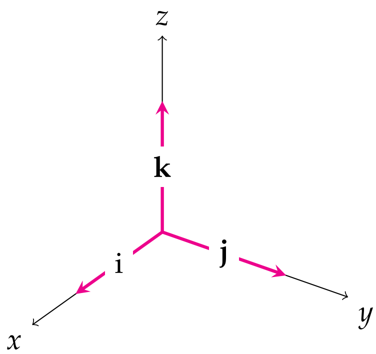
(This extends in the natural way, namely the ith basis vector has a 1 in its ith position and 0's everywhere else.)

For instance any vector $\mathbf{u} = \langle u_0, u_1 \rangle$ in $\mathbb{R}^2$ can be written as $u_0\mathrm{i} + u_1\mathbf{j}$ because



$$
\begin{aligned}
&u_0\mathrm{i} + u_1\mathbf{j} \\
&\quad = u_0\langle 1, 0, \rangle + u_1\langle 0, 1 \rangle \\
&\quad = \langle u_0, u_1 \rangle \\
&\quad = \mathbf{u}.
\end{aligned}
$$

And any vector $\mathbf{u} = \langle u_0, u_1, u_2 \rangle$ in $\mathbb{R}^3$ can be written as $u_0\mathrm{i} + u_1\mathbf{j} + u_2\mathbf{k}$ because



$$
\begin{aligned}
&u_0\mathrm{i} + u_1\mathbf{j} + u_2\mathbf{k} \\
&\quad = u_0\langle 1, 0, 0 \rangle + u_1\langle 0, 1, 0 \rangle + u_2\langle 0, 0, 1 \rangle \\
&\quad = \langle u_0, u_1, u_2 \rangle \\
&\quad = \mathbf{u}.
\end{aligned}
$$

Example 2.37. $\langle 1, -2, 6 \rangle = \mathrm{i} - 2\mathbf{j} + 6\mathbf{k}$.

2.5 Dot Product

Is it possible to multiply two vectors so that their product is a useful quantity? Yes, but it is up to us to define one.

Definition 2.38 (Dot Product). The *dot product* also called the *scalar product* of two vectors $\mathbf{a} = \langle a_0, \ldots, a_n \rangle$ and $\mathbf{b} = \langle b_0, \ldots, b_n \rangle$ is a function denoted $\mathbf{a} \cdot \mathbf{b}$ given by

$$\mathbf{a} \cdot \mathbf{b} := a_0 b_0 + \cdots + a_n b_n$$

or equivalently

$$\mathbf{a} \cdot \mathbf{b} := \sum_{i=0}^{n} a_i b_i.$$

Note the dot product $\cdot$ is a function given by

$$\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}.$$

Example 2.39.

$$\langle 2, 4 \rangle \cdot \langle 3, -1 \rangle = 2(3) + 4(-1) = 2$$

$$\langle -1, 7, 4 \rangle \cdot \langle 6, 2, -\tfrac{1}{2} \rangle = (-1)(6) + 7(2) + 4(\tfrac{1}{2}) = 6$$

$$(\mathrm{i} + 2\mathbf{j} - 3\mathbf{k}) \cdot (2\mathbf{j} - \mathbf{k}) = 1(0) + 2(2) + (-3)(-1) = 7.$$

2.5.1 Properties of the Dot Product

Proposition 2.40. If $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are vectors in $\mathbb{R}^n$ and $c \in \mathbb{R}$ is a scalar, then

$$\begin{aligned}
\mathbf{a} \cdot \mathbf{a} &= |\mathbf{a}|^2 && \\
\mathbf{a} \cdot \mathbf{b} &= \mathbf{b} \cdot \mathbf{a} && \text{Commutativity,} \\
\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) &= \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c} && \text{Distributivity,} \\
(c\mathbf{a}) \cdot \mathbf{b} &= c(\mathbf{a} \cdot \mathbf{b}) = \mathbf{a} \cdot (c\mathbf{b}) && \\
\mathbf{0} \cdot \mathbf{a} &= \langle 0, \cdots, 0 \rangle \cdot \mathbf{a} = 0 && \text{Zero.}
\end{aligned}$$

PROOF OF 1..

$$\mathbf{a} \cdot \mathbf{a} = a_0^2 + \cdots + a_n^2 = |\mathbf{a}|^2.$$

■

PROOF OF 3..

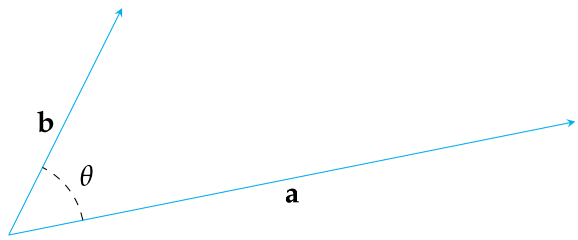
$$\begin{aligned}
&\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})\\
&\quad=\langle a_0,\ldots,a_n\rangle\cdot\langle b_0+c_0,\cdots,b_n+c_n\rangle\\
&\quad=a_0(b_0+c_0)+\cdots+a_n(b_n+c_n)\\
&\quad=a_0b_0+a_0c_0+\cdots+a_nb_n+a_nc_n\\
&\quad=(a_0b_0+\cdots a_nb_n)+(a_0c_0+\cdots+a_nc_n)\\
&\quad=\mathbf{a}\cdot\mathbf{b}+\mathbf{a}\cdot\mathbf{c}.
\end{aligned}$$

■

In physics, the following theorem is typically used as the *definition* of dot product.

Theorem 2.41. If θ is the angle between $\mathbf{a}$ and $\mathbf{b}$, then

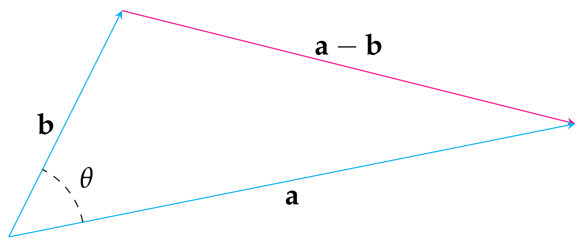
$$\mathbf{a}\cdot\mathbf{b}=|\mathbf{a}||\mathbf{b}|\cos\theta.$$



PROOF. Using the law of cosines

$$|\mathbf{a}-\mathbf{b}|^2=|\mathbf{a}|^2+|\mathbf{b}|^2-2|\mathbf{a}||\mathbf{b}|\cos\theta \tag{2.2}$$

(Observe this still works when $\theta=0$ or $\theta=\pi$ or when $\mathbf{a}=\mathbf{0}$ or $\mathbf{b}=\mathbf{0}$.



Using properties of 1, 2, and 3 of the dot product we get

$$\begin{aligned}
&|\mathbf{a}-\mathbf{b}|^2\\
&\quad=(\mathbf{a}-\mathbf{b})\cdot(\mathbf{a}-\mathbf{b})
\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{a}\cdot\mathbf{a} - \mathbf{a}\cdot\mathbf{b} - \mathbf{b}\cdot\mathbf{a} + \mathbf{b}\cdot\mathbf{b} \\
&= |\mathbf{a}|^2 - 2\mathbf{a}\cdot\mathbf{b} + |\mathbf{b}|^2 \\
&= |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2|\mathbf{a}||\mathbf{b}|\cos\theta
\end{aligned}
$$

which gives

$$
\begin{aligned}
|\mathbf{a}|^2 - 2\mathbf{a}\cdot\mathbf{b} + |\mathbf{b}|^2 &= |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2|\mathbf{a}||\mathbf{b}|\cos\theta \\
\implies -2\mathbf{a}\cdot\mathbf{b} &= -2|\mathbf{a}||\mathbf{b}|\cos\theta \\
\implies \mathbf{a}\cdot\mathbf{b} &= |\mathbf{a}||\mathbf{b}|\cos\theta
\end{aligned}
$$

■

Example 2.42. If vectors $\mathbf{a}$ and $\mathbf{b}$ have lengths 4 and 6 with angle $\pi/3$ between them, what is $\mathbf{a}\cdot\mathbf{b}$?

ANSWER. $\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos(\frac{\pi}{3}) = (4)(6)(\frac{1}{2}) = 12.$ ◆

Corollary 2.43. [1] If θ is the angle between two nonzero vectors $\mathbf{a}$ and $\mathbf{b}$, then

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}||\mathbf{b}|}.$$

Question 2.44. What is the angle between $\mathbf{a} = \langle 2, 1\rangle$ and $\mathbf{b} = \langle 1, 3\rangle$?

ANSWER. We have

$$|\mathbf{a}| = \sqrt{2^2 + 1^2} = \sqrt{5} \qquad \text{and} \qquad |\mathbf{b}| = \sqrt{1^2 + 3^2} = \sqrt{10}$$

and $\mathbf{a}\cdot\mathbf{b} = 2(1) + 1(3) = 5$. So it follows from the corollary that

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}||\mathbf{b}|} = \frac{5}{5\sqrt{2}} = \frac{1}{\sqrt{2}}$$

and thus $\theta = \arccos\frac{1}{\sqrt{2}} = \frac{\pi}{4}$. ◆

Question 2.45. What is the angle between $\mathbf{a} = \langle 2, 2, -1\rangle$ and $\mathbf{b} = \langle 5, -3, 2\rangle$?

ANSWER. We have

$$|\mathbf{a}| = \sqrt{2^2 + 2^2 + (-1)^2} = 3 \quad \text{and} \quad |\mathbf{b}| = \sqrt{5^2 + (-3)^2 + 2^2} = \sqrt{38}$$

and $\mathbf{a}\cdot\mathbf{b} = 2(5) + 2(-3) + (-1)(2) = 2$. So it follows from the corollary that

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}||\mathbf{b}|} = \frac{2}{3\sqrt{38}}$$

[1] A corollary is something that follows immediately from a theorem or proposition.

and thus $\theta = \arccos \frac{2}{3\sqrt{38}}$. ◆

Definition 2.46 (Orthogonal). Two vectors **a** and **b** are called *orthogonal* or *perpendicular* if the angle between them is $\frac{\pi}{2}$.

Proposition 2.47. For **a** and **b** vectors from $\mathbb{R}^n$

$$\mathbf{a} \cdot \mathbf{b} = 0 \iff \mathbf{a} \perp \mathbf{b}$$

PROOF.

$$\mathbf{a} \cdot \mathbf{b} = 0 \iff |\mathbf{a}||\mathbf{b}| \cos\theta = 0 \iff \arccos\theta = 0 \iff \theta = \frac{\pi}{2}.$$

■

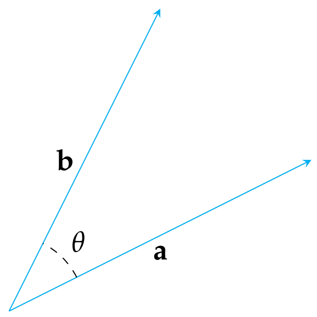
Question 2.48. Show $2\mathrm{i} + 2\mathbf{j} - \mathbf{k}$ is perpendicular to $5\mathrm{i} - 4\mathbf{j} + 2\mathbf{k}$.

ANSWER. $\langle 2, 2, -1\rangle \cdot \langle 5, -4, 2\rangle = 2(5) + 2(-4) + (-1)(2) = 0$ and thereby the vectors are perpendicular. ◆

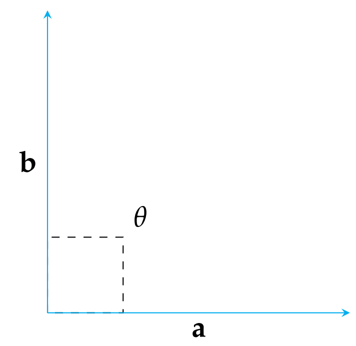
Notice that

$$\cos\theta > 0 \text{ when } \theta \in [0, \tfrac{\pi}{2}), \text{ and}$$
$$\cos\theta < 0 \text{ when } \theta \in (\tfrac{\pi}{2}, \pi].$$

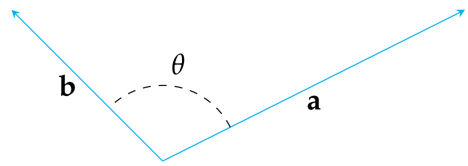
So the dot product can be thought of as a measure of the extent to which two vectors are pointing in the same direction.



$\mathbf{a} \cdot \mathbf{b} > 0$ when θ is acute.



$\mathbf{a}\cdot\mathbf{b} = 0$ when $\theta = \frac{\pi}{2}$.



$\mathbf{a}\cdot\mathbf{b} < 0$ when θ is obtuse.

2.6 *Projections*

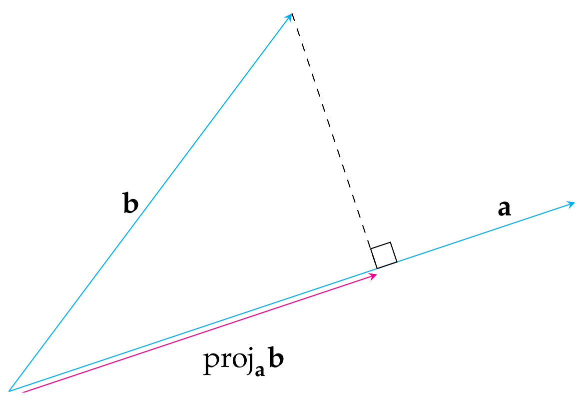
Definition 2.49 (Unit Direction Vector)**.** The *unit direction vector* of $\mathbf{a}$ is the vector pointing in the same direction that has unit length. It is denoted $\hat{\mathbf{a}}$ and given by

$$\hat{\mathbf{a}} := \frac{\mathbf{a}}{|\mathbf{a}|}.$$

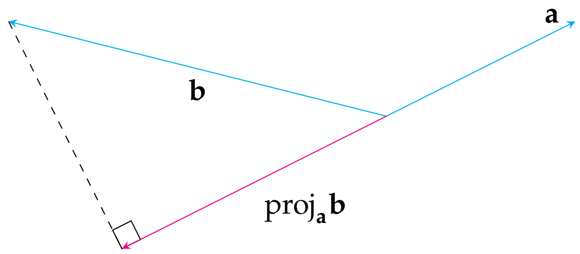
Definition 2.50 (Vector Projection)**.** The *vector projection* of $\mathbf{b}$ onto $\mathbf{a}$ is denoted $\text{proj}_{\mathbf{a}}\mathbf{b}$ and *is the vector* given by

$$\text{proj}_{\mathbf{a}}\mathbf{b} := \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|}\hat{\mathbf{a}}.$$

Example 2.51.

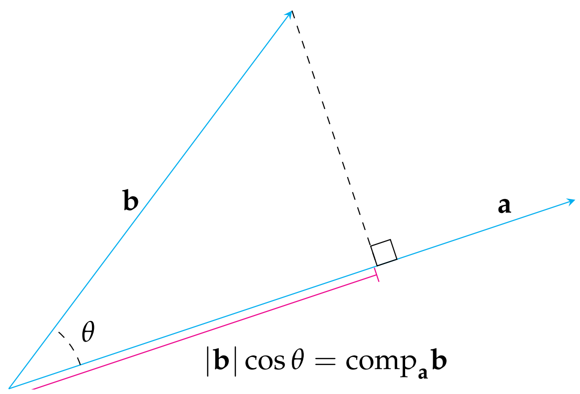


Example 2.52.



Definition 2.53 (Scalar Projection). The *scalar projection* of $\mathbf{b}$ onto $\mathbf{a}$, also called the *component of* $\mathbf{b}$ *along* $\mathbf{a}$, is denoted by $\text{comp}_\mathbf{a}\mathbf{b}$ and is the signed magnitude (i.e. the magnitude can be negative) of the vector projection of $\mathbf{b}$ onto $\mathbf{a}$:

$$\text{comp}_\mathbf{a}\mathbf{b} := \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|}.$$



Proposition 2.54. For vectors $\mathbf{a}$ and $\mathbf{b}$ we have

$$\text{proj}_\mathbf{a}\mathbf{b} = (\text{comp}_\mathbf{a}\mathbf{b})\,\hat{\mathbf{a}}.$$

PROOF. Follows directly from definitions. ■

Question 2.55. What is the scalar projection of $\mathbf{b} = \langle 1, 1, 2\rangle$ onto $\mathbf{a} = \langle -2, 3, 1\rangle$?

ANSWER.

$$\begin{aligned}\text{comp}_\mathbf{a}\mathbf{b} &= \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|}\\ &= \frac{(-2)(1)+3(1)+1(2)}{\sqrt{(-2)^2+3^2+1^2}}\\ &= \frac{3}{\sqrt{14}}\end{aligned}$$

◆

Question 2.56. What is the vector projection of $\mathbf{b} = \langle 1, 1, 2 \rangle$ onto $\mathbf{a} = \langle -2, 3, 1 \rangle$?

ANSWER.

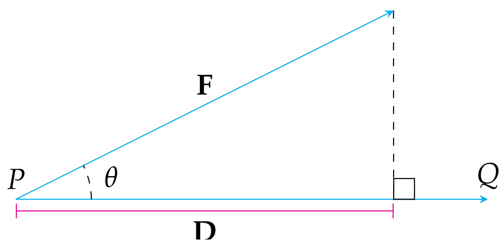
$$\begin{aligned} \operatorname{proj}_{\mathbf{a}} \mathbf{b} &= (\operatorname{comp}_{\mathbf{a}} \mathbf{b})\, \hat{\mathbf{a}} \\ &= \frac{3}{\sqrt{14}} \frac{\langle -2, 3, 1 \rangle}{\sqrt{(-2)^2 + 3^2 + 1}} \\ &= \frac{3}{\sqrt{14}} \frac{\langle -2, 3, 1 \rangle}{\sqrt{14}} \\ &= \left\langle -\frac{3}{7}, \frac{9}{14}, \frac{3}{14} \right\rangle \end{aligned}$$

◆

2.6.1 Application

Definition 2.57 (Work). If a *force* $\mathbf{F}$ moves an object from P to Q so that it is *displaced* by $\mathbf{D}$ then the *work*, W, done by this force is given by

$$W = \mathbf{F} \cdot \mathbf{D}.$$



Question 2.58. What is the work done pulling a wagon 100 m along a horizontal path with 70N of force when the angle of the wagons handle is $\frac{\pi}{4}$?

ANSWER. $W = (70)(100) \cos \frac{\pi}{4} = 3500\sqrt{2} \approx 4949.75J$. ◆

2.7 Vector Product

Given two (nonzero) vectors $\mathbf{a} = \langle a_0, a_1, a_2 \rangle$ and $\mathbf{b} = \langle b_0, b_1, b_2 \rangle$ what is the vector $\mathbf{c} = \langle c_0, c_1, c_2 \rangle$ perpendicular to *both* $\mathbf{a}$ and $\mathbf{b}$?

We need to solve $\mathbf{a} \cdot \mathbf{c} = \mathbf{b} \cdot \mathbf{c} = 0$ for $\mathbf{c}$:

$$a_0 c_0 + a_1 c_1 + a_2 c_2 = 0, \tag{2.3}$$

$$b_0c_0 + b_1c_1 + b_2c_2 = 0. \tag{2.4}$$

First, multiply (2.3) by b_2 and (2.4) by a_2 and subtract to eliminate c_2

$$\begin{aligned}(a_0c_0 + a_1c_1 + a_2c_2)b_2 &- (b_0c_0 + b_1c_1 + b_2c_2)a_2 \\ &= a_0b_2c_0 + a_1b_2c_1 + a_2b_2c_2 - a_2b_0c_0 - a_2b_1c_1 - a_2b_2c_2 \\ &= (a_0b_2 - a_2b_0)c_0 + (a_1b_2 - a_2b_1)c_1 \qquad (2.5) \\ &= 0\end{aligned}$$

and then notice (2.5) has form $pc_0 + qc_1$ which has obvious solution $c_0 = q$ and $c_1 = p$. Thus:

$$c_0 = a_1b_2 - a_2b_1 \qquad\qquad c_1 = a_2b_0 - a_0b_2$$

and thereby $c_2 = a_0b_1 - a_1b_0$.

So the vector $\mathbf{c}$ perpendicular to both $\mathbf{a}$ and $\mathbf{b}$ is

$$\mathbf{c} = \langle a_1b_2 - a_2b_1,\ a_2b_0 - a_0b_2,\ a_0b_1 - a_1b_0\rangle.$$

Definition 2.59 (Cross Product). The *cross product* or *vector product* of $\mathbf{a} = \langle a_0, a_1, a_2\rangle$ and $\mathbf{b} = \langle b_0, b_1, b_2\rangle$, denoted $\mathbf{a}\times\mathbf{b}$ is given by

$$\mathbf{a}\times\mathbf{b} := \langle a_1b_2 - a_2b_1,\ a_2b_0 - a_0b_2,\ a_0b_1 - a_1b_1\rangle.$$

Question 2.60. What is the vector perpendicular to i and $\mathbf{j}$?

ANSWER.

$$\begin{aligned}\mathrm{i}\times\mathbf{j} &= \langle 1,0,0\rangle \times \langle 0,1,0\rangle \\ &= (0)(1) - (0)(0),\ (0)(1) - (1)(0),\ (1)(1) - (0)(1) \\ &= \langle 0,0,1\rangle\end{aligned}$$

◆

The cross product was discovered as a side-effect of Hamilton's efforts to generalize $\mathbb{C}$ to higher dimensions. Recall $\mathbb{H}$, the Quaternions, had the form

$$\underbrace{a}_{\text{scalar part}} + \underbrace{b\mathrm{i} + c\mathbf{j} + d\mathbf{k}}_{\text{vector part}}.$$

Also recall

$$\mathrm{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathrm{i}\mathbf{j}\mathbf{k} = -1$$

and let us multiply two quaternions with *zero scalar part*. Notice that the dot product and scalar product show up naturally when we do this:

$$\begin{aligned}&(b\mathbf{i}+c\mathbf{j}+d\mathbf{k})(f\mathbf{i}+g\mathbf{j}+h\mathbf{k})\\&\quad=-bf+bg\mathbf{k}-bh\mathbf{j}-cf\mathbf{k}-cg+ch\mathbf{i}+df\mathbf{j}-dg\mathbf{i}-dh\\&\quad=\underbrace{(bf+cg+dh)}_{\text{dot product}}+\underbrace{(ch-dg)\mathbf{i}+(df-bh)\mathbf{j}+(bg-cf)\mathbf{k}}_{\text{cross product}}.\end{aligned}$$

Definition 2.61 (Determinant)**.** The *determinant* of *order two* is a *matrix operation* defined by

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

and the determinant of *order three* is defined by

$$\begin{vmatrix} a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 \\ c_0 & c_1 & c_2 \end{vmatrix} = a_0 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} - a_1 \begin{vmatrix} b_0 & b_2 \\ c_0 & c_2 \end{vmatrix} + a_2 \begin{vmatrix} b_0 & b_1 \\ c_0 & c_1 \end{vmatrix}.$$

Example 2.62.

$$\begin{aligned}\begin{vmatrix} 1 & 2 & -1 \\ 3 & 0 & 1 \\ -5 & 4 & 2 \end{vmatrix} &= 1 \begin{vmatrix} 0 & 1 \\ 4 & 2 \end{vmatrix} - 2 \begin{vmatrix} 3 & 1 \\ -5 & 2 \end{vmatrix} + (-1) \begin{vmatrix} 3 & 0 \\ -5 & 4 \end{vmatrix}\\ &= 1(0-4) - 2(6+5) + (-1)(12-0)\\ &= -38.\end{aligned}$$

2.7.1 *Alternate Forms for Cross Product*

Proposition 2.63. For vectors $\mathbf{a} = \langle a_0, a_1, a_2 \rangle$ and $\mathbf{b} = \langle b_0, b_1, b_2 \rangle$ we have

$$\begin{aligned}\mathbf{a} \times \mathbf{b} &= \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \mathbf{i} - \begin{vmatrix} a_0 & a_2 \\ b_0 & b_2 \end{vmatrix} \mathbf{j} + \begin{vmatrix} a_0 & a_1 \\ b_0 & b_1 \end{vmatrix} \mathbf{k}\\ &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 \end{vmatrix}.\end{aligned}$$

Example 2.64. Let $\mathbf{a} = \langle 1, 3, 4 \rangle$ and $\mathbf{b} = \langle 2, 7, -5 \rangle$, then

$$\begin{aligned} \mathbf{a} \times \mathbf{b} &= \begin{vmatrix} \mathrm{i} & \mathbf{j} & \mathbf{k} \\ 1 & 3 & 4 \\ 2 & 7 & -5 \end{vmatrix} \\ &= \begin{vmatrix} 3 & 4 \\ 7 & -5 \end{vmatrix} \mathrm{i} - \begin{vmatrix} 1 & 4 \\ 2 & -5 \end{vmatrix} \mathbf{j} + \begin{vmatrix} 1 & 3 \\ 2 & 7 \end{vmatrix} \mathbf{k} \\ &= -43\mathrm{i} - 13\mathbf{j} + \mathbf{k} \end{aligned}$$

Proposition 2.65. For any vector $\mathbf{a} = \langle a_0, a_1, a_2 \rangle \in \mathbb{R}^3$

$$\mathbf{a} \times \mathbf{a} = \mathbf{0}.$$

PROOF.

$$\begin{aligned} \mathbf{a} \times \mathbf{a} &= \begin{vmatrix} \mathrm{i} & \mathbf{j} & \mathbf{k} \\ a_0 & a_1 & a_2 \\ a_0 & a_1 & a_2 \end{vmatrix} \\ &= (a_1 a_2 - a_2 a_1)\,\mathrm{i} - (a_0 a_2 - a_2 a_0)\,\mathbf{j} + (a_0 a_1 - a_1 a_0)\,\mathbf{k} \\ &= 0\mathrm{i} - 0\mathbf{j} + 0\mathbf{k}. \end{aligned}$$

■

Theorem 2.66. For any vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$

$$\mathbf{a} \times \mathbf{b} \perp \mathbf{a} \text{ and } \mathbf{a} \times \mathbf{b} \perp \mathbf{b}.$$

PROOF.

$$\begin{aligned} (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{a} &= \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} a_0 - \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} a_1 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} a_2 \\ &= a_0(a_1 b_2 - a_2 b_1) - a_1(a_1 b_2 - a_2 b_1) - a_2(a_1 b_2 - a_2 b_1) \\ &= 0. \end{aligned}$$

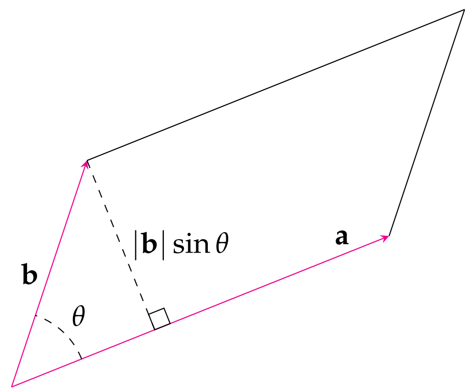
The same argument works for $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{b}$. ■

Theorem 2.67. If $\theta \in [0, \pi]$ is the angle between $\mathbf{a}$ and $\mathbf{b}$, two vectors from

$\mathbb{R}^3$, then

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin\theta.$$

Or, equivalently, $|\mathbf{a} \times \mathbf{b}|$ is equal to the area of the parallelogram determined by $\mathbf{a}$ and $\mathbf{b}$.



PROOF. (Here the dots represent an expansion and regrouping that we omit for brevity.)

$$\begin{aligned} |\mathbf{a} \times \mathbf{b}|^2 &= \cdots \\ &\vdots \\ &= |\mathbf{a}|^2|\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2 \\ &= |\mathbf{a}|^2|\mathbf{b}|^2 - |\mathbf{a}|^2|\mathbf{b}|^2 \cos^2\theta && \text{We will prove this.} \\ &= |\mathbf{a}|^2|\mathbf{b}|^2(1 - \cos^2\theta) \\ &= |\mathbf{a}|^2|\mathbf{b}|^2(\sin^2\theta) \end{aligned}$$

Thereby $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin\theta$. ■

Question 2.68. Find the vector perpendicular to the plane that passes through the points $P = (1, 4, 6)$, $Q = (-2, 5, -1)$, and $R = (1, -1, 1)$.

ANSWER. We need only find the vector perpendicular to $\overrightarrow{PQ}$ and $\overrightarrow{PR}$.

$$\overrightarrow{PQ} = \langle -3, 1, -7 \rangle \qquad \overrightarrow{PR} = \langle 0, -5, -5 \rangle$$

which is given by

$$\overrightarrow{PQ} \times \overrightarrow{PR} = \begin{vmatrix} \mathrm{i} & \mathbf{j} & \mathbf{k} \\ -3 & 1 & -7 \\ 0 & -5 & -5 \end{vmatrix} = -40\mathrm{i} - 15\mathbf{j} + 15\mathbf{k}.$$

◆

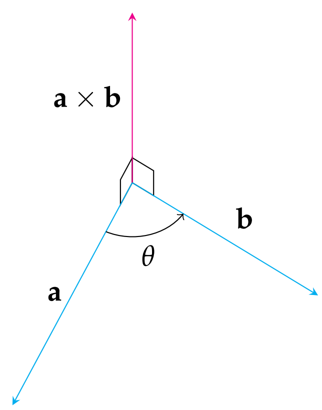
Question 2.69. Find the area of the triangle that passes through the points $P = (1, 4, 6)$, $Q = (-2, 5, -1)$, and $R = (1, -1, 1)$.

ANSWER. We already know $\overrightarrow{PQ} \times \overrightarrow{PR} = \langle -40, -15, 15 \rangle$. The area of the parallelogram formed by PQ and PR is the length of this cross product:

$$|\langle -40, -15, 15 \rangle| = 5\sqrt{82}.$$

The area of the triangle PQR is half of this: $\frac{5}{2}\sqrt{82}$. ◆

Proposition 2.70. $\mathbf{a} \times \mathbf{b}$ is the vector to perpendicular *both* to $\mathbf{a}$ and $\mathbf{b}$ whose direction is determined by "*the right hand rule*" and whose magnitude is $|\mathbf{a}||\mathbf{b}| \sin \theta$. (This is how physicists *define* $\mathbf{a} \times \mathbf{b}$.)



Corollary 2.71. Two nonzero vectors $\mathbf{a}$ and $\mathbf{b}$ are parallel, denoted $\mathbf{a} \parallel \mathbf{b}$, when their cross product is zero:

$$\mathbf{a} \times \mathbf{b} = \mathbf{0} \iff \mathbf{a} \parallel \mathbf{b}.$$

PROOF. Two nonzero vectors $\mathbf{a}$ and $\mathbf{b}$ are parallel only when θ, the angle between them, is 0 or π. In either case $\sin \theta = 0$ so $|\mathbf{a} \times \mathbf{b}| = 0$ and therefore $\mathbf{a} \times \mathbf{b} = \mathbf{0}$ (only the zero vector has magnitude zero). ■

2.7.2 Properties of the Cross Product

Question 2.72. Is the cross product commutative?

ANSWER. No! Consider $\mathbf{i} = \langle 1, 0, 0 \rangle$ and $\mathbf{j} = \langle 0, 1, 0 \rangle$.

$$\mathbf{i} \times \mathbf{j} = \langle 0, 0, 1 \rangle = \mathbf{k} \qquad \mathbf{j} \times \mathbf{i} = \langle 0, 0, -1 \rangle = -\mathbf{k}.$$

◆

Question 2.73. Is the cross product associative?

ANSWER. No! Consider $\mathbf{i} = \langle 1, 0, 0 \rangle$, $\mathbf{j} = \langle 0, 1, 0 \rangle$, $\mathbf{k} = \langle 0, 0, 1 \rangle$.

$$\mathbf{i} \times (\mathbf{i} \times \mathbf{j}) = \mathbf{i} \times \mathbf{k} = -\mathbf{j}$$

whereas

$$(\mathbf{i} \times \mathbf{i}) \times \mathbf{j} = \mathbf{0} \times \mathbf{j} = \mathbf{0}.$$

◆

Proposition 2.74. If $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are vectors in $\mathbb{R}^3$ and $c \in \mathbb{R}$ is a scalar then

1. $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$,
2. $(c\mathbf{a}) \times \mathbf{b} = c(\mathbf{a} \times \mathbf{b}) = \mathbf{a} \times (c\mathbf{b})$,
3. $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$,
4. $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$,
5. $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$,
6. $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$.

PROOF. Exercise. ■

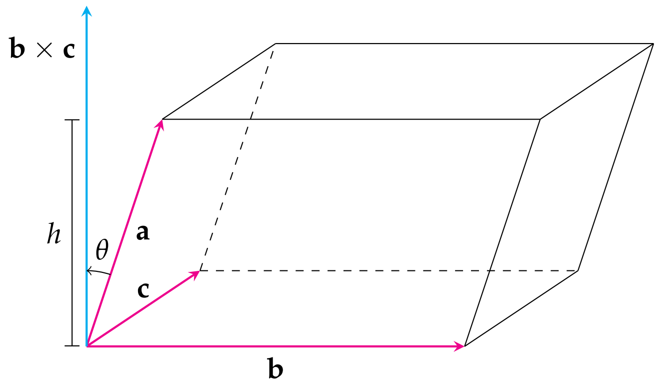
Definition 2.75 (Scalar triple product)**.** The product $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ is called the *triple scalar product* of the vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$.

Proposition 2.76.

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 \\ c_0 & c_1 & c_2 \end{vmatrix}.$$

PROOF. Exercise. ■

The scalar triple product determines the volume of the *parallelepiped* determined by **a**, **b**, and **c**.



$$V = |\mathbf{b} \times \mathbf{c}||\mathbf{a}||\cos\theta| = |\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|.$$

Question 2.77. Show the vectors $\mathbf{a} = \langle 1, 4, -7 \rangle$, $\mathbf{b} = \langle 2, -1, 4 \rangle$, and $\mathbf{c} = \langle 0, -9, 18 \rangle$ are coplanar.

ANSWER.

$$\begin{aligned} \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) &= \begin{vmatrix} 1 & 4 & -7 \\ 2 & -1 & 4 \\ 0 & -9 & 18 \end{vmatrix} \\ &= 1 \begin{vmatrix} -1 & 4 \\ -9 & 18 \end{vmatrix} - 4 \begin{vmatrix} 2 & 4 \\ 0 & 18 \end{vmatrix} - 7 \begin{vmatrix} 2 & 4 \\ 0 & 18 \end{vmatrix} \\ &= 0. \end{aligned}$$

Thereby the volume of the parallelepiped is 0 and this can only be the case when **a**, **b**, and **c** are coplanar. ◆

2.7.3 *Application*

Definition 2.78 (Torque). The *torque* $\boldsymbol{\tau}$ (relative to the origin) is defined to be the cross product of a force vector and position vector **r**:

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}.$$

3 Lines, Planes, and Hyperplanes

3.1 Lines

3.1.1 Parametric Form

Definition 3.1 (Vector Equations of a Line)**.** Let $\mathbf{r}_0$ (position) and $\mathbf{v}$ (direction/slope) be vectors of $\mathbb{R}^n$ and $t \in \mathbb{R}$ a scalar. Then

$$\mathbf{r}(t) = \mathbf{r}_0 + t\mathbf{v}$$

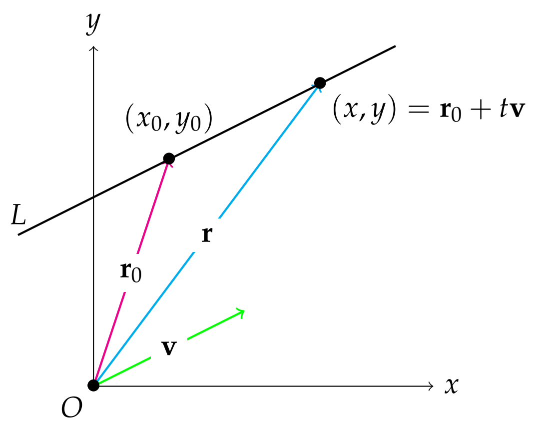
is the *vector equation* of a line.

This is also called the *parametric form* of the line because the line's points are parameterized by t.

Definition 3.2 (Direction Numbers)**.** When $\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}$ the components of the line $\mathbf{v}$ are called the *direction numbers* of L.

Note *any* vector parallel to $\mathbf{v}$ could be used to define the same line.

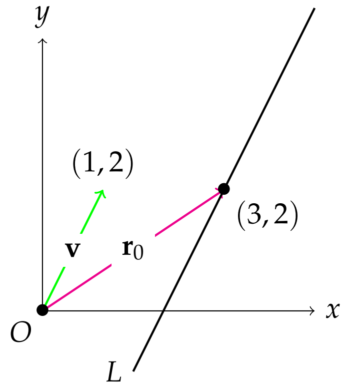
The line L as described by two vectors (i.e. a point $\mathbf{r}_0$ and direction $\mathbf{v}$).



$$\mathbf{r}(t) = \mathbf{r}_0 + t\mathbf{v}$$

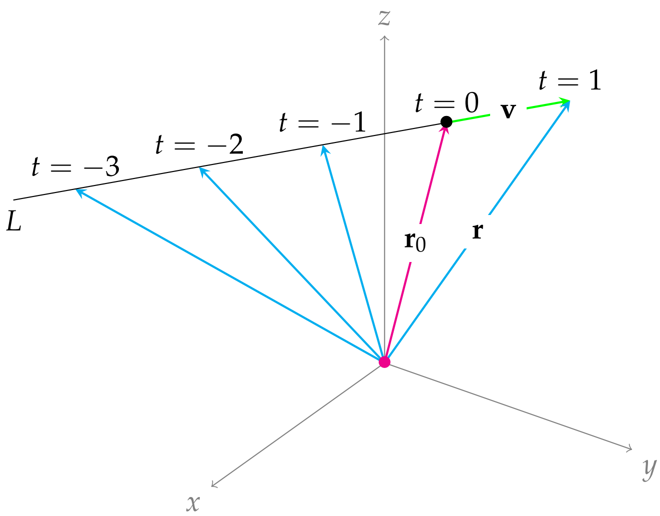
Question 3.3. What is the parametric equation for the line with slope 2 going through $(3,2)$?

ANSWER. $\mathbf{r} = \langle 3,2\rangle + t\langle 1,2\rangle$



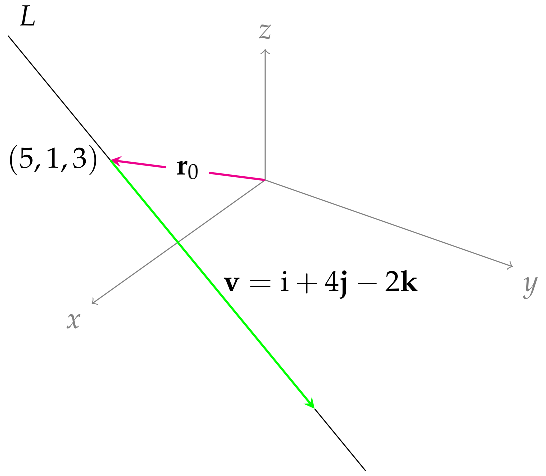
◆

The line L as described by two vectors (i.e. a point $\mathbf{r}_0$ and direction $\mathbf{v}$) in $\mathbb{R}^3$.



$$\mathbf{r}(t) = \mathbf{r}_0 + t\mathbf{v}$$

Question 3.4. Find a vector equation and parametric equations for the line that passes through the point $(5,1,3)$ and is parallel to the vector $\mathbf{i} + 4\mathbf{j} - 2\mathbf{k}$. Find two other points on the line.



ANSWER. Here $\mathbf{r}_0 = \langle 5, 1, 3\rangle$ and $v = \mathrm{i} + 4\mathbf{j} - 2\mathbf{k} = \langle 1, 4, -2\rangle$.

$$\begin{aligned}\mathbf{r} &= \mathbf{r}_0 + t\mathbf{v}\\ &= \langle 5, 1, 3\rangle + t\langle 1, 4, -2\rangle\end{aligned}$$

Two other points are given by $t = -1$ and $t = 1$:

$$\langle 4, -3, 5\rangle \qquad \langle 6, 5, 1\rangle.$$

◆

3.1.2 Symmetric Form

Notice $\mathbf{r} = \langle 5, 1, 3\rangle + t\langle 1, 4, -2\rangle$ is equivalent to

$$\langle x, y, z\rangle = \langle 5, 1, 3\rangle + t\langle 1, 4, -2\rangle$$

which means we have the set of equations

$$x = 5 + t \qquad y = 1 + 4t \qquad z = 3 - 2t.$$

Solving for t gives

$$t = \frac{x-5}{1} \qquad t = \frac{y-1}{4} \qquad t = \frac{z-3}{-2}.$$

and therefore

$$\frac{x-5}{1} = \frac{y-1}{4} = \frac{z-3}{-2}$$

is another description of the line.

Generally, in three space, when $a, b, c \neq 0$,

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}$$

defines a line through (x_0, y_0, z_0) with slope $\langle a, b, c \rangle$.

Definition 3.5 (Symmetric Form of a Line)**.** In $\mathbb{R}^n$ when $\mathbf{x} = \langle x_0, \ldots, x_{n-1} \rangle$ is a vector-valued variable, $\mathbf{p} = \langle p_0, \ldots, p_{n-1} \rangle$ is a fixed, and $\mathbf{a} = \langle a_0, \ldots, a_{n-1} \rangle \in (\mathbb{R} \setminus \{0\})^n$ then

$$\frac{x_0 - p_0}{a_0} = \cdots = \frac{x_{n-1} - p_{n-1}}{a_{n-1}}$$

defines a line through $\mathbf{p}$ with direction $\mathbf{a}$.

Question 3.6. Find the *parametric* and *symmetric* equations of the line through $(2, 4, -3)$ and $(3, -1, 1)$. Where does this line intersect the xy-plane?

ANSWER. We are not explicitly given a direction vector but notice

$$\mathbf{v} = \langle 3, -1, 1 \rangle - \langle 2, 4, -3 \rangle = \langle 1, -5, 4 \rangle$$

is the direction of the line. We need only pick either $(2, 4, -3)$ or $(3, -1, 1)$ as $\mathbf{r}_0$. Therefore the *parametric* equation of the line is given by

$$\langle x, y, z \rangle = \langle 2, 4, -3 \rangle + t \langle 1, -5, 4 \rangle$$

for $t \in \mathbb{R}$ a parameter. The *symmetric* equation is

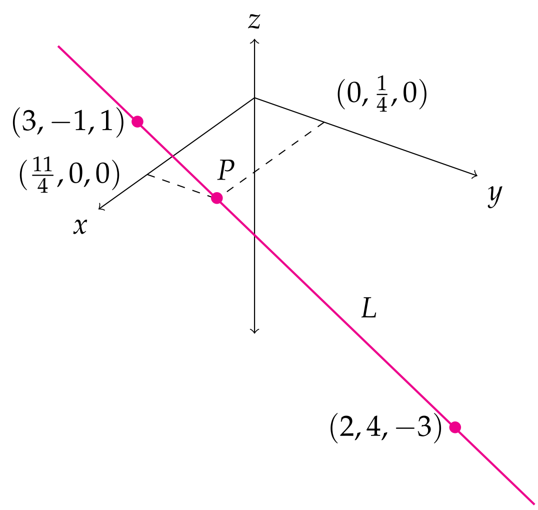
$$\frac{x - 2}{1} = \frac{y - 4}{-5} = \frac{z + 3}{4}.$$

and thus when in the xy-plane where $z = 0$, x and y are given by

$$\frac{x - 2}{1} = \frac{y - 4}{-5} = \frac{3}{4}$$

which implies $x = \frac{11}{4}$ and $y = \frac{1}{4}$.

The line intersects the xy axis when $z = 0$.

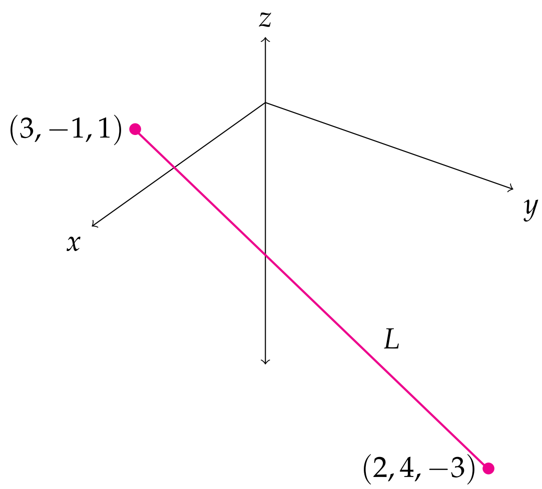


◆

3.1.3 Line Segment

We can also use parameterized curves to describe *line segments*:

$$\mathbf{r}(t) = \langle 2 + t,\ 4 - 5t,\ -3 + 4t \rangle \qquad t \in [0, 1]$$



Proposition 3.7. The line through the (tail of the) vectors $\mathbf{r}_0$ and $\mathbf{r}_1$ is given by

$$\mathbf{r}(t) = \mathbf{r}_0 + t(\mathbf{r}_1 - \mathbf{r}_0)$$

where the line segment given by $\mathbf{r}_0$ and $\mathbf{r}_1$ is in the interval $t \in [0, 1]$.

3.1.4 Skew Lines

Definition 3.8 (Skew)**.** Two lines L_0 and L_1 are skew when they *do not intersect* and *are not parallel*.

Example 3.9. Show that the lines L_0 (parameterized by t) and L_1 (parameterized by s) with the parametric equations

$$x = 1 + t \qquad y = -2 + 3t \qquad z = 4 - t$$
$$x = 2s \qquad y = 3 + s \qquad z = -3 + 4s$$

are skew.

ANSWER. The corresponding direction vectors for L_0 and L_1

$$\langle 1, 3, -1 \rangle \qquad \langle 2, 1, 4 \rangle$$

are not scalar multiples of one another — thus the lines cannot be parallel.

It remains to show the lines do not intersect. Towards a contradiction, suppose the lines *do* have a point of intersection given by

$$\begin{aligned} 1 + t &= 2s \\ -2 + 3t &= 3 + s \\ 4 - t &= -3 + 4s \end{aligned}$$

Notice substituting the second ($s = -5 + 3t$) into the first gives

$$1 + t = -10 + 6t \implies 5t = 11 \implies t = \frac{11}{5}$$

which means $s = \frac{1+t}{2} = \frac{8}{5}$. This implies, by the third equation, that

$$4 - \frac{11}{5} = 2\frac{8}{5} \implies \frac{-39}{20} = 0 \ ↯$$

◆

3.2 Planes

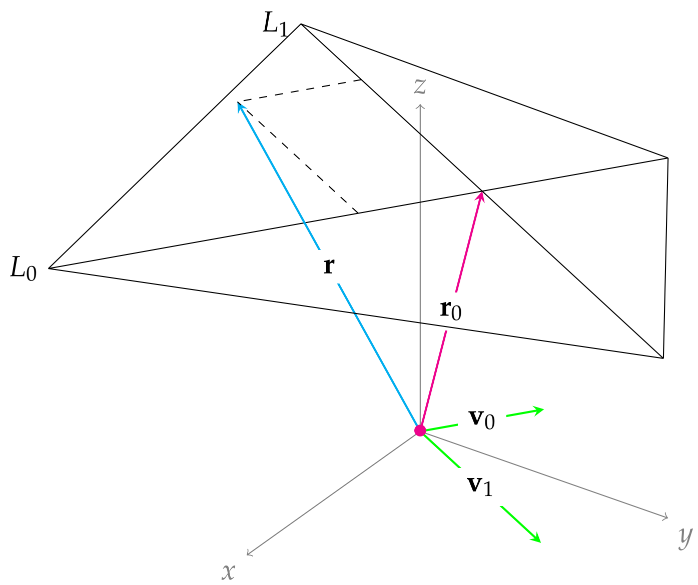
A plane is a surface defined by three points.

3.2.1 Parametric Form

Definition 3.10 (Parametric Equation of Plane)**.** Let $\mathbf{r}_0$ (position) and $\mathbf{v}_0, \mathbf{v}_1$ (direction) be vectors of $\mathbb{R}^n$ and $s, t \in \mathbb{R}$ be scalars. Then

$$\mathbf{r}(s,t) = \mathbf{r}_0 + s\mathbf{v}_0 + t\mathbf{v}_1$$

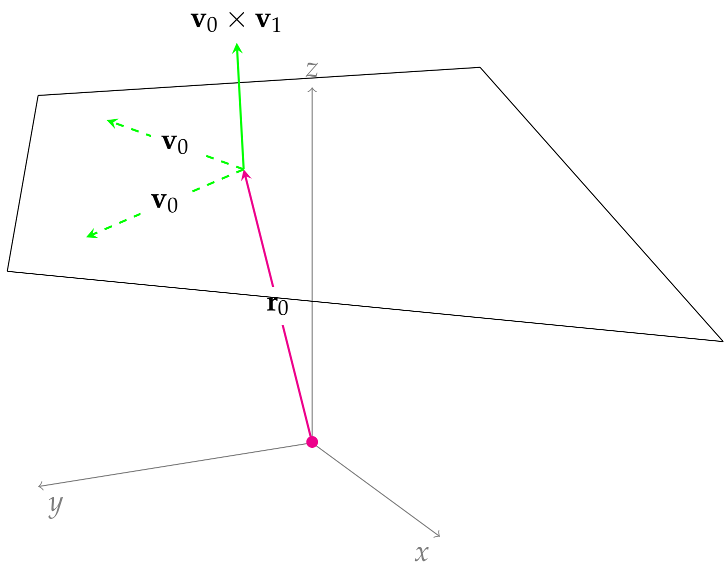
defines a *plane* in $\mathbb{R}^n$.



Notice however, that the two vectors $\mathbf{v}_0$ and $\mathbf{v}_1$ uniquely (up to scalar multiple) define a cross product and that this cross product can instead be used to define the vector. This cross product is called the *normal* of the plane given by $\mathbf{v}_0$ and $\mathbf{v}_1$. We denote by

$$\mathbf{n} = \mathbf{v}_0 \times \mathbf{v}_1.$$

Note *every* vector in the plane is orthogonal to this normal vector.



3.2.2 Vector Equation

Definition 3.11 (Vector Equation of the Plane)**.** Let $\mathbf{n}$ be a vector and $\mathbf{r}_0$ be a fixed position vector. The plane through $\mathbf{r}_0$ with normal $\mathbf{n}$ is given by

$$\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) = 0$$

for vector-valued variable $\mathbf{r}$. Equivalently we may also write

$$\mathbf{n} \cdot \mathbf{r} = \mathbf{n} \cdot \mathbf{r}_0$$

for the plane.

To obtain a scalar equation for the plane write

$$\mathbf{n} = \langle a, b, c \rangle \qquad \mathbf{r} = \langle x, y, z \rangle \qquad \mathbf{r}_0 = \langle x_0, y_0, z_0 \rangle$$

and recall the vector equation of the plane is given by $\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0)$ or

$$\langle a, b, c \rangle \cdot \langle x - x_0, y - y_0, z - z_0 \rangle.$$

Expanding yields

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

as the equation of the plane in $\mathbb{R}^3$ — a kind of "point-normal" analogue of the point-slope equation for the line.

Definition 3.12 (Scalar Equation of the Plane)**.** The *scalar equation of the plane* through (x_0, y_0, z_0) with normal vector $\mathbf{n} = \langle a, b, c\rangle$ is given by

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

3.2.3 Standard Form

Definition 3.13 (Linear Equation of the Plane)**.** Let a, b, c, d be reals and x, y, z real valued variables. Then

$$ax + by + cz + d = 0$$

defines the plane in three-space.

Example 3.14. Consider the plane given by

$$2(x - 1) - 3(y - 2) + 7(z) = 0.$$

The planes normal is $\langle 2, -3, 7\rangle$ and the plane passes through the point $(1, 2, 0)$.

Question 3.15. Find an equation of the plane through the point $(2, 4, -1)$ with normal $\mathbf{n} = \langle 2, 3, 4\rangle$. Find where this plane intersects the x, y, and z axis and sketch the plane.

ANSWER. Trivially, the plane is given by $\mathbf{n} \cdot \langle x - 2, y - 4, z + 1\rangle$ or equivalently

$$2(x - 2) + 3(y - 4) + 4(z + 1) = 0$$
$$\implies 2x + 3y + 4z = 12.$$

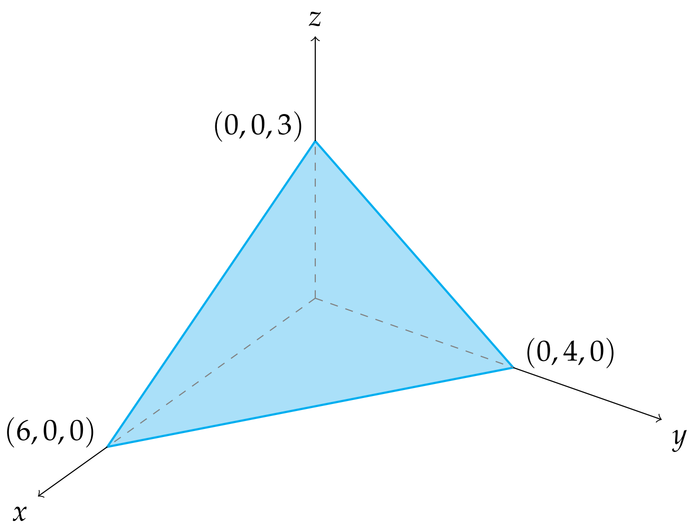
The x-intercept is found by setting $y = z = 0$ (and so on). Doing so yields

$$x = 6 \qquad y = 4 \qquad z = 3.$$

So we have the plane passes through

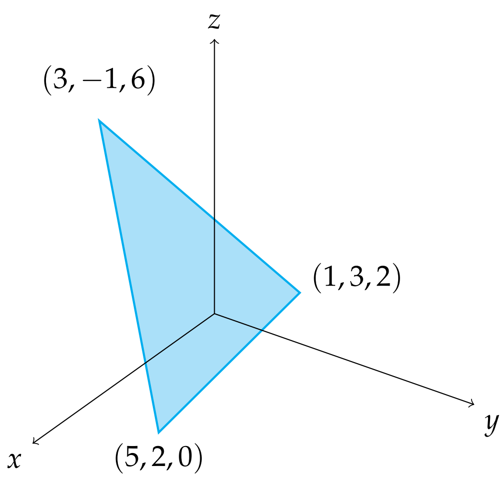
$$(6, 0, 0) \qquad (0, 4, 0) \qquad (0, 0, 3)$$

which we can sketch.



◆

Question 3.16. Find an equation of the plane that passes through the points $(1, 3, 2)$, $(3, -1, 6)$, and $(5, 2, 0)$.



ANSWER. We can get two (arbitrary) direction vectors from these three points. (The vectors should be given from the same tail.)

$$\mathbf{v}_0 = \langle 3, -1, 6 \rangle - \langle 1, 3, 2 \rangle = \langle 2, -4, 4 \rangle$$
$$\mathbf{v}_1 = \langle 5, 2, 0 \rangle - \langle 1, 3, 2 \rangle = \langle 4, -1, -2 \rangle$$

The normal to the plane is then $\langle 2, -4, 4 \rangle \times \langle 4, -1, -2 \rangle$ which we compute by

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & -4 & 4 \\ 4 & -1 & -2 \end{vmatrix} = \begin{vmatrix} -4 & 4 \\ -1 & -2 \end{vmatrix} \mathbf{i} - \begin{vmatrix} 2 & 4 \\ 4 & -2 \end{vmatrix} \mathbf{j} + \begin{vmatrix} 2 & -4 \\ 4 & -1 \end{vmatrix} \mathbf{k}$$
$$= \langle 12, 20, 14 \rangle.$$

Choosing the point $(1, 3, 2)$ (though the other two are fine as well) we can give the equation of the plane as

$$12(x-1) + 20(y-3) + 14(z-2) = 0$$

which simplifies to the linear equation

$$6x + 10y + 7z = 50.$$

◆

3.2.4 Intersections

Question 3.17. Find the point at which the line

$$x = 2 + 3t \qquad y = -4t \qquad z = 5 + t$$

intersects the plane $4x + 5y - 2z = 18$.

ANSWER. The point or line of intersection must be those points (x, y, z) satisfying both equations simultaneously. Thus we substitute the points of the line into the equation of the plane

$$4(2 + 3t) + 5(-4t) - 2(5 + t) = 18$$

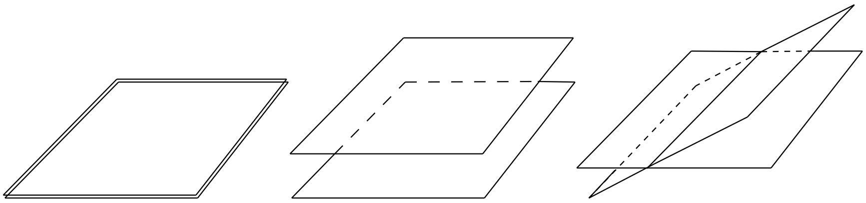
and solve for $t = -2$. Thus the *point* or intersection is

$$(2 + 3(-2),\ -4(-2),\ 5 + (-2)) = (-4,\ 8,\ 3).$$

◆

Definition 3.18. Two *planes* are *parallel* if their normal vectors are parallel. (Note this does not preclude that the planes are identical.)

In fact, the only types of intersection the plane can have are:



Question 3.19. Are the two planes given by

$$x + 2y - 3z = 4 \qquad\qquad 2x + 4y - 6z = 3$$

parallel?

ANSWER. Yes. The normals are $\langle 1, 2, -3\rangle$ and $\langle 2, 4, -6\rangle$ respectively — which are clearly parallel because they only differ by the scalar multiple 2. ◆

3.2.5 Angle Between Planes

Question 3.20. Find the angle between the planes $x + y + z = 1$ and $x - 2y + 3z = 1$ and then give the line of intersection as a symmetric equation.

ANSWER. The normal vector of these planes are

$$\mathbf{n}_0 = \langle 1, 1, 1\rangle \qquad\qquad \mathbf{n}_1 = \langle 1, -2, 3\rangle$$

Notice the angle between the planes is the same as the angle between the normals which is given by

$$\begin{aligned}\cos\theta &= \frac{\mathbf{n}_0 \cdot \mathbf{n}_1}{|\mathbf{n}_0||\mathbf{n}_1|} \\ &= \frac{1(1) + 1(-2) + 1(3)}{\sqrt{1+1+1}\sqrt{1+4+9}} \\ &= \frac{2}{\sqrt{42}} \implies \theta \approx 72^\circ.\end{aligned}$$

Now for the line of intersection: Remember, every line of a plane is perpendicular to the plane's normal. Thus a line in *two* planes must be perpendicular to *both* normals. However, there is a unique (up to scalar multiple) vector $\mathbf{v}$ perpendicular to $\mathbf{n}_0$ and $\mathbf{n}_1$ and that is

$$\mathbf{v} = \mathbf{n}_0 \times \mathbf{n}_1 = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 1 & 1 \\ 1 & -2 & 3 \end{vmatrix} = \langle 5, -2, -3 \rangle.$$

That is, this is the *direction vector* for the line! All we need now is a point.

As any point on both planes will do let us solve for when $z = 0$ in both equations (i.e. a solution in the xy-plane). That is, we want a solution of

$$x + y - 1 = 0 \qquad\qquad x - 2y - 1 = 0$$

Subtracting the equations gives $3y = 0$ and thus $(1, 0, 0)$ is a point on both planes. The line of intersection is given by

$$\langle x, y, z \rangle = \langle 1, 0, 0 \rangle + t\langle 5, -2, -3 \rangle$$

which corresponds to the symmetric equation

$$\frac{x-1}{5} = \frac{y-0}{-2} = \frac{z-0}{-3}.$$

◆

3.2.6 Lines as Plane Intersections

It stands to reason that we can define lines in three-space by the intersection of two planes.

Proposition 3.21. In general, the equation of a line given by

$$\frac{x-a_0}{a} = \frac{y-y_0}{b} = \frac{z-z_0}{c}$$

can be regarded to be the line of intersection of the two planes

$$\frac{x-x_0}{a} = \frac{y-y_0}{b} \qquad\qquad \frac{y-y_0}{b} = \frac{z-z_0}{c}.$$

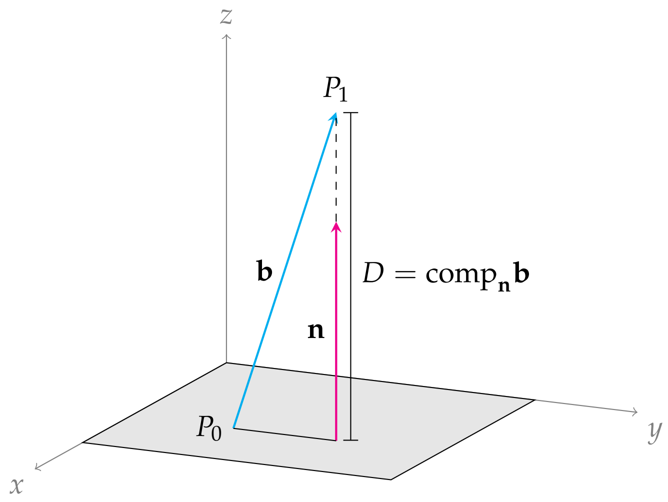
3.2.7 Shortest Distance to Plane

Question 3.22. Find a formula for the distance D from point $P = (x_1, y_1, z_1)$ to the plane $ax + by + cz + d = 0$.

ANSWER. Let (x_0, y_0, z_0) be a point from the plane and let

$$\mathbf{b} = \langle x_1 - x_0, y_1 - y_0, z_1 - z_0 \rangle.$$

The shortest distance to the plane is given by the projection of the vector $\mathbf{b}$ onto the normal $\mathbf{n} = \langle a, b, c \rangle$ of the plane.



So, we need only calculate the *component* (i.e. the length of the projection) of $\mathbf{b} = \langle x_1 - x_0,\ y_1 - y_0,\ z_1 - z_0 \rangle$ onto $\mathbf{n} = \langle a,\ b,\ c \rangle$:

$$\begin{aligned} D &= |\mathrm{comp}_{\mathbf{n}} \mathbf{b}| \\ &= \frac{|\mathbf{n} \cdot \mathbf{b}|}{|\mathbf{n}|} && \text{By definition.} \\ &= \frac{|a(x_1 - x_0) + b(y_1 - y_0) + c(z_1 - z_0)|}{\sqrt{a^2 + b^2 + c^2}} \\ &= \frac{|(ax_1 + by_1 + cz_1) - (ax_0 + by_0 + cz_0)|}{\sqrt{a^2 + b^2 + c^2}}. \end{aligned}$$

We have calculated the distance from P_1 to *any* point P_0 on the plane with normal $\langle a, b, c \rangle$ is

$$D = \frac{|(ax_1 + by_1 + cz_1) - (ax_0 + by_0 + cz_0)|}{\sqrt{a^2 + b^2 + c^2}}.$$

However, as we know (x_0, y_0, z_0) is on the plane we must have

$$ax_0 + by_0 + cz_0 + d = 0$$

and thereby $ax_0 + by_0 + cz_0 = -d$. (Notice that the point (x_0, y_0, z_0) has been eliminated from the equation!) Thus, this is the "distance to the plane" is

$$D = \frac{|ax_1 + by_1 + cz_1 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

◆

Let us repeat this answer using the algebra rules instead of using explicit points. That is, we calculate the distance from arbitrary point $\mathbf{x}_1$ to the plane given by $ax + by + cz + d = 0$ with normal $\mathbf{n} = \langle a, b, c \rangle$: Let $\mathbf{x}_0$ lie on the plane (thus $\mathbf{n} \cdot \mathbf{x}_0 = -d$) and let $\mathbf{b} = \mathbf{x}_0 - \mathbf{x}_1$

$$\begin{aligned} D = |\text{comp}_{\mathbf{n}} \mathbf{b}| &= \frac{|\mathbf{n} \cdot \mathbf{b}|}{|\mathbf{n}|} \\ &= \frac{\mathbf{n}\mathbf{x}_0 - \mathbf{n}\mathbf{x}_1}{|\mathbf{n}|} && \text{distribution} \\ &= \frac{|-d - \mathbf{n}\mathbf{x}_1|}{|\mathbf{n}|} \\ &= \frac{|\mathbf{n}\mathbf{x}_1 + d|}{|\mathbf{n}|}. \end{aligned}$$

Question 3.23. Find the distance between the two *parallel* planes $10x + 2y - 2z = 5$ and $5x + y - z = 1$. (If they were not parallel the distance would be zero.)

ANSWER. First notice the normals are $\langle 10, 2, -2 \rangle$ and $\langle 5, 1, -1 \rangle$ which indeed correspond to parallel planes. We need only calculate the distance from *any* point on the first plane to the second plane. We *just* devised a formula for this.

So, let us pick an arbitrary point on $10x + 2y - 2z = 5$, say $(\frac{1}{2}, 0, 0)$, and find its distance to the plane $5x + 1y - 1z = 1$:

$$D = \frac{|(5)(\frac{1}{2}) + (1)(0) + (-1)(0) + (-1)|}{|\langle 5, 1, -1 \rangle|} = \frac{3/2}{3\sqrt{3}} = \frac{\sqrt{3}}{6}.$$

◆

Question 3.24. We previously showed the lines

$$\begin{aligned} L_0: &\quad x = 1 + t \quad && y = -2 + 3t \quad && z = 4 - t \\ L_1: &\quad x = 2s && y = 3 + s && z = -3 + 4s \end{aligned}$$

were skew. What then, is the distance between them?

ANSWER. If the lines L_0 and L_1 are skew then they be viewed as laying on two separate parallel planes P_0 and P_1. The distance between the lines is the same as the distance between the planes.

The normal of these planes, for them to be parallel, should be the cross-product of the line's direction vectors:

$$\langle 1,3,-1\rangle \times \langle 2,1,4\rangle = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & -1 \\ 2 & 1 & 4 \end{vmatrix} = \langle 13,-6,-5\rangle.$$

We're now free to choose a point from one line and calculate the distance to the plane given by the other line. Setting $s = t = 0$ we see the point $(1,-2,4)$ lies on L_0 and $(0,3,-3)$ on L_1. The plane defined by L_1 is given by

$$\begin{aligned} 13(x-0) - 6(y-3) - 5(z+3) &= 0 \\ \implies 13x - 6y - 5z + 3 &= 0 \end{aligned}$$

By our equation, the distance from the point $(1,-2,4)$ to the plane $13x - 6y - 5z + 3 = 0$ is

$$D = \frac{|13(1) - 6(-2) - 5(4) + 3|}{|\langle 13,-6,-5\rangle|} = \frac{8}{\sqrt{230}}$$

◆

3.3 Systems of Linear Equations

3.3.1 Linear Equations

Linear equations are those defined by the dot-product of a variable-valued vector $\mathbf{x}$ and a real-valued vector $\mathbf{a}$:

$$\langle x_1, \ldots, x_n\rangle \cdot \langle a_0, \ldots, a_n\rangle = b.$$

Definition 3.25 (Linear Equation). A *linear equation* in n variables $x_1, \ldots, x_n$ is given by

$$a_1x_1 + \cdots + a_nx_n = b$$

for $a_0, \ldots, a_n, b \in \mathbb{R}$. (The variables are sometimes called *unknowns*.)

Question 3.26. Which of the following are linear equations?

1. $x + 3y = 7$, *linear*
2. $x_0 - 3x_2 + x_3 = 7$, *linear*
3. $y - \sin x = 0$, *not linear*
4. $x + 3y^2 = 7$, *not linear*

5. $y = \frac{1}{2}x + 3z + 1$, *linear*
6. $x_0 + x_1 + \cdots + x_{n-1} = 1$, *linear*
7. $3x + 2y - z + xz = 4$, *not linear*
8. $\sqrt{x_0} + 2x_1 + x_2 = 1$. *not linear*

Definition 3.27 (Solution Set)**.** The *solution set* of the linear equation $f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{a} - b$ is all points $p = (p_1, \ldots, p_n) \in \mathbb{R}^n$ such that $f(p) = 0$. This set is sometimes denoted by the *zero set* $\mathbf{V}(f)$ and is given by

$$\mathbf{V}(f) := \{p \in \mathbb{R}^n \,:\, a_1 p_1 + \cdots + a_n p_n - b = 0\}.$$

Example 3.28. Let $f = x - 2$ in $\mathbb{R}^3$ then

$$\mathbf{V}(f) = \{(2, 0, 0)\}.$$

(Be mindful of the *ambient space* $\mathbb{R}^3$.)

Example 3.29. Let $f = 4x - 2y - 1$ in $\mathbb{R}^2$ then

$$\mathbf{V}(f) = \left\{ \left(t, 2t - \frac{1}{2}\right) \,:\, t \in \mathbb{R} \right\}.$$

This is called a *parameterization* of the solution set. One can find such a solution by setting one of the variables, say x, to t and solving for the remaining ones.

Alternatively, the solution is the *line* given by

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix} + t \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Example 3.30. Let $f = x_0 + 4x_1 + 7x_2 - 5$ then letting $x_1 = s$ and $x_2 = t$ we get $x_0 = 5 + 4s - 7t$ and thus

$$\mathbf{V}(f) = \{(5 + 4s - 7t,\, s,\, t) \,:\, s, t \in \mathbb{R}\}.$$

Generally speaking, one can use one fewer number of parameters than the dimension of the space (i.e. number of variables) to describe a line.

3.3.2 Linear Systems

Definition 3.31 (Linear Systems)**.** A finite set of linear equations $\mathbf{f} = \{f_1, \ldots, f_n\}$ in $x_1, \ldots, x_n$ forms a *system of linear equations* or *linear system*. The solution set for a linear system is denoted $\mathbf{V}(\mathbf{f})$ and given by

$$\mathbf{V}(f_1, \ldots, f_n) := \{p \in \mathbb{R}^n \,:\, f_1(p) = \cdots = f_n(p) = 0\}.$$

That is, the solution set contains all points which *simultaneously* zero all the linear systems.

Example 3.32. Consider the system $\mathbf{f}$ defined by

$$\begin{cases} 4x - y + 3z = -1 \\ 3x + y + 9z = -4 \end{cases}$$

then we have, for example,

$$(1, 2, -1) \in \mathbf{V}(\mathbf{f}).$$

But how do we find the rest of the solutions? (More on this later.)

Question 3.33. Let $f = x + y - 4$ and $g = 2x + 2y - 6$. What is $\mathbf{V}(f, g)$?

ANSWER. Consider that a solution to f is also a solution to $2f$. That is

$$f(p) = 0 \implies 2f(p) = 0.$$

Thus our system is equivalent to

$$\begin{aligned} 2x + 2y &= 8 \\ 2x + 2y &= 6 \end{aligned}$$

which clearly has no solutions. ($3x + 2y$ can never be simultaneously equal to 8 and 6.) ◆

Definition 3.34 (Inconsistent System). A linear system $\mathbf{f}$ satisfying

$$\mathbf{V}(\mathbf{f}) = \varnothing$$

is called an *inconsistent system*. Otherwise, the system is called *consistent*.

Proposition 3.35. Every system of linear equations has either

1. no solutions,
2. exactly one solution, or
3. infinitely many solutions.

(Remember, if two lines share two points then they must be the same line!) PROOF. ■

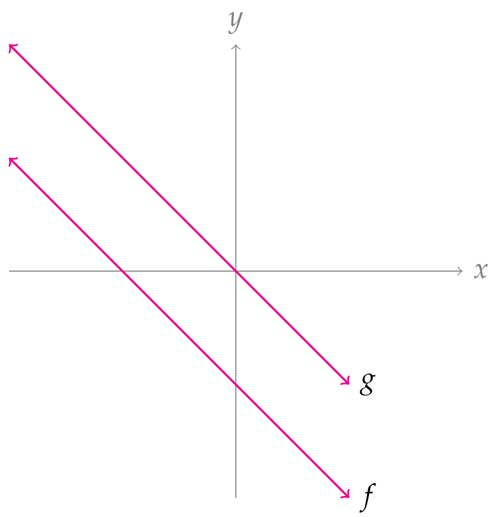


Figure 3.1: No solution. Inconsistent system.

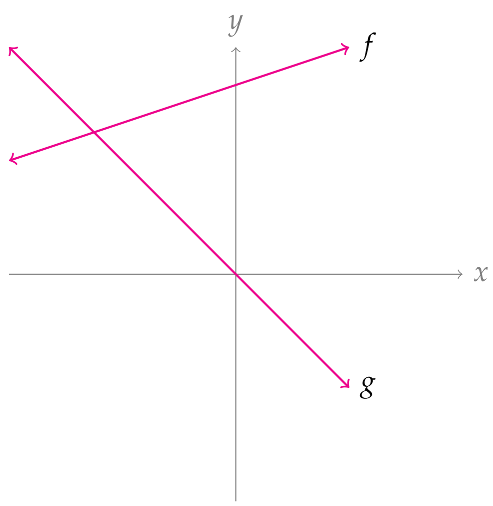


Figure 3.2: One solution. Consistent system.

3.4 Solving Linear Systems

In this setting "solving" a system $\mathbf{f}$ means finding $\mathbf{f}'$ such that

$$\mathbf{V}(\mathbf{f}) = \mathbf{V}(\mathbf{f}')$$

and $\mathbf{f}'$ is a trivially soluble (i.e. solveable) linear system like

$$\begin{aligned} x_n + a_{n-1}x_{n-1} + \cdots + a_1x_1 &= b_n \\ x_{n-1} + \cdots + a_1x_1 &= b_{n-1} \\ &\vdots \\ x_2 + a_1x_1 &= b_2. \\ x_1 &= b_1. \end{aligned}$$

Namely, $\mathbf{f}'$ has trivial *back substitution* (notice there are no coefficients on the 'largest' x_i in each equation).

Example 3.36. In $\mathbb{R}^3$ we have

$$\mathbf{V}\begin{pmatrix} x+y+2z-9 \\ 2x+4y-3z-1 \\ 3x+6y-5z \end{pmatrix} = \mathbf{V}\begin{pmatrix} x+y+2z-9 \\ 2y-7z-17 \\ z-3 \end{pmatrix} = \mathbf{V}\begin{pmatrix} x-1 \\ y-2 \\ z-3 \end{pmatrix} = \{(1,2,3)\}.$$

We say these systems are *similar* for which the symbol $\sim$ is used.

An arbitrary system of m linear equations in n unknowns $x_1, \ldots, x_n$ can be written like

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

Definition 3.37 (Augmented Matrix). Let $\mathbf{f}$ be the linear system defined above. The *augmented matrix* of $\mathbf{f}$ is given by

$$[\mathbf{A} \mid \mathbf{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

(Make sure each column contains the coefficient for the *same* variable.)

We will not use the line to separate $\mathbf{A}$ from $\mathbf{b}$ although a lot of texts do.

Question 3.38. Suppose we have a system of linear equations $\mathbf{f}$. How can we modify $\mathbf{f}$ so that $\mathbf{V}(\mathbf{f})$ does not change?

ANSWER. For $f \in \mathbf{f}$ we can replace f with

1. cf for any $c \in \mathbb{R} \setminus \{0\}$, and
2. $f - g$ for any $g \in \mathbf{f} \setminus \{f\}$.

(Note, combining these rules means we can replace f with $f - bg$ for any nonzero $b \in \mathbb{R}$ and $g \neq f \in \mathbf{f}$.) ◆

3.4.1 The Elementary Row Operations

There are three actions on matrices are called the *elementary row operations* and are sufficient for systematically solving linear systems.

Definition 3.39 (Elementary Row Operations). Let $\mathbf{A}$ be a matrix and c a scalar. Then the following are *elementary row operations* on $\mathbf{A}$:

1. Swapping two rows of $\mathbf{A}$,
2. Multiplying a row of $\mathbf{A}$ by c, and
3. Adding a row of $\mathbf{A}$ to another row of $\mathbf{A}$.

Notation. Swap rows i and j.

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{i1} & \cdots & a_{in} & b_i \\ \vdots & & \vdots & \vdots \\ a_{j1} & \cdots & a_{jn} & b_j \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{bmatrix} \begin{array}{c} \\ \\ \leftarrow \\ \\ \leftarrow \\ \\ \\ \end{array} = \begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{j1} & \cdots & a_{jn} & b_j \\ \vdots & & \vdots & \vdots \\ a_{i1} & \cdots & a_{in} & b_i \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{bmatrix}$$

Example 3.40.

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{bmatrix} \begin{array}{l} \leftarrow\!\urcorner \\ \quad | \\ \leftarrow\!\lrcorner \end{array} = \begin{bmatrix} 3 & 6 & -5 & 0 \\ 2 & 4 & -3 & 1 \\ 1 & 1 & 2 & 9 \end{bmatrix}$$

Notation. Multiplication of row ℓ by constant c.

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{\ell 1} & \cdots & a_{\ell n} & b_\ell \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{bmatrix} \begin{array}{l} \\ \\ \mid c \\ \\ \\ \end{array} = \begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ c\,a_{\ell 1} & \cdots & c\,a_{\ell n} & c\,b_\ell \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{bmatrix}$$

Example 3.41.

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{bmatrix} \begin{array}{l} \\ \mid 3 \\ \\ \end{array} = \begin{bmatrix} 1 & 1 & 2 & 9 \\ 6 & 12 & -9 & 3 \\ 3 & 6 & -5 & 0 \end{bmatrix}$$

Notation. Add c-times row i to row j.

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{i1} & \cdots & a_{in} & b_i \\ \vdots & & \vdots & \vdots \\ a_{j1} & \cdots & a_{jn} & b_j \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{bmatrix} \begin{array}{l} \\ \\ \urcorner^{c} \\ | \\ \leftarrow\!\lrcorner_{+} \\ \\ \\ \end{array} = \begin{bmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{i1} & \cdots & a_{in} & b_i \\ \vdots & & \vdots & \vdots \\ a_{j1} + c\,a_{i1} & \cdots & a_{jn} + c\,a_{in} & b_j + c\,b_i \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{bmatrix}$$

Example 3.42.

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{bmatrix} \begin{array}{l} \urcorner^{-2} \\ \leftarrow\!\lrcorner_{+} \\ \\ \end{array} = \begin{bmatrix} 1 & 1 & 2 & 9 \\ 2+(-2) & 4+(-2) & -3+(-4) & 1+(-18) \\ 3 & 6 & -5 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 3 & 6 & -5 & 0 \end{bmatrix}$$

Example 3.43. Combining notations

$$\begin{bmatrix}1&1&2&9\\2&4&-3&1\\3&6&-5&0\end{bmatrix} \begin{matrix} \rceil^{-2}\ \rceil^{-3} & & \\ \leftarrow^{+} & \rceil^{-\frac{3}{2}} & \\ \leftarrow^{+} & \leftarrow^{+} & \mid(-2)\end{matrix} = \begin{bmatrix}1&1&2&9\\0&2&-7&-17\\0&0&1&3\end{bmatrix}.$$

3.4.2 *Back-Substitution*

Definition 3.44 (Leading Coefficient). The *leading coefficient* or *pivot* of a matrix row is the “left-most” nonzero element of that row. We say a zero row has a zero pivot.

Question 3.45. What are the row pivots of the following matrix?

$$\begin{bmatrix}1&2&3\\0&0&0\\0&5&6\end{bmatrix}$$

ANSWER. The pivot of row 1, 2, and 3 is 1, 0, and 5 respectively. ◆

Definition 3.46 (Row-Echelon Form). A matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ is in *row-echelon form* or *upper-triangular form* if

1. no row with a zero pivot is above a row with a non-zero pivot, and
2. each pivot is strictly to the right of the pivot above it.

Proposition 3.47. A matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ is in *row-echelon form* when all entries in a column below a pivot are zeros.

Example 3.48. The following is in *row-echelon form*

$$\begin{bmatrix}1&\times&\times&\times&\times\\0&0&2&\times&\times\\0&0&0&1&\times\end{bmatrix}.$$

(The $\times$s are meant to emphasize that the values are irrelevant.)

Question 3.49. What is special about row-echelon form?

ANSWER. Back substitution. ◆

The following system can be solved with *back substitution*

$$x + y + 2z = 9$$

$$\begin{aligned} 2y - 7z &= -17 \\ z &= 3. \end{aligned}$$

because y can be recovered from

$$y = \frac{7(3) - 17}{2} = 2$$

and x from

$$x = 9 - 2(3) - 2 = 1.$$

In general any matrix in row-echelon form will have this property.

3.4.3 Gaussian Elimination

Using the *elementary row operations* any matrix $\mathbb{A}$ can be reduced to row-echelon form.

Question 3.50. Solve the linear system given by

$$\begin{aligned} -2z + t &= 12 \\ 2x + 4y - 10z + 6s + 12t &= 28 \\ 2x + 4y - 5z + 6s - 5t &= -1 \end{aligned}$$

by reducing the augmented matrix to row-echelon form and performing back-substitution.

ANSWER. The augmented matrix of

$$\begin{aligned} -2z + t &= 12 \\ 2x + 4y - 10z + 6s + 12t &= 28 \\ 2x + 4y - 5z + 6s - 5t &= -1 \end{aligned}$$

is

$$\left[\begin{array}{ccccc|c} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{array}\right].$$

(Note the ordering x, y, z, s, t of the columns.)

$$\begin{bmatrix} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \\ 0 & 0 & -2 & 0 & 7 & 12 \end{bmatrix}$$

$$\sim \begin{bmatrix} 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \\ 0 & 0 & -2 & 0 & 7 & 12 \end{bmatrix} \begin{matrix} \urcorner^{-1} \\ \hookleftarrow_{+} \\ \end{matrix} = \begin{bmatrix} 2 & 4 & -10 & 6 & 12 & 28 \\ 0 & 0 & 5 & 0 & -17 & -29 \\ 0 & 0 & -2 & 0 & 7 & 12 \end{bmatrix}$$

$$\sim \begin{bmatrix} 2 & 4 & -10 & 6 & 12 & 28 \\ 0 & 0 & 5 & 0 & -17 & -29 \\ 0 & 0 & -2 & 0 & 7 & 12 \end{bmatrix} \begin{matrix} \\ \mid 2 \\ \mid 5 \end{matrix} = \begin{bmatrix} 2 & 4 & -10 & 6 & 12 & 28 \\ 0 & 0 & 10 & 0 & -34 & -58 \\ 0 & 0 & -10 & 0 & 35 & 60 \end{bmatrix}$$

$$\sim \begin{bmatrix} 2 & 4 & -10 & 6 & 12 & 28 \\ 0 & 0 & 10 & 0 & -34 & -58 \\ 0 & 0 & -10 & 0 & 35 & 60 \end{bmatrix} \begin{matrix} \\ \urcorner \\ \hookleftarrow_{+} \end{matrix} = \begin{bmatrix} 2 & 4 & -10 & 6 & 12 & 28 \\ 0 & 0 & 10 & 0 & -34 & -58 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

(This is row-echelon form.)

Returning to a system of linear equations gives

$$\begin{aligned} 2x + 4y - 10z + 6s + 12t &= 28 \\ 10z - 34t &= -58 \\ t &= 2 \end{aligned}$$

and applying back substitution yields

$$z = \frac{34(2) - 58}{10} = 1$$

and

$$2x + 4y + 6s = 28 + 10(1) - 12(2) = 26.$$

Thus

$$\mathbf{V}(\mathbf{f}) = \{(x, y, 1, s, 2) \,:\, 2x + 4y + 6s = 26\}$$

(i.e. there are an infinite number of solutions). ◆

Definition 3.51 (Gaussian Elimination). Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ be a matrix. To convert $\mathbf{A}$ into row-echelon form do

1. If two rows have a non-zero pivot in the same column use one to eliminate the pivot of the other by elementary row operations.

2. Repeat 1. until no column has two non-zero pivots.

3. Swap rows so that each pivot is strictly right of the one above it.

Suppose $\mathbf{a}$ and $\mathbf{b}$ are two rows from matrix $\mathbf{A}$ that look like

$$\mathbf{a} = [0, \ldots, a_i, \ldots, a_n] \qquad \mathbf{b} = [0, \ldots, b_i, \ldots, b_n]$$

then

$$b_i\mathbf{a} - a_i\mathbf{b} = [0, \ldots, 0, a_{i+1}b_i - a_ib_{i+1}, \ldots, a_nb_i - a_ib_n]$$

or

$$\frac{1}{a_i}\mathbf{a} - \frac{1}{b_i}\mathbf{b} = \left[0, \ldots, 0, \frac{a_{i+1}}{a_i} + \frac{b_{i+1}}{b_i}\right]$$

does not have a pivot in column i.

Question 3.52. Solve $\mathbf{f} \subset \mathbb{R}[x, y, z]$ given by

$$\begin{aligned} x + y + 2z &= 9 \\ 2x + 4y - 3z &= 1 \\ 3x + 6y - 5z &= 0 \end{aligned}$$

by Gaussian elimination and back substitution.

ANSWER. The augmented system is given by $\begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{bmatrix}$ and the Gaussian elimination can be carried out like

$$\begin{aligned}
\begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{bmatrix}
\begin{array}{l} \urcorner^{-3} \\ | \\ \leftarrow\lrcorner_{+} \end{array}
&= \begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 0 & 3 & -11 & -27 \end{bmatrix} \\
\sim \begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 0 & 3 & -11 & -27 \end{bmatrix}
\begin{array}{l} \urcorner^{-2} \\ \leftarrow\lrcorner_{+} \\ \end{array}
&= \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 0 & 3 & -11 & -27 \end{bmatrix} \\
\sim \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 0 & 3 & -11 & -27 \end{bmatrix}
\begin{array}{l} \\ \urcorner^{-3} \\ |2 \leftarrow\lrcorner_{+} \end{array}
&= \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 0 & 0 & -1 & -3 \end{bmatrix}
\end{aligned}$$

(This is row echelon form and in particular *Hermite normal form*.)

Returning to a system, $\mathbf{f}$, of linear equations gives

$$\begin{aligned} x + y + 2z &= 9 \\ 2y - 7z &= -17 \\ -z &= -3 \end{aligned}$$

and applying back substitution yields

$$z = 3$$

$$y = \frac{7(3) - 17}{2} = 2$$
$$x = 9 - 2(3) - 2 = 1.$$

Thus $\mathbf{V}(\mathbf{f}) = \{(1, 2, 3)\}$ which is indeed a solution to

$$\{x + y + 2z = 9,\ 2x + 4y - 3z = 1,\ 3x + 6y - 5z = 0\}.$$

◆

Let us repeat the same question using fractions instead of least-common-multiples for pivoting.

ANSWER.

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{bmatrix} \begin{array}{l} R_1 \cdot (-3) \to R_3,\ R_1 \cdot (-2) \to R_2 \end{array} = \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 0 & 3 & -11 & -27 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 0 & 3 & -11 & -27 \end{bmatrix} \begin{array}{l} \\ \mid \frac{1}{2} \\ \mid \frac{1}{3} \end{array} = \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 1 & -\frac{11}{3} & -9 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 1 & -\frac{11}{3} & -9 \end{bmatrix} \begin{array}{l} R_2 \cdot (-1) \to R_3 \end{array} = \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & -\frac{1}{6} & -\frac{1}{2} \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & -\frac{1}{6} & -\frac{1}{2} \end{bmatrix} \begin{array}{l} \\ \\ \mid -6 \end{array} = \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Which gives (back substituting)

$$z = 3$$
$$y = \frac{7}{2}(3) - \frac{17}{2} = 2$$
$$x = -1(2) - 2(3) + 9 = 1.$$

But why stop here? Notice how z was easy to retrieve because it does not depend on the other variables — let us do this with every variable.

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & 1 & 3 \end{bmatrix} \begin{array}{l} R_3 \cdot \frac{7}{2} \to R_2 \end{array} = \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & 0 & \frac{21}{2} - \frac{17}{2} \\ 0 & 0 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

$$\sim \begin{bmatrix}1&1&2&9\\0&1&0&2\\0&0&1&3\end{bmatrix} \begin{matrix}\leftarrow^{+} & \leftarrow^{+}\\ \rfloor_{-1} & \\ & \rfloor_{-2}\end{matrix} = \begin{bmatrix}1&0&2&7\\0&1&0&2\\0&0&1&3\end{bmatrix} \begin{matrix}\leftarrow^{+}\\ \\ \rfloor_{-2}\end{matrix} = \begin{bmatrix}1&0&0&1\\0&1&0&2\\0&0&1&3\end{bmatrix}$$

(This is *reduced row echelon form*.)

Correspondingly, the solutions are given by

$$x = 1 \qquad y = 2 \qquad z = 3$$

which requires no back-substitution. ◆

Definition 3.53 (Hermite normal form). A matrix $\mathbf{A}$ is in *Hermite normal form* when it is in row echelon form and all of its entries are integers.

Definition 3.54 (Reduced Row Echelon Form). A matrix $\mathbf{A}$ is in *reduced row echelon form* when it is in row echelon form and each nonzero pivot is 1 and the only non-zero element of the column.

Example 3.55. The following is in *reduced row-echelon form*

$$\begin{bmatrix}1&\times&0&0&\times\\0&0&1&0&\times\\0&0&0&1&\times\end{bmatrix}.$$

(The $\times$s are meant to emphasize that the values are irrelevant.)

3.5 Curve Fitting / Interpolation

Proposition 3.56. Suppose there is no degree-$n-1$ polynomial through $p_0, \dots, p_n \in \mathbb{R}^2$, then there is exactly one curve of degree n though the $n+1$ points. That is, there is a unique degree-n polynomial $f \in \mathbb{R}[x]$ such that

$$f(p_0) = \cdots = f(p_n) = 0.$$

(Intuitively, two points define a line, three points a parabola, and so on.)

Question 3.57. What parabola passes through the points $(1, 1)$, $(2, -1)$ and $(3, -1)$? (See Figure 3.4.)

ANSWER. Substituting our (x, y) points into the general equation of the parabola

$$y = ax^2 + bx + c$$

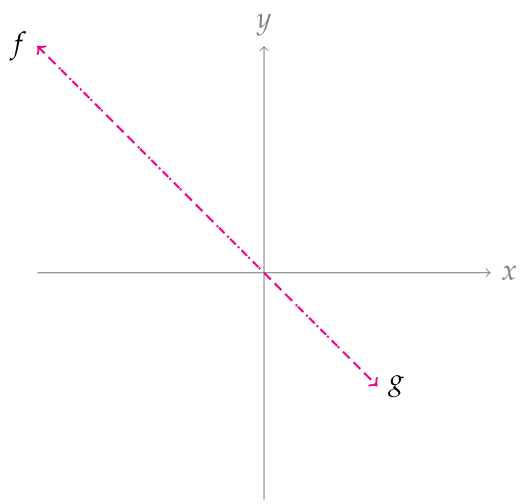


Figure 3.3: Infinite solutions. Consistent system.

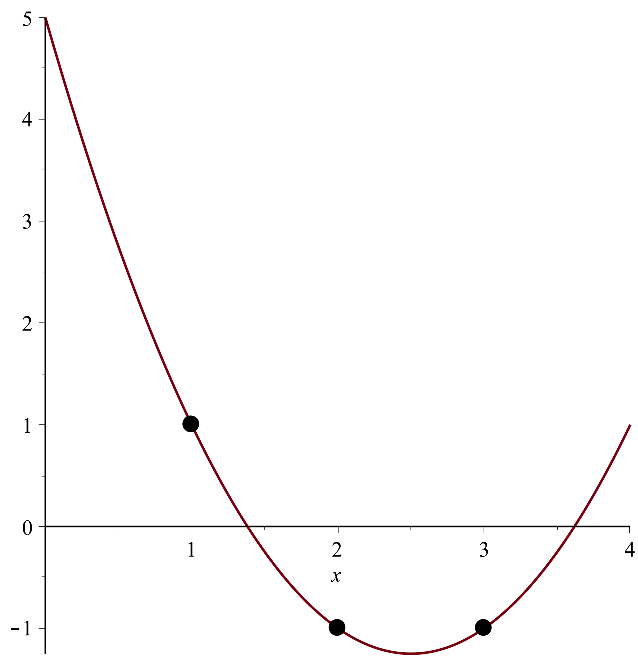


Figure 3.4: The parabola through the points $(1,1)$, $(2,-1)$, and $(3,-1)$.

produces the system of linear equations

$$\begin{aligned} a + b + c &= 1 \\ 4a + 2b + c &= -1 \\ 9a + 3b + c &= -1. \end{aligned}$$

Let us solve this using Gaussian elimination.

$$\begin{aligned}
\begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & 2 & 1 & -1 \\ 9 & 3 & 1 & -1 \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & -3 & -5 \\ 0 & -6 & -8 & -10 \end{bmatrix} \\
\sim \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & -3 & -5 \\ 0 & -6 & -8 & -10 \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & -3 & -5 \\ 0 & 0 & 1 & 5 \end{bmatrix} \\
\sim \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & -3 & -5 \\ 0 & 0 & 1 & 5 \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 0 & -4 \\ 0 & -2 & 0 & 10 \\ 0 & 0 & 1 & 5 \end{bmatrix} \\
\sim \begin{bmatrix} 1 & 1 & 0 & -4 \\ 0 & -2 & 0 & 10 \\ 0 & 0 & 1 & 5 \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & 5 \end{bmatrix}
\end{aligned}$$

Giving the solution

$$a = 1 \qquad b = -5 \qquad c = 5.$$

which is the parabola $x^2 - 5x + 5$. ◆

4 *Matrix Algebra*

Intuitively, a *matrix* $\mathbf{A}$ is a vector or vectors. $\mathbf{A}_{ij}$ then denotes the jth entry of the vector in the ith position.

Definition 4.1 (Matrix). A *matrix* $\mathbf{A}$ is a rectangular array (or vector) of *entries*. For instance

$$\mathbf{A} := \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}.$$

is an array of *dimension* $n \times m$. Its entires are a_{11} through a_{mn}. The set of all $n \times m$ matrices with entries from $\mathbb{R}$ is denoted $\mathbb{R}^{n \times m}$.

Question 4.2. What are the dimensions of the following matrices? (Remember: row $\times$ column).

1. $\begin{bmatrix} 1 & 2 \\ 3 & 0 \\ -1 & 4 \end{bmatrix}$ 3×2.
2. $\begin{bmatrix} 2 & 1 & 0 & -3 \end{bmatrix}$ 1×4.
3. $\begin{bmatrix} -\sqrt{2} & \pi & e \\ 3 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$ 3×3.

Example 4.3. Let $\mathbf{A} \in \mathbb{Z}^{2 \times 2}$ be

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

then $\mathbf{A}_{11} = 1$, $\mathbf{A}_{12} = 2$, $\mathbf{A}_{21} = 3$, and $\mathbf{A}_{22} = 4$.

Definition 4.4 (Row and Column Matrices). Matrices with dimension $1 \times n$ or $n \times 1$ are called (respectively) *row* and *column* matrices. For such matrices double subscripting is unnecessary. Instead we have

$$\mathbf{a} := \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \qquad \mathbf{b} := \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Definition 4.5. Two matrices are *equal* if they have the same size and the same entries at the same positions. That is, when $\mathbf{A}$ and $\mathbf{B}$ are from $\mathbb{R}^{n \times m}$

$$\mathbf{A} = \mathbf{B} \iff \forall i \in [1, n]\ \forall j \in [1, m];\ \mathbf{A}_{ij} = \mathbf{B}_{ij}.$$

Example 4.6. Consider the matrices

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 3 & x \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 2 & 1 & 0 \\ 3 & x & 0 \end{bmatrix}.$$

If $x = 5$ then $A = B$ but not for any other value of x. There is no value of x for which $\mathbf{A}$ and $\mathbf{C}$ are equal since they have different sizes.

4.1 Operations on Matrices

4.1.1 Matrix Addition

Definition 4.7 (Matrix Sum). Let $\mathbf{A}$ and $\mathbf{B}$ be matrices of the same dimension. The *sum* $A + B$ is given by

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{bmatrix}.$$

The *difference* $A - B$ is given in a similar fashion.

Alternatively we can view a matrix $\mathbf{A}$ as a column with vector entries $\mathbf{a}_i = \langle a_{i1}, \ldots, a_{in} \rangle$. With this interpretation we have

$$\begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix} + \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 + \mathbf{b}_1 \\ \vdots \\ \mathbf{a}_n + \mathbf{b}_n \end{bmatrix}$$

Example 4.8. Let

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 0 & 3 \\ -1 & 0 & 2 & 4 \\ 4 & -2 & 7 & 0 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} -4 & 3 & 5 & 1 \\ 2 & 2 & 0 & -1 \\ 3 & 2 & -4 & 5 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

then

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} -2 & 4 & 5 & 4 \\ 1 & 2 & 2 & 3 \\ 7 & 0 & 3 & 5 \end{bmatrix} \qquad \mathbf{A} - \mathbf{B} = \begin{bmatrix} 6 & -2 & -5 & 2 \\ -3 & -2 & 2 & 5 \\ 1 & -4 & 11 & -5 \end{bmatrix}$$

Note that each of $A + C$, $B + C$, $A - C$, and $B - C$ is undefined because they do not have the appropriate dimensions.

4.1.2 *Scalar Matrix Product*

Definition 4.9 (Scalar Matrix Product)**.** Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ be a matrix $c \in \mathbb{R}$ a scalar. Then the *scalar matrix product* is given by

$$c\mathbf{A} := \begin{bmatrix} c\,a_{11} & \cdots & c\,a_{1n} \\ \vdots & \ddots & \vdots \\ c\,a_{m1} & \cdots & c\,a_{mn} \end{bmatrix} = \begin{bmatrix} c\,\mathbf{a}_1 \\ \vdots \\ c\,\mathbf{a}_m \end{bmatrix}.$$

Example 4.10. Let

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 3 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 2 & 7 \\ -1 & 3 & -5 \end{bmatrix} \quad C = \begin{bmatrix} 9 & -6 & 3 \\ 3 & 0 & 12 \end{bmatrix}$$

then

$$2A = \begin{bmatrix} 4 & 6 & 8 \\ 2 & 6 & 2 \end{bmatrix} \quad -B = \begin{bmatrix} 0 & -2 & -7 \\ 1 & -3 & 5 \end{bmatrix} \quad \frac{1}{3}C = \begin{bmatrix} 3 & -2 & 1 \\ 1 & 0 & 4 \end{bmatrix}.$$

4.1.3 Matrix Product

Definition 4.11 (Row/Column). Let $\mathbf{A} \in \mathbb{R}^{n\times m}$ be a matrix given by

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$$

The ith row and jth column of $\mathbf{A}$ is given by

$$\text{row}_i\mathbf{A} := \begin{bmatrix} a_{i1} & \cdots & a_{im} \end{bmatrix} \qquad \text{col}_j\mathbf{A} := \begin{bmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{bmatrix}$$

Note that when $n = m$ we allow ourselves to take dot-product of these column and row matrices so that, for instance,

$$\text{row}_i\mathbf{A} \cdot \text{col}_j\mathbf{A} = a_{i1}a_{1j} + \cdots + a_{im}a_{nj}.$$

Definition 4.12 (Matrix Product). Let $\mathbf{A} \in \mathbb{R}^{n\times m}$ and $\mathbf{B} \in \mathbb{R}^{m\times \ell}$ be matrices (note the differing dimensions). The *matrix product* $\mathbf{A} \times \mathbf{B}$ is the $n \times \ell$ matrix given by

$$(\mathbf{A} \times \mathbf{B})_{ij} = \text{row}_i\mathbf{A} \cdot \text{col}_j\mathbf{B}$$

where $\cdot$ is the dot product.

Notice that in order for the dot product to be well-defined the number of entries of $\text{row}_i\mathbf{A}$ must be the same as $\text{col}_j\mathbf{A}$.

Example 4.13.

$$\begin{bmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{bmatrix} \begin{bmatrix} 4 & 1 & 4 & 3 \\ 0 & -1 & 3 & 1 \\ 2 & 7 & 5 & 2 \end{bmatrix} = \begin{bmatrix} & & & \\ & & 26 & \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{bmatrix} \begin{bmatrix} 4 & 1 & 4 & 3 \\ 0 & -1 & 3 & 1 \\ 2 & 7 & 5 & 2 \end{bmatrix} = \begin{bmatrix} & & & 13 \\ & & & \end{bmatrix}$$

Question 4.14. What is the matrix product of $\mathbf{A}$ and $\mathbf{B}$ when

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 4 & 1 & 4 & 3 \\ 0 & -1 & 3 & 1 \\ 2 & 7 & 5 & 2 \end{bmatrix}$$

ANSWER. First notice $\mathbf{A}$ is 2×3 and $\mathbf{B}$ is 3×4 so the product will be 2×4. Namely, it is

$$\begin{bmatrix} 4+0+8 & 1-2+28 & 4+6+20 & 3+2+8 \\ 8+0+0 & 2-6+0 & 8+18+0 & 6+6+0 \end{bmatrix} = \begin{bmatrix} 12 & 27 & 30 & 13 \\ 8 & -4 & 26 & 12 \end{bmatrix}$$

◆

Question 4.15. Suppose $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are matrices with the following dimensions

$$\mathbf{A} : 3 \times 4 \qquad \mathbf{B} : 4 \times 7 \qquad \mathbf{C} : 7 \times 3$$

What dimension is $\mathbf{A} \times \mathbf{B} \times \mathbf{C}$? Are there any other well-defined matrix products here?

ANSWER. 3×3. And yes, $\mathbf{C} \times \mathbf{A} \times \mathbf{B}$ and $\mathbf{B} \times \mathbf{C} \times \mathbf{A}$ are also well-defined.
◆

Proposition 4.16 (Properties of the Matrix Product)**.** Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be matrices and c a scalar. The matrix product is

1. *Not* commutative in general

$$\mathbf{AB} \neq \mathbf{BA}$$

2. Distributive

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \qquad (\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$$

3. Associativity of scalar multiplication

$$c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B} \qquad (\mathbf{AB})c = \mathbf{A}(\mathbf{B}c).$$

Example 4.17. Matrix multiplication is *not* commutative. Consider

$$\mathbf{A} = \begin{bmatrix} -1 & 0 \\ 2 & 3 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}$$

then

$$\mathbf{AB} = \begin{bmatrix} -1 & -2 \\ 11 & 4 \end{bmatrix} \qquad \mathbf{BA} = \begin{bmatrix} 3 & 6 \\ -3 & 0 \end{bmatrix}$$

and thus $\mathbf{AB} \neq \mathbf{BA}$.

Since matrix multiplication is not commutative we distinguish between *left-multiplication* of $\mathbf{A}$ on $\mathbf{B}$: $\mathbf{AB}$ and *right-multiplication* of $\mathbf{A}$ on $\mathbf{B}$: $\mathbf{BA}$.

4.2 Matrix Aritmetic

Proposition 4.18 (Rules of Matrix Arithmetic)**.** Assuming the dimensions of the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are such that the corresponding operations are well-defined, and that a and b are scalars, then

$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$	Commutativity of addition.
$\mathbf{A}+(\mathbf{B}+\mathbf{C})=(\mathbf{A}+\mathbf{B})+\mathbf{C}$	Associativity of addition.
$\mathbf{A}(\mathbf{BC})=(\mathbf{AB})\mathbf{C}$	Associativity of multiplication.
$(\mathbf{B}+\mathbf{C})\mathbf{A}=\mathbf{BA}+\mathbf{CA}$	Left distributive law.
$\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}$	Right distributive law.
$(a+b)\mathbf{C}=a\mathbf{C}+b\mathbf{C}$	
$a(b\mathbf{C})=(ab)\mathbf{C}$	
$a(\mathbf{BC})=(a\mathbf{B})\mathbf{C}=\mathbf{B}(a\mathbf{C})$	

LEFT DISTRIBUTIVE LAW.. Recall $\mathbf{A}_{ij}$ denotes the entry in the ith row and jth column of a matrix.

$$\begin{aligned}
&[\mathbf{A}(\mathbf{B}+\mathbf{C})]_{ij} \\
&\quad = \mathrm{row}_i\mathbf{A}\cdot\mathrm{col}_j(\mathbf{B}+\mathbf{C}) \\
&\quad = \mathrm{row}_i\mathbf{A}\cdot(\mathrm{col}_j\mathbf{B}+\mathrm{col}_j\mathbf{C}) \\
&\quad = \mathrm{row}_i\mathbf{A}\cdot\mathrm{col}_j(\mathbf{B})+\mathrm{row}_i\mathbf{A}\cdot\mathrm{col}_j(\mathbf{C}) && \text{Distributivity of dot product.} \\
&\quad = [\mathbf{AB}]_{ij}+[\mathbf{AC}]_{ij} && \text{Definition of } \times. \\
&\quad = [\mathbf{AB}+\mathbf{AC}]_{ij}
\end{aligned}$$

■

Example 4.19. To illustrate associativity of matrix multiplication consider

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} 1 & 20 \\ 2 & 3 \end{bmatrix}.$$

We expect that $(\mathbf{AB})\mathbf{C}$ and $\mathbf{A}(\mathbf{BC})$ are equal.

$$\mathbf{AB} = \begin{bmatrix} 8 & 5 \\ 20 & 13 \\ 2 & 1 \end{bmatrix} \qquad (\mathbf{AB})\mathbf{C} = \begin{bmatrix} 18 & 15 \\ 46 & 39 \\ 4 & 3 \end{bmatrix}.$$

$$\mathbf{BC} = \begin{bmatrix} 10 & 9 \\ 4 & 3 \end{bmatrix} \qquad \mathbf{A}(\mathbf{BC}) = \begin{bmatrix} 18 & 15 \\ 46 & 39 \\ 4 & 3 \end{bmatrix}.$$

And we see indeed that $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

4.2.1 *Additive and Multiplicative Zero and Identity*

There is an additive and multiplicative zero for matrix arithmetic.

Definition 4.20 (Zero Matrix). Let $\mathbf{0}^{n\times m}$ denote a $n \times m$ matrix comprised only of zeros:

$$\mathbf{0} = \underbrace{\begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}}_{m \text{ columns}} \left.\vphantom{\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}}\right\} n \text{ rows.}$$

Proposition 4.21. Assuming the dimensions of the matrices $\mathbf{A}$ and $\mathbf{0}$ are such that the corresponding operations are well-defined, then

1. $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$,
2. $\mathbf{A} - \mathbf{A} = \mathbf{0}$,
3. $\mathbf{0} - \mathbf{A} = -\mathbf{A}$, and
4. $\mathbf{A0} = \mathbf{0}$ and $\mathbf{0A} = \mathbf{0}$.

PROOF. Straight from definitions. ■

Do we have an additive and multiplicative inverse for matrix arithmetic? For addition, yes, because the additive inverse of $\mathbf{A}$ is $-\mathbf{A}$. But for multiplication, it depends. First consider what our multiplicative identity is: it is a matrix, say $\mathbf{I}$, with the property that

$$\mathbf{AI} = \mathbf{A} \qquad \text{or} \qquad \mathbf{IA} = \mathbf{A}?$$

Definition 4.22 (Identity Matrix). The *identity* matrix $\mathbf{I}_n$ is the $n \times n$ matrix with ones on its diagonal an zeros everywhere else. That is,

$$\mathbf{I}_n := \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}.$$

(We typically just write $\mathbf{I}$ instead of $\mathbf{I}_n$ because the dimension is implied.)

Proposition 4.23. Let $\mathbf{A} \in \mathbb{R}^{n\times m}$ be a $n \times m$ matrix, then

$$\mathbf{AI}_m = \mathbf{A} \qquad \text{and} \qquad \mathbf{I}_n\mathbf{A} = \mathbf{A}.$$

PROOF. Notice

$$\begin{aligned} [\mathbf{AI}_m]_{ij} &= \text{row}_i\mathbf{A} \cdot \text{col}_j\mathbf{I}_m \\ &= \mathbf{A}_{i1} \cdot 0 + \cdots + \mathbf{A}_{ij} \cdot 1 + \cdots + \mathbf{A}_{im} \cdot 0 \\ &= \mathbf{A}_{ij} \end{aligned}$$

■

Example 4.24. Consider multiplying on the left by the identity gives

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x & y & z \\ r & s & t \end{bmatrix} = \begin{bmatrix} x & y & z \\ r & s & t \end{bmatrix}$$

and from the right

$$\begin{bmatrix} x & y & z \\ r & s & t \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} x & y & z \\ r & s & t \end{bmatrix}.$$

4.3 Matrix Inverse

Definition 4.25 (Invertibility). The square matrix $\mathbf{A}$ is invertible when there is another square matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

Note the left *and* right multiplication of $\mathbf{A}^{-1}$ on $\mathbf{A}$. For this to be well defined it must be the case that $\mathbf{A}$, $\mathbf{A}^{-1}$, and $\mathbf{I}$ are square matrices of the equal dimension.

Example 4.26. $\mathbf{B} = \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$ is the inverse of $\mathbf{A} = \begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix}$ because

$$\mathbf{AB} = \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

and

$$\mathbf{BA} = \begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}.$$

Not all matrices are invertible. Consider $\mathbf{0}$ for instance.

Proposition 4.27. The matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 0 \end{bmatrix}$$

is not invertible.

PROOF. For any $\mathbf{B} \in \mathbb{R}^{3\times 3}$ we have that

$$\text{col}_3(\mathbf{BA}) = \begin{bmatrix} \text{row}_1\mathbf{B} \cdot \text{col}_3\mathbf{A} \\ \text{row}_2\mathbf{B} \cdot \text{col}_3\mathbf{A} \\ \text{row}_3\mathbf{B} \cdot \text{col}_3\mathbf{A} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \text{col}_3\mathbf{I}.$$

Thus $\mathbf{BA} \neq \mathbf{I}$ and so $\mathbf{A}$ cannot be invertible. ■

Proposition 4.28. If $\mathbf{B}$ and $\mathbf{C}$ are inverses of $\mathbf{A}$ then $\mathbf{B} = \mathbf{C}$. That is to say, matrix inverses are unique:

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I} \text{ and } \mathbf{AC} = \mathbf{CA} = \mathbf{I} \implies \mathbf{B} = \mathbf{C}.$$

PROOF. Towards a contradiction suppose $\mathbf{B} \neq \mathbf{C}$ are both inverses of $\mathbf{A}$. Thereby we have $(\mathbf{BA})\mathbf{C} = \mathbf{B}(\mathbf{AC})$ by associativity,

$$\mathbf{BA} = \mathbf{I} \implies (\mathbf{BA})\mathbf{C} = \mathbf{IC} = \mathbf{C},$$

and

$$\mathbf{AC} = \mathbf{I} \implies \mathbf{B}(\mathbf{AC}) = \mathbf{BI} = \mathbf{B}.$$

Thus $\mathbf{B} = \mathbf{C}$. $\lightning$ ■

Since inverses are unique we can now just simply refer to "the" inverse.

Notation. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a square matrix. The *inverse of* $\mathbf{A}$ is denoted $\mathbf{A}^{-1}$ and satisfies

$$\mathbf{AA}^{-1} = \mathbf{I}_n \qquad \text{and} \qquad \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n.$$

Proposition 4.29. If $\mathbf{A}$ and $\mathbf{B} \in \mathbb{R}^{n\times n}$ are invertible matrices, then:

1. $\mathbf{AB}$ is invertible, and
2. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

PROOF. Notice

$$\mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{AB} = \mathbf{I}$$

and

$$\mathbf{ABB}^{-1}\mathbf{A}^{-1} = \mathbf{I}$$

so $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

Provided $\mathbf{A}$ and $\mathbf{B}$ are invertible (that is, $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ exist) then $\mathbf{AB}$ is invertible. ■

More generally, the product of any number of invertible matrices is invertible, and the inverse of the product is the product of the inverses in reverse order.

Proposition 4.30. Let $\mathbf{A}_0, \ldots, \mathbf{A}_n \in \mathbb{R}^{n\times n}$ be invertible, then

$$(\mathbf{A}_0 \cdots \mathbf{A}_n)^{-1} = \mathbf{A}_n^{-1} \cdots \mathbf{A}_0^{-1}$$

PROOF. Exercise. ■

4.3.1 Inverting 2×2 matrices

Finding the inverse of a 2×2 matrix is relatively straightforward using the following proposition.

Proposition 4.31 (2×2 Matrix Inversion)**.** Let $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and recall $\det \mathbf{A} = ad - cb$. Provided $\det \mathbf{A} \neq 0$

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

PROOF. Consider

$$\begin{aligned} \mathbf{A}^{-1}\mathbf{A} &= \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \frac{1}{ad - bc} \begin{bmatrix} ad - bc & bd - bd \\ -ac + ca & ad - bc \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

(Do the right-multiplication of $\mathbf{A}^{-1}$ as an exercise.) ■

Question 4.32. What is the inverse of $\mathbf{AB}$ when

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix} \qquad \mathbf{AB} = \begin{bmatrix} 7 & 6 \\ 9 & 8 \end{bmatrix}?$$

ANSWER. Applying our formula for 2×2 determinants we find

$$\mathbf{A}^{-1} = \frac{1}{1} \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} \quad \mathbf{B}^{-1} = \frac{1}{2} \begin{bmatrix} 2 & -2 \\ -2 & 3 \end{bmatrix} \quad (\mathbf{AB})^{-1} = \begin{bmatrix} 4 & -3 \\ -\frac{9}{2} & \frac{7}{2} \end{bmatrix}.$$

To confirm our proposition notice

$$\mathbf{B}^{-1}\mathbf{A}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ -\frac{9}{2} & \frac{7}{2} \end{bmatrix}$$

and also $\mathbf{B}^{-1}\mathbf{A}^{-1} = (\mathbf{AB})^{-1}$. ◆

Definition 4.33 (Matrix Powers)**.** Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $m \in \mathbb{Z}$ be nonnegative. "$\mathbf{A}$ to the power of m" or the "mth power of $\mathbf{A}$" is given by

$$\mathbf{A}^m := \mathbf{A}\mathbf{A}^{m-1}$$

where $\mathbf{A}^0 := \mathbf{I}$. Moreover $\mathbf{A}^{-m} = (\mathbf{A}^{-1})^m$.

Proposition 4.34. If $\mathbf{A}$ is a square matrix and $r, s \in \mathbb{Z}$ then (as usual)

1. $\mathbf{A}^r\mathbf{A}^s = \mathbf{A}^{r+s}$, and
2. $(\mathbf{A}^r)^s = \mathbf{A}^{rs}$.

PROOF. Exercise. ■

Proposition 4.35. If $\mathbf{A}$ is invertible and k is a nonzero scalar then

1. $\mathbf{A}^{-1}$ is invertible and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$,
2. $\mathbf{A}^n$ is invertible and $(\mathbf{A}^n)^{-1} = (\mathbf{A}^{-1})^n$ for $n = 0, 1, 2, \ldots$ and
3. $k\mathbf{A}$ is invertible and $(k\mathbf{A})^{-1} = \frac{1}{k}\mathbf{A}^{-1}$.

OF 1.. Clearly $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ so by definition $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$. ■

OF 3.. Using the rules of matrix arithmetic we can write

$$(k\mathbf{A})\left(\frac{1}{k}\mathbf{A}^{-1}\right) = \frac{1}{k}(k\mathbf{A})\mathbf{A}^{-1} = \left(\frac{1}{k}k\right)\mathbf{A}\mathbf{A}^{-1} = (1)\mathbf{I} = \mathbf{I}.$$

Analogously $\left(\frac{1}{k}\mathbf{A}^{-1}\right)(k\mathbf{A}) = \mathbf{I}$ so $(k\mathbf{A})^{-1} = \frac{1}{k}\mathbf{A}^{-1}$. ■

Example 4.36. Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$ so that $\mathbf{A}^{-1} = \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}$ then

$$\begin{aligned} \mathbf{A}^3 &= \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 8 \\ 4 & 11 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 11 & 30 \\ 15 & 41 \end{bmatrix} \end{aligned}$$

Thus we expect $(\mathbf{A}^{-1})^3$ to be $(\mathbf{A}^3)^{-1}$ (note $\det \mathbf{A}^3 = 1$).

$$\begin{aligned}\mathbf{A}^{-3} = (\mathbf{A}^{-1})^3 &= \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 11 & -8 \\ -4 & 3 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 41 & -30 \\ -15 & 11 \end{bmatrix}\end{aligned}$$

Definition 4.37 (Transpose). Let $\mathbf{A} \in \mathbb{R}^{n\times m}$ be a matrix. The *matrix transpose* of $\mathbf{A}$, denoted $\mathbf{A}^{\mathrm{T}}$, is the $m \times n$ matrix obtained by interchanging the rows and columns of $\mathbf{A}$. That is

$$\mathrm{col}_i(\mathbf{A}^{\mathrm{T}}) = \mathrm{row}_i(\mathbf{A})$$

and in particular $\mathbf{A}^{\mathrm{T}}_{ij} = \mathbf{A}_{ji}$.

Question 4.38. Find the transposes of the following matrices:

$$\mathbf{A} = \begin{bmatrix} a_{00} & \cdots & a_{0m} \\ \vdots & \ddots & \vdots \\ a_{n0} & \cdots & a_{nm} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 2 & 3 \\ 1 & 4 \\ 5 & 6 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1 & 3 & 5 \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} 4 \end{bmatrix}$$

ANSWER.

$$\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} a_{00} & \cdots & a_{0n} \\ \vdots & \ddots & \vdots \\ a_{m0} & \cdots & a_{mn} \end{bmatrix} \quad \mathbf{B}^{\mathrm{T}} = \begin{bmatrix} 2 & 1 & 5 \\ 3 & 4 & 6 \end{bmatrix} \quad \mathbf{C}^{\mathrm{T}} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} \quad \mathbf{D}^{\mathrm{T}} = \begin{bmatrix} 4 \end{bmatrix}$$

◆

Definition 4.39 (Trace). The *trace* of the *square matrix* $\mathbf{A} \in \mathbb{R}^{n\times n}$ with main diagonal

$$\mathbf{A} = \begin{bmatrix} a_{00} & & \\ & \ddots & \\ & & a_{nn} \end{bmatrix}$$

is given by

$$\mathrm{trace}\,(\mathbf{A}) := a_{00} + \cdots + a_{nn}.$$

Example 4.40.

$$\mathbf{B} = \begin{bmatrix} -1 & 2 & 7 & 0 \\ 3 & 5 & -8 & 4 \\ 1 & 2 & 7 & -3 \\ 4 & -2 & 1 & 0 \end{bmatrix} \implies \text{trace}\,(\mathbf{B}) = -1+5+7+0 = 11.$$

Proposition 4.41. Assuming the dimensions of the matrices $\mathbf{A}$ and $\mathbf{B}$ are such that the corresponding operations are well-defined, then

1. $\mathbf{A}^{\mathrm{T}^{\mathrm{T}}} = \mathbf{A}$,
2. $(\mathbf{A}+\mathbf{B})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}} + \mathbf{B}^{\mathrm{T}}$ and $(\mathbf{A}-\mathbf{B})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}} - \mathbf{B}^{\mathrm{T}}$,
3. $(k\mathbf{A})^{\mathrm{T}} = k\mathbf{A}^{\mathrm{T}}$ where k is a scalar, and
4. $(\mathbf{A}\mathbf{B})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$.

OF 4..

$$\begin{aligned} [(\mathbf{AB})^{\mathrm{T}}]_{ij} &= [\mathbf{AB}]_{ji} \\ &= \text{row}_j\mathbf{A} \cdot \text{col}_i\mathbf{B} \\ &= \text{col}_j(\mathbf{A}^{\mathrm{T}}) \cdot \text{row}_i(\mathbf{B}^{\mathrm{T}}) \\ &= \text{row}_i(\mathbf{B}^{\mathrm{T}}) \cdot \text{col}_j(\mathbf{A}^{\mathrm{T}}) \\ &= [\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}]_{ij} \end{aligned}$$

and thus $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$. ■

More generally we have the transpose of a product of any number of matrices is equal to the product of their transposes in reverse order.

Proposition 4.42. Let $\mathbf{A}_0, \ldots, \mathbf{A}_n$ be matrices such that $\mathbf{A}_0 \cdots \mathbf{A}_n$ is a well-defined product. Then

$$(\mathbf{A}_0 \cdots \mathbf{A}_n)^{\mathrm{T}} = \mathbf{A}_n{}^{\mathrm{T}} \cdots \mathbf{A}_0{}^{\mathrm{T}}.$$

PROOF. Exercise. ■

Proposition 4.43. If $\mathbf{A}$ is an invertible matrix then $\mathbf{A}^{\mathrm{T}}$ is also invertible and in particular

$$(\mathbf{A}^{\mathrm{T}})^{-1} = (\mathbf{A}^{-1})^{\mathrm{T}}.$$

PROOF. Notice

$$\mathbf{A}^T(\mathbf{A}^{-1})^T = (\mathbf{A}^{-1}\mathbf{A})^T = \mathbf{I}^T = \mathbf{I}$$

and similarly

$$(\mathbf{A}^{-1})^T\mathbf{A}^T = (\mathbf{A}\mathbf{A}^{-1})^T = \mathbf{I}^T = \mathbf{I}.$$

The result follows. ■

Example 4.44. Consider the inverses of

$$\mathbf{A} = \begin{bmatrix} -5 & -3 \\ 2 & 1 \end{bmatrix} \qquad \mathbf{A}^T = \begin{bmatrix} -5 & 2 \\ -3 & 1 \end{bmatrix}$$

which, in particular, are

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & 3 \\ -2 & -5 \end{bmatrix} \qquad (\mathbf{A}^T)^{-1} = \begin{bmatrix} 1 & -2 \\ 3 & -5 \end{bmatrix}.$$

Notice $(\mathbf{A}^T)^{-1}$ and $(\mathbf{A}^{-1})^T$ are equal.

4.3.2 *A General Method for Finding* $\mathbf{A}^{-1}$

We know how to invert 2×2 matrices, but what about larger ones?

Definition 4.45 (Elementary Matrix). An $n \times n$ matrix is called an *elementary matrix* if it can be obtained from $\mathbf{I}_n$ by performing elementary row operations.

(These correspond exactly to linear systems with single solutions.)

Proposition 4.46. Every elementary matrix is invertible. Thereby all inverses must also be elementary.

Proposition 4.47. If $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a square matrix, then the following statements are equivalent.

1. $\mathbf{A}$ is invertible,
2. $\mathbf{A}\mathbf{x} = \mathbf{0}$ has only the trivial solution (i.e. $\mathbf{x} = \langle 0, \ldots, 0\rangle$),
3. The reduced row-echelon form of $\mathbf{A}$ is $\mathbf{I}_n$, and
4. $\mathbf{A}$ is expressible as a product of elementary matrices.

Proposition 4.48. Let $\mathbf{A}$ and $\mathbf{B}$ be from $\mathbb{R}^{n\times n}$. If

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} & 0 & \cdots & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & \cdots & 0 & b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & b_{n1} & \cdots & b_{nn} \end{bmatrix}$$

then $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$. That is, $\mathbf{A}^{-1} = \mathbf{B}$. (Recall $\mathbf{C} \sim \mathbf{D}$ ($\mathbf{C}$ is "similar" to $\mathbf{D}$) when there are elementary row operations that take $\mathbf{C}$ to $\mathbf{D}$.)

Question 4.49. What is the inverse of $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}$?

ANSWER. First construct

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 3 & 0 & 1 & 0 \\ 1 & 0 & 8 & 0 & 0 & 1 \end{bmatrix}$$

then use row operations to produce the 3×3 identity matrix in first 3 columns.

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 3 & 0 & 1 & 0 \\ 1 & 0 & 8 & 0 & 0 & 1 \end{bmatrix} \begin{array}{l} \rceil^{-2}\ \rceil^{-1} \\ \leftarrow_{+} \\ \leftarrow\!\!-\!\!-_{+} \end{array} = \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & -2 & 5 & -1 & 0 & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & -2 & 5 & -1 & 0 & 1 \end{bmatrix} \begin{array}{l} \\ \rceil^{2} \\ \leftarrow_{+} \end{array} = \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & -1 & -5 & 2 & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & -1 & -5 & 2 & 1 \end{bmatrix} \begin{array}{l} \\ \\ \mid -1 \end{array} = \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{bmatrix} \begin{array}{l} \leftarrow\!\!-\!\!-_{+} \\ \leftarrow_{+} \\ \rfloor_{3}\ \rfloor_{-3} \end{array} = \begin{bmatrix} 1 & 2 & 0 & -14 & 6 & 3 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 2 & 0 & -14 & 6 & 3 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{bmatrix} \begin{array}{l} \leftarrow_{+} \\ \rfloor_{-2} \\ \end{array} = \begin{bmatrix} 1 & 0 & 0 & -40 & 16 & 9 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{bmatrix}.$$

We are finished and thus

$$\mathbf{A}^{-1} = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix}$$

which we can verify

$$\begin{aligned} \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix} &= \begin{bmatrix} -40+32+9 & -80+80+0 & -120+48+72 \\ 13-10-3 & 26-25+0 & 39-15-24 \\ 5-4-1 & 10-10+0 & 15-6-8 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

◆

It is not always possible to invert a matrix.

Question 4.50. What is the inverse of

$$\mathbf{A} = \begin{bmatrix} 1 & 6 & 4 \\ 2 & 4 & -1 \\ -1 & 2 & 5 \end{bmatrix}?$$

ANSWER. We construct

$$\left[\begin{array}{ccc|ccc} 1 & 6 & 4 & 1 & 0 & 0 \\ 2 & 4 & -1 & 0 & 1 & 0 \\ -1 & 2 & 5 & 0 & 0 & 1 \end{array}\right]$$

proceed with elementary row operations on

$$\begin{bmatrix} 1 & 6 & 4 & 1 & 0 & 0 \\ 2 & 4 & -1 & 0 & 1 & 0 \\ -1 & 2 & 5 & 0 & 0 & 1 \end{bmatrix} \begin{array}{l} -2 \quad 1 \\ + \\ + \end{array} = \begin{bmatrix} 1 & 6 & 4 & 1 & 0 & 0 \\ 0 & -8 & -9 & -2 & 1 & 0 \\ 0 & 8 & 9 & -1 & 0 & 1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 6 & 4 & 1 & 0 & 0 \\ 0 & -8 & -9 & -2 & 1 & 0 \\ 0 & 8 & 9 & 1 & 0 & 1 \end{bmatrix} \begin{matrix} \urcorner 1 \\ \hookleftarrow + \end{matrix} = \begin{bmatrix} 1 & 6 & 4 & 1 & 0 & 0 \\ 0 & -8 & -9 & -2 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

And since we have obtained a row of zeros $\mathbf{A}$ cannot be invertible (it was never elementary in the first place). ◆

4.4 Linear Systems

Recall an arbitrary system of m linear equations in n unknowns $x_1, \ldots, x_n$ can be written like

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

or even more compactly as $\mathbf{Ax} = \mathbf{b}$. This means, if $\mathbf{A}$ is invertible, that the solutions to the system can be recovered by multiplying on the left by $\mathbf{A}^{-1}$:

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \implies \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}.$$

(Note $\mathbf{A}$ is invertible only when there is only a single solution to $\mathbf{Ax} = \mathbf{b}$.)

Example 4.51. The polynomial system

$$\begin{aligned} x + 2y + 3z &= 1 \\ 2x + 5y + 3z &= 2 \\ x + 8z &= 3 \end{aligned}$$

can be expressed using matrices as

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Recall,

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}^{-1} = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix}$$

so we have

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 19 \\ -6 \\ -2 \end{bmatrix}$$

which means $(x, y, z) = (10, -6, -2)$ is the solution to the linear system.

4.5 Linear Transformations

We study (vector) functions of the form

$$\begin{aligned} F : \mathbb{R}^n &\to \mathbb{R}^m \\ \mathbf{x} &\mapsto \mathbf{w}. \end{aligned}$$

Example 4.52. Let $F(\mathbf{x})$ be a mapping from $\mathbb{R}^3$ to $\mathbb{R}^2$ given by

$$F\left(\begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_0 \\ x_2 \end{bmatrix}.$$

Then

$$F\left(\begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 7 \end{bmatrix}.$$

Note, from now on, to save vertical space we write $F(x_1, x_1, x_2) = (x_1, x_2)$ instead.

Definition 4.53 (Transform). If f is a function with domain $\mathbb{R}^n$ and codomain $\mathbb{R}^m$ then f is called a *transformation* and we say f *maps* $\mathbb{R}^n$ into $\mathbb{R}^m$. This is denoted by

$$\begin{aligned} f : \mathbb{R}^n &\to \mathbb{R}^m \\ \mathbf{x} &\mapsto \mathbf{w} \end{aligned}$$

One way to give a transform from $\mathbb{R}^n \to \mathbb{R}^m$ is to give m equations from $\mathbb{R}^n \to \mathbb{R}$:

$$w_1 = f_1(x_1, \ldots, x_n)$$

$$\vdots$$
$$w_m = f_m(x_1, \ldots, x_n).$$

That is, if we denote this transform by T and let $\mathbf{x} := (x_1, \ldots, x_n)$, then

$$T(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})).$$

Example 4.54. The equations

$$\begin{aligned} w_1 &= x_0 + x_1 \\ w_2 &= 3x_0x_1 \\ w_3 &= x_0^2 - x_1^2 \end{aligned}$$

define a transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ and in particular

$$T(x_0, x_1) = (x_0 + x_1,\, 3x_0x_1,\, x_0^2 - x_1^2).$$

Thus, for example, $T(1, -2) = (-1, -6, -3)$.

Definition 4.55 (Linear Transform). When a transform T is given by the *linear equations*

$$\begin{aligned} w_1 &= a_{11}x_1 + \cdots + a_{1n}x_n \\ &\vdots \\ w_m &= a_{m1}x_1 + \cdots + a_{mn}x_n \end{aligned}$$

or, equivalently, the matrix expression

$$\begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

(i.e. $\mathbf{w} = \mathbf{A}\mathbf{x}$) the transform is called a *Linear Transform*.

Definition 4.56 (Standard Matrix). Let $\mathbf{A} \in \mathbb{R}^{n \times m}$ define a linear system which itself defines a linear transformation. This matrix $\mathbf{A}$ is called the *standard matrix*.

Example 4.57. The linear transformation defined by the equations

$$\begin{aligned} w_1 &= 2x_1 - 3x_2 + x_3 - 5x_4 \\ w_2 &= 4x_1 + x_2 - 2x_3 + x_4 \end{aligned} \tag{4.1}$$

$$w_3 = 5x_1 - x_2 + 4x_3$$

can be expressed in matrix from as

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 2 & -3 & 1 & -5 \\ 4 & 1 & -2 & 1 \\ 5 & -1 & 4 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}. \tag{4.2}$$

Thus the *image* of a point $\mathbf{x}$ can be computed using (4.1) or (4.2). For example, let $\mathbf{w} = (1, 3, 8)$, then

$$w_1 = 2(1) - 3(-3) + (0) - 5(2) = 1 \qquad w_2 = 3 \qquad w_3 = 8$$

or equivalently

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 2 & -3 & 1 & -5 \\ 4 & 1 & -2 & 1 \\ 5 & -1 & 4 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix}.$$

Notation. When $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transform given by multiplication by $\mathbf{A}$ we sometimes denote T by

$$T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$$

to emphasis the standard matrix. Thus we have

$$T_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$$

where $\mathbf{x}$ is understood to be a column matrix.

Example 4.58. Let 0 denote the zero-matrix and $\mathbf{0}$ the zero-vector in $\mathbb{R}^n$, then for any vector $\mathbf{x} \in \mathbb{R}^n$

$$T_0(\mathbf{x}) = 0\mathbf{x} = \mathbf{0}.$$

We call T_0 the *zero transformation*.

Example 4.59. Let $\mathbf{I}$ denote the identity-matrix, then for any vector $\mathbf{x} \in \mathbb{R}^n$

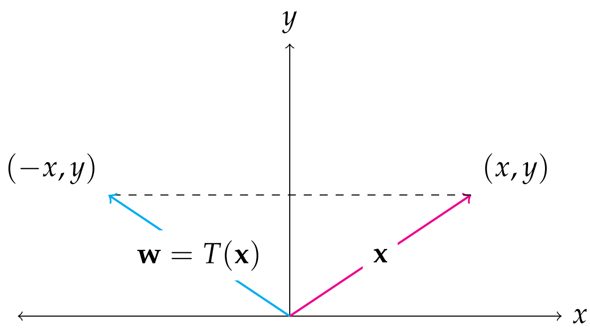
$$T_{\mathbf{I}}(\mathbf{x}) = \mathbf{I}\mathbf{x} = \mathbf{x}.$$

We call $T_{\mathbf{I}}$ the *the identity transform* on $\mathbb{R}^m$.

4.5.1 Transforming points in $\mathbb{R}^n$

Among the most important linear operations on $\mathbb{R}^2$ and $\mathbb{R}^3$ are those that rotate, reflect project, and rotate points in Euclidean space.

Question 4.60. Consider an operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ that maps each point into its reflection about the y-axis:



What is the *standard matrix* for this transform?

ANSWER. Recall we want $\mathbf{A}$ such that $T_{\mathbf{A}}(x, y) = (y, x) = \mathbf{w}$. A linear system which does this reflection is given by

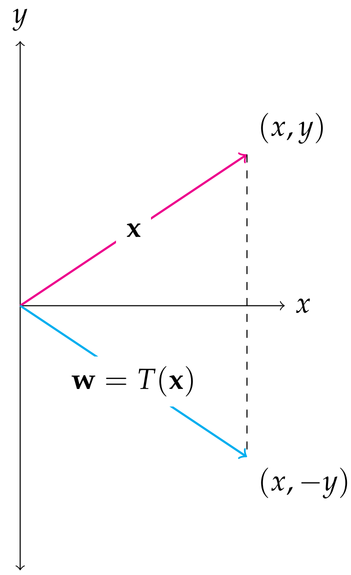
$$\begin{aligned} w_1 &= -x + 0y \\ w_2 &= 0x + y. \end{aligned}$$

corresponding to the standard matrix $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$. Notice (just to confirm)

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x \\ y \end{bmatrix}.$$

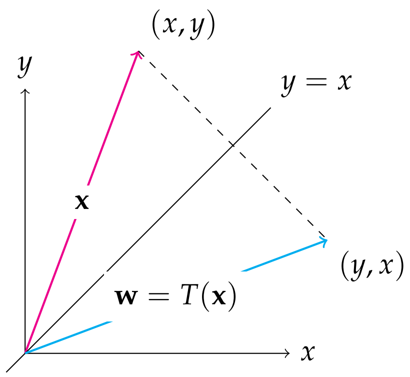
◆

Proposition 4.61 (Reflection about x-axis)**.**



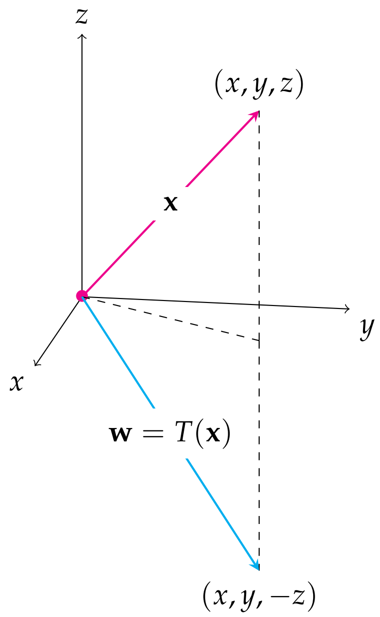
$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix}$$

Proposition 4.62 (Reflection about the $y = x$ line)**.**



$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$$

Proposition 4.63 (Reflection about the xy plane)**.**



$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ -z \end{bmatrix}.$$

Similarly we have a reflection about the xz plane

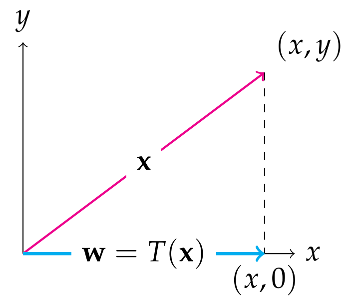
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ -y \\ z \end{bmatrix}$$

and about the yz plane

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -x \\ y \\ z \end{bmatrix}.$$

4.5.2 Projections

Question 4.64. Consider an operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ that maps each vector into its projection into the x-axis:



What is the *standard matrix* for this transform?

ANSWER. Recall we want $\mathbf{A}$ such that $T_{\mathbf{A}}(x, y) = (x, 0) = \mathbf{w}$. A linear system which does this projection is given by

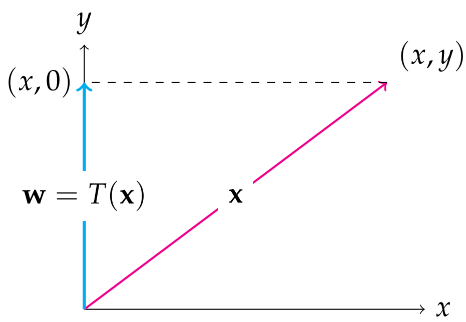
$$w_1 = x + 0y$$
$$w_2 = 0x + 0y.$$

corresponding to the standard matrix $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. Notice (just to confirm)

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix}.$$

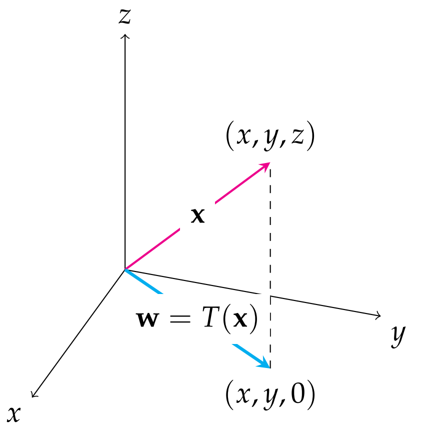
◆

Proposition 4.65 (Projection into y-axis)**.**



$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}.$$

Proposition 4.66 (Projection into the xy plane)**.**



$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}.$$

Similarly we have projection into the xz plane

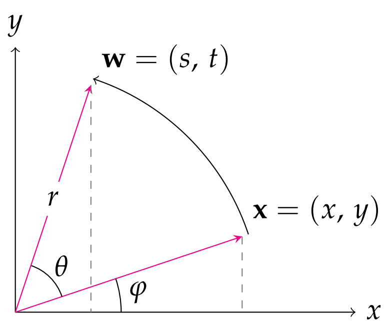
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ 0 \\ z \end{bmatrix}$$

and projection into the yz plane

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ y \\ z \end{bmatrix}.$$

4.5.3 Rotations

Definition 4.67 (Rotation)**.** A *transform* which *rotates* each vector in $\mathbb{R}^2$ through by a fixed angle θ is called a *rotation transform* on $\mathbb{R}^2$.



Suppose $|\mathbf{w}| = |\mathbf{v}| = r$ then, from basic trig, we have

$$(x, y) = (r\cos\varphi,\, r\sin\varphi) \qquad (s, t) = (r\cos[\theta+\varphi],\, r\sin[\theta+\varphi])\,.$$

The appropriate trig identities produces further that

$$(s, t) = (r\cos\theta\cos\varphi - r\sin\theta\sin\varphi,\ r\sin\theta\cos\varphi + r\cos\theta\sin\varphi). \tag{4.3}$$

Recalling that $(x, y) = (r\cos\varphi,\, r\sin\varphi)$, we can reduce (4.3) to

$$(s, t) = (x\cos\theta - y\sin\theta,\ x\sin\theta + y\cos\theta).$$

Finally, since these are linear equations in x and y, we can write

$$\begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Producing a standard matrix for the rotation of $\mathbf{x}$ into $\mathbf{w}$.

Proposition 4.68 (Rotation in $\mathbb{R}^2$)**.** To rotate a point $(x, y) \in \mathbb{R}^2$ through by an angle θ apply the linear transformation

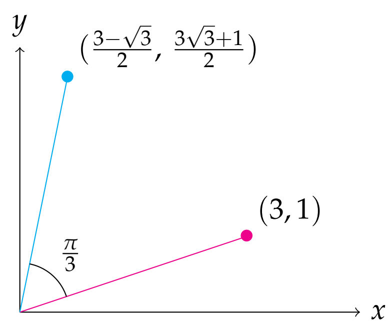
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} s \\ t \end{bmatrix}.$$

PROOF. Above. ■

Question 4.69. What is the image of $(3, 1)$ after rotating the point about the origin by $\frac{\pi}{3}$?

ANSWER.

$$\begin{bmatrix} \cos\frac{\pi}{3} & -\sin\frac{\pi}{3} \\ \sin\frac{\pi}{3} & \cos\frac{\pi}{3} \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{3-\sqrt{3}}{2} \\ \frac{3\sqrt{3}+1}{2} \end{bmatrix} \approx \begin{bmatrix} 0.63 \\ 3.10 \end{bmatrix}.$$



◆

Proposition 4.70. The *counter-clockwise* rotation by θ radians about the x, y, or z axis in $\mathbb{R}^3$ is (respectively) given by the standard matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \quad \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \quad \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

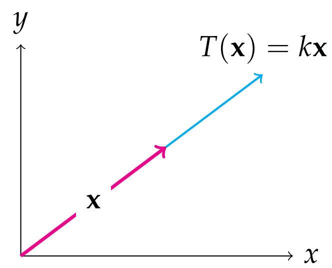
Theorem 4.71. The *counter-clockwise* rotation about the arbitrary *unit vector* $\hat{\mathbf{u}} = \langle a, b, c \rangle$ is given by

$$\begin{bmatrix} a^2(1-\cos\theta)+\cos\theta & ab(1-\cos\theta)-c\sin\theta & ac(1-\cos\theta)+b\sin\theta \\ ab(1-\cos\theta)+c\sin\theta & b^2(1-\cos\theta)+\cos\theta & bc(1-\cos\theta)-a\sin\theta \\ ac(1-\cos\theta)-b\sin\theta & bc(1-\cos\theta)+a\sin\theta & c^2(1-\cos\theta)+\cos\theta \end{bmatrix}.$$

PROOF. ■

4.5.4 Dilations and Contractions

Question 4.72. Consider an operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ which *contracts* or *dilates* (i.e. scales positively or negatively) a vector $\mathbf{x}$ by some scalar k.



What is the *standard matrix* for this transform?

ANSWER. Recall we want $\mathbf{A}$ such that $T_{\mathbf{A}}(x, y) = (kx, ky) = \mathbf{w}$. A linear system which does this scaling is

$$w_1 = kx + 0y$$
$$w_2 = 0x + ky$$

corresponding to the standard matrix $\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$ Notice (just to confirm)

$$\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} kx \\ ky \end{bmatrix}.$$

◆

Proposition 4.73. The *scaling* of a vector $\mathbf{x}$ in $\mathbb{R}^n$ by the scalar k is given by the linear transformation with standard matrix

$$\begin{bmatrix} k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & k \end{bmatrix}.$$

4.5.5 Combining Linear Transforms

Definition 4.74 (Composition). If $T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^k$ and $T_{\mathbf{B}} : \mathbb{R}^k \to \mathbb{R}^m$ are linear transforms then the *composition of* $T_{\mathbf{B}}$ *with* $T_{\mathbf{A}}$ is the function $T_{\mathbf{B}} \circ T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ satisfying

$$(T_{\mathbf{B}} \circ T_{\mathbf{A}})(\mathbf{x}) = T_{\mathbf{B}}(T_{\mathbf{A}}(\mathbf{x})).$$

Note the order of the transformations. First $\mathbf{x}$ is transformed by $T_{\mathbf{A}}$ then this image is transformed by $T_{\mathbf{B}}$.

Proposition 4.75. The composition $T_{\mathbf{B}} \circ T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transform and in particular

$$T_{\mathbf{B}} \circ T_{\mathbf{A}} = T_{\mathbf{BA}}.$$

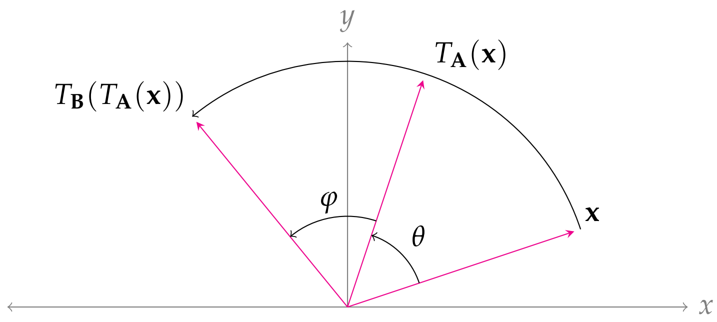
Note this result says we can do a multi-step linear transformation with a single standard matrix.

PROOF.

$$(T_{\mathbf{B}} \circ T_{\mathbf{A}})(\mathbf{x}) = T_{\mathbf{B}}(T_{\mathbf{A}}(\mathbf{x})) = T_{\mathbf{B}}(A\mathbf{x}) = B(A\mathbf{x}) = (BA)\mathbf{x}.$$

■

To test Proposition 4.75 let us find a single standard matrix which does a rotation by θ then φ. We should find this standard matrix is equal to that of the standard matrix for rotation by $\theta + \varphi$.



Example 4.76. Let $T_{\mathbf{A}} : \mathbb{R}^2 \to \mathbb{R}^2$ and $T_{\mathbf{B}} = \mathbb{R}^2 \to \mathbb{R}^2$ be linear transformations which rotate a point by θ and φ respectively. By Proposition 4.75 we have

$$T_{\mathbf{B}} \circ T_{\mathbf{A}} = T_{\mathbf{BA}}$$

where

$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}.$$

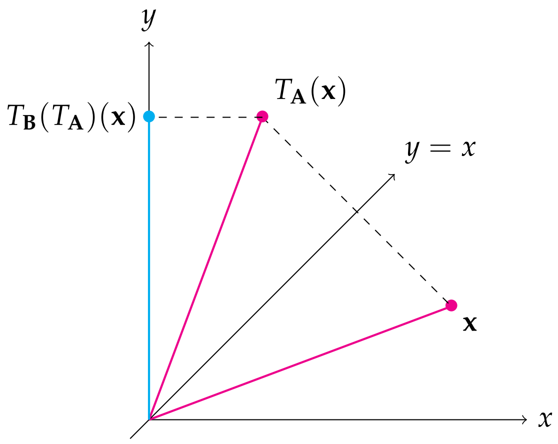
To obtain the standard matrix for rotation by $\theta + \varphi$ we do

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \\ &= \begin{bmatrix} \cos\varphi\cos\theta - \sin\varphi\sin\theta & -\cos\varphi\sin\theta - \sin\varphi\cos\theta \\ \sin\varphi\cos\theta + \cos\varphi\sin\theta & -\sin\varphi\sin\theta + \cos\varphi\cos\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos(\theta+\varphi) & -\sin(\theta+\varphi) \\ \sin(\theta+\varphi) & \cos(\theta+\varphi) \end{bmatrix} \end{aligned}$$

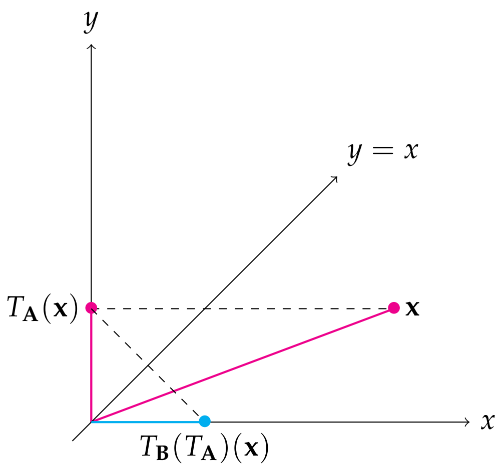
Notice that $\mathbf{AB}$ is a rotation by $\theta + \varphi$ as we have defined.

The order in which linear transformations are composed matters. This should be obvious as we know matrix multiplication is not commutative.

Example 4.77. Let $T_{\mathbf{A}} : \mathbb{R}^2 \to \mathbb{R}^2$ be a reflection about the $y = x$ line and $T_{\mathbf{B}} : \mathbb{R}^2 \to \mathbb{R}^2$ be the orthogonal projection into the y-axis. The following figures illustrates that $(T_{\mathbf{A}} \circ T_{\mathbf{B}})(\mathbf{x}) \neq (T_{\mathbf{B}} \circ T_{\mathbf{A}})(\mathbf{x})$ and thereby $\mathbf{AB} \neq \mathbf{BA}$.



$$T_{\mathbf{B}} \circ T_{\mathbf{A}} = T_{\mathbf{C}} \text{ where } \mathbf{C} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

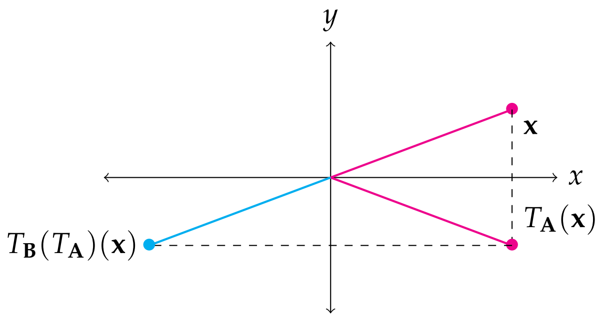


$$T_{\mathbf{B}} \circ T_{\mathbf{A}} = T_{\mathbf{C}} \text{ where } \mathbf{C} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

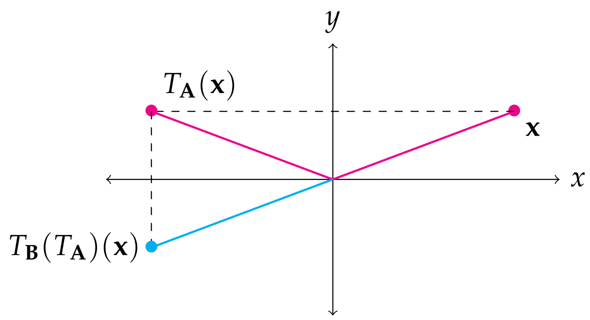
(Note the difference.)

Composing two linear transforms *can* yield the same result independent of the ordering of the composition.

Example 4.78. Let $T_{\mathbf{A}} : \mathbb{R}^2 \to \mathbb{R}^2$ be a reflection about the y axis and $T_{\mathbf{B}} : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection about the x axis. The following figures illustrates that $(T_{\mathbf{A}} \circ T_{\mathbf{B}})(\mathbf{x}) = (T_{\mathbf{B}} \circ T_{\mathbf{A}})(\mathbf{x})$ and thereby $\mathbf{AB} = \mathbf{BA}$.



$$T_{\mathbf{B}} \circ T_{\mathbf{A}} = T_{\mathbf{C}} \text{ where } \mathbf{C} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$



$$T_{\mathbf{B}} \circ T_{\mathbf{A}} = T_{\mathbf{C}} \text{ where } \mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

(Note they are the same.)

We can compose as many linear transforms together as we like.

Theorem 4.79. Suppose we have m linear transforms given by

$$T_{\mathbf{A}_m} : \mathbb{R}^{n_{m-1}} \to \mathbb{R}^{n_m},$$

that is

$$\mathbb{R}^{n_0} \xrightarrow{T_{\mathbf{A}_1}} \mathbb{R}^{n_1} \xrightarrow{T_{\mathbf{A}_2}} \cdots \mathbb{R}^{n_{m-1}} \xrightarrow{T_{\mathbf{A}_m}} \mathbb{R}^{n_m}$$

where $n_0, \ldots, n_m$ is a sequence of nonzero naturals. We have that

$$T_{\mathbf{A}_m} \circ T_{\mathbf{A}_{m-1}} \circ \cdots \circ T_{\mathbf{A}_m} = T_{\mathbf{A}_m \mathbf{A}_{m-1} \cdots \mathbf{A}_1}$$

is a transform from $\mathbb{R}^{n_0} \to \mathbb{R}^{n_m}$.

Question 4.80. Find the standard matrix for a linear operator that

1. rotates counterclockwise about the z-axis by θ, then

2. reflects the resulting vector about the z, and then

3. projects orthogonally onto the xy-plane.

ANSWER. The standard matrices corresponding to the listed transforms are (respectively)

$$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

And multiplying them (in reverse order) gives

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -\cos\theta & \sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

◆

Definition 4.81 (one-to-one). A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is *one-to-one* if T maps distinct points in $\mathbb{R}^n$ to distinct points in $\mathbb{R}^m$. Namely, T is one-to-one when

$$T(\mathbf{x}) = T(\mathbf{y}) \iff \mathbf{x} = \mathbf{y}.$$

Example 4.82. The linear transformation rotating a point by θ in $\mathbb{R}^2$ *is* one-to-one.

Example 4.83. Projections are *not* one-to-one transformations because lots of points project into the same point.

Theorem 4.84. If $\mathbf{A}$ is an $n \times n$ matrix and $T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transform, then the following are equivalent

1. $\mathbf{A}$ is invertible,

2. The range of $T_{\mathbf{A}}$ is $\mathbb{R}^n$, and

3. $T_{\mathbf{A}}$ is one-to-one.

Let us confirm that rotation in $\mathbb{R}^2$ satisfies the last Theorem.

Example 4.85. The linear transform for rotation in $\mathbb{R}^2$ is given by the standard matrix

$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

It is clear this transform is one-to-one (every point can be rotated to and no two distinct point can rotate to the same place). It suffices to check $\mathbf{A}$ is invertible. Notice though, that

$$\det \mathbf{A} = \cos^2\theta + \sin^2\theta = 1 \neq 0.$$

Example 4.86. The linear transform for projection into the x-axis in $\mathbb{R}^2$ is given by the standard matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Clearly $\mathbf{A}$ is *not* one-to-one as many points project into the same place on the x-axis. This means $\mathbf{A}$ is non-invertible, as we confirm:

$$\det \mathbf{A} = 0.$$

Proposition 4.87. Let $T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The *inverse transform* $(T_{\mathbf{A}})^{-1}$ is given by $\mathbf{A}^{-1}$ provided this standard matrix is invertible.

$$(T_{\mathbf{A}})^{-1} = T_{\mathbf{A}^{-1}}.$$

PROOF. Let $(T_{\mathbf{A}})^{-1} = T_{\mathbf{B}}$. Then, we know

$$\begin{aligned} T_{\mathbf{B}}(T_{\mathbf{A}}(\mathbf{x})) &= \mathbf{x} \\ \iff T_{\mathbf{BA}}(\mathbf{x}) &= \mathbf{x} \\ \iff \mathbf{BAx} &= \mathbf{x} \\ \iff \mathbf{BA} &= \mathbf{I}. \end{aligned}$$

A similar argument gives $\mathbf{AB} = \mathbf{I}$ and thereby $\mathbf{B} = \mathbf{A}^{-1}$ by definition. ■

Question 4.88. Let $T_{\mathbf{A}} : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transform rotating each point of $\mathbb{R}^2$ by θ. What is the inverse transform $(T_{\mathbf{A}})^{-1}$?

ANSWER. It is evident geometrically that to invert a rotation by θ is to do another rotation by $-\theta$.

We have

$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

and

$$\mathbf{A}^{-1} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix}$$

so the standard matrix for $(T_{\mathbf{A}})^{-1}$. (Check $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$.) ◆

Question 4.89. Show that the linear operator $T_{\mathbf{A}} : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}$$

is one-on-one, and find $(T_{\mathbf{A}})^{-1}(\mathbf{y})$.

ANSWER. Notice $\det \mathbf{A} = 5$ implies $\mathbf{A}$ is invertible so by our Theorem $T_{\mathbf{A}}$ is one-to-one.

In particular $\mathbf{A}^{-1} = \begin{bmatrix} \frac{4}{5} & -\frac{1}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{bmatrix}$ and

$$\begin{bmatrix} \frac{4}{5} & -\frac{1}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{4}{5}y_1 - \frac{1}{5}y_2 \\ -\frac{3}{5}y_1 - \frac{3}{5}y_2 \end{bmatrix}$$

thus

$$(T_{\mathbf{A}})^{-1}(\mathbf{y}) = \left(\frac{4}{5}y_1 - \frac{1}{5}y_2, -\frac{3}{5}y_1 - \frac{3}{5}y_2\right).$$

◆

Theorem 4.90. A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation *if and only if* the following hold for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and scalar $c \in \mathbb{R}$

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$, and
2. $T(c\mathbf{u}) = cT(\mathbf{u})$.

PROOF. ■