Contents

Chapter 1

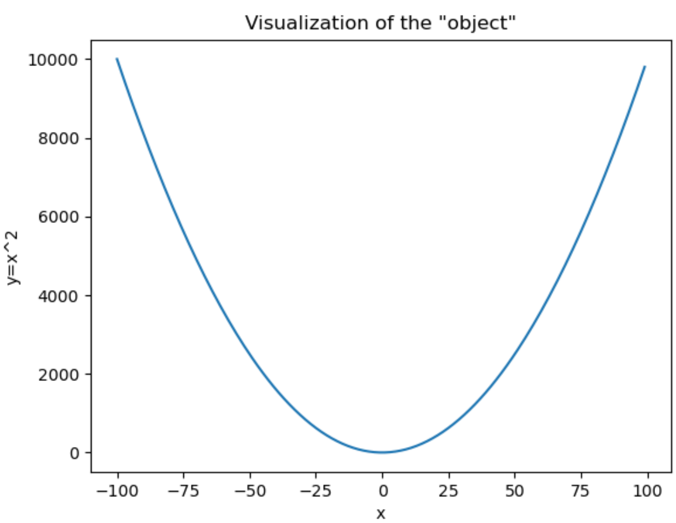
Coordinate Geometry

1.1 Introduction

Many of you have encountered some form of coordinate geometry in high school. For instance, the “standard” way to visualize a graph e.g. $f(x) = x^2$ is to visualize the points in 2-D space (x, y) where $y = x^2$. We give a demonstration in Python code.

```
import matplotlib.pyplot as plt
import numpy as np
X=np.arange(-100,100) #create list of numbers from -100 to 100
Y= X**2 #calculate the square of at each x
plt.plot(X,Y) #plot all the pairs of points in 2d plane
plt.xlabel('x')
plt.ylabel('y=x^2')
plt.title('Visualization of the "object"')
plt.show()
plt.show()
```

$x \longrightarrow x$
the “graph” object



This is also known as the Cartesian plane, named after René Descartes who invented it in the 17th century.

1.2 Visualization of geometric objects

The Cartesian plane allows us to describe shapes with equations and perform calculations with them. We first define the playing field (the Cartesian plane and higher dimensional analogues) and the players.

Definition 1.1 (Real numbers)

The set of **real numbers**, denoted as $\mathbb{R}$, is (informally) the set of all the numbers that can be written out in decimal form.

Example 1.2

The following are real numbers:

1. The integers 0, ± 1, ± 2, ...
2. Fractions in the form $\frac{a}{b}$, where a and $b \neq 0$ are integers.
3. Irrational numbers $\sqrt{2}$, π.

Remark. *The set of real numbers is known as a* ***complete field****. The definition of a complete field will be swept under the rug, but it guarantees a few things. The most important property: We will not "escape" the set by performing operations, possibly infinitely many.*

Definition 1.3 (N-dimensional space)

Let n be a positive integer. We denote the **n-dimensional real space** to be $\mathbb{R}^n$, consisting of all the n-tuples $(x_1, x_2, x_3, ..., x_n)$, where each x_j is a real number.
We call an n-tuple $(x_1, ..., x_n)$ a **point**, and two points $(x_1, ..., x_n)$ and $(y_1, ..., y_n)$ are equal if $x_j = y_j$ for all j-th entries of the tuples.

Remark. *We sometimes use* $\mathbf{x}$ *to denote* $(x_1, ..., x_n)$ *to make notation cleaner.*

1.2.1 Lines

Now that we have introduced the playing field of n-dimensional space, we can start translating the axioms of euclidean geometry to this coordinate system.

Definition 1.4 (Lines)

In euclidean geometry, a line is defined by two points. We let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. The **line** going from $\mathbf{x}$ to $\mathbf{y}$ is denoted $\overrightarrow{\mathbf{xy}}$.

What would this line look like? To get from $\mathbf{x}$ to $\mathbf{y}$, we have to traverse $y_1 - x_1$ units in the first coordinate, $y_2 - x_2$ units in the second, ..., $y_n - x_n$ in the last. We thus have a natural notation for the line $\overrightarrow{\mathbf{xy}}$.

$$\overrightarrow{\mathbf{xy}} = (y_1 - x_1, y_2 - x_2, ..., y_n - x_n).$$

This is very similar to a point as an n-tuple, but this is "spiritually" different to a point. This tuple represents the direction of line. One way to think of the correspondence between $(x_1, ..., x_n)$

point and $(x_1, ..., x_n)$ line is that $(x_1, ..., x_n)$ line is the line connecting $(0, 0, ...0)$ to $(x_1, ..., x_n)$ point. Because of this, we can identify a tuple as both the point and the line, and we call it a "vector" to abstract away from the actual geometric meaning.

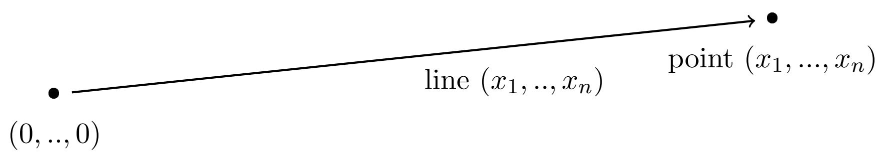


Figure 1.1: The correspondence between a line "vector" and a point "vector".

Notation. *Vectors are a very general notation of* $n-tuples$. *Depending on context, we use both of the following notations to denote the entries of* $\vec{v} \in \mathbb{R}^n$

- *"Ordered sets"* $(v_1, ..., v_n)$, *suitable dealing with points (and other geometric objects). It also looks cleaner when writing inline.*
- *"Column vectors"* $\begin{bmatrix} v_1 \\ . \\ . \\ . \\ v_n \end{bmatrix}$ *when the ordered sets are complicated to read, and when working with matrix algebra.*

1.2.2 Operation with lines

We need to translate a few more things from euclidean geometry.

Proposition 1.5

Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, then

$$\overrightarrow{\mathbf{xy}} + \overrightarrow{\mathbf{yz}} = \overrightarrow{\mathbf{xz}},$$

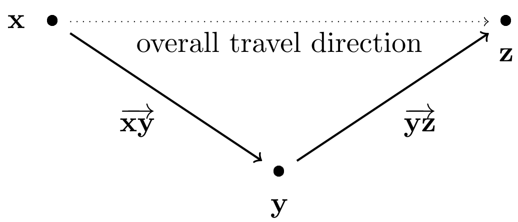
where $(a_1, ..., a_n) + (b_1, ..., b_n) \stackrel{\text{def}}{=} (a_1 + b_1, ..., a_n + b_n)$.

Proof.

$$\begin{aligned} \overrightarrow{\mathbf{xy}} + \overrightarrow{\mathbf{yz}} &= (y_1 - x_1, ..., y_1 - x_n) + (z_1 - y_1, ..., z_1 - y_n) \\ &= (y_1 - x_1 + z_1 - y_1, ..., y_n - x_n + z_n - y_n) \\ &= (z_1 - x_1, ..., z_n - x_n) \\ &= \overrightarrow{\mathbf{xz}} \end{aligned}$$

□

Geometrically, this means if you connect $\mathbf{x}$ to $\mathbf{y}$ to $\mathbf{z}$, the overall "direction of travel" you make is $\mathbf{x}$ to $\mathbf{z}$. This gives us a natural extension for addition of vectors by considering each entry. Similarly for scaling vectors, we just scale the entries along each dimension.



Definition 1.6 (Addition and scaling of vectors)

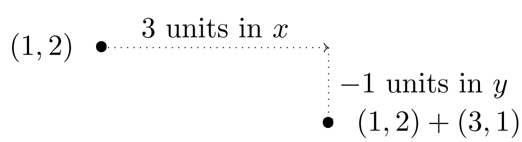
Let $\vec{a}, \vec{b}$ be two vectors in $\mathbb{R}^n$. We define the sum/difference of $\vec{a}$ and $\vec{b}$

$$\vec{a} + \vec{b} \stackrel{\text{def}}{=} (a_1 + b_1, ..., a_n + b_n), \quad \vec{a} - \vec{b} \stackrel{\text{def}}{=} (a_1 - b_1, ..., a_n - b_n)$$

and the scaling of $\vec{a}$ by a real number $c \in \mathbb{R}$

$$c\vec{a} \stackrel{\text{def}}{=} (ca_1, ca_2, ..., ca_n).$$

Remark. *Here we use the term "vectors", as we can in essence add points and lines together. How does one make sense of adding a line to a point? We can view this as translating the point along the path of the line, for instance, let us translate the point* $(1, 2)$ 3 *units in the first coordinate and* -1 *units in the second coordinate. This will give us* $(4, 1)$.



This way, we can write the line from $\vec{x}$ to $\vec{y}$ as $\vec{y} - \vec{x}$. The proof is a computational exercise.

Notation. *We now transferred for talking about points and the lines between points to addition. Therefore, we can overload the notation for points and lines as a vector* $\vec{v}$, *keeping in mind that they have the same arithmetic structure.*

In fact, most of our intuition for the real numbers translates to $\mathbb{R}$. For formality, we will list them here; in practice, we (almost always) take these properties for granted.

Proposition 1.7

Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$, and $a, b \in \mathbb{R}$. Then the following hold:

- *(Associativity)* $(\vec{u}+\vec{v})+\vec{w} = \vec{u}+(\vec{v}+\vec{w})$.
- *(Commutativity)* $\vec{u}+\vec{v} = \vec{v}+\vec{u}$.
- *(Identity)* The zero vector $\vec{0} \stackrel{\text{def}}{=} (0,0,...,0) \in \mathbb{R}$ satisfies $\vec{v}+\vec{0} = \vec{v}$.
- *(Inverse)* The inverse of $\vec{v}$, $-\vec{v} \stackrel{\text{def}}{=} (-v_1, ..., -v_n)$ satisfies $\vec{v}+(-\vec{v}) = \vec{0}$.
- *(Scalar multiplication)* $a(b\vec{v}) = (ab)\vec{v}$.
- *(Scalar Identity)* $1\vec{v} = \vec{v}$.
- *(Distributivity 1)* $a(\vec{u}+\vec{v}) = a\vec{u}+a\vec{v}$.
- *(Distributivity 2)* $(a+b)\vec{v} = a\vec{v}+b\vec{v}$.

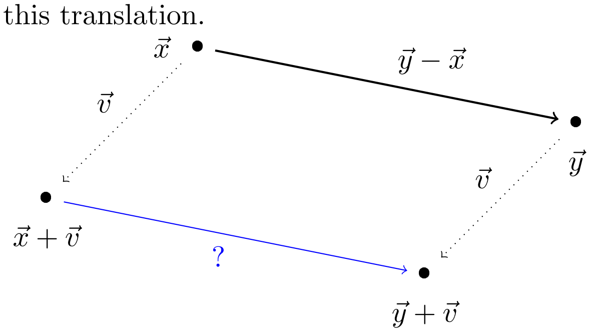
Remark. *All of these have good geometric intuition behind. For instance, the zero vector $\vec{0}$ is the "don't move" vector, corresponding to the point at the origin, or the "too short to be a line". The inverse of $\overrightarrow{\mathbf{xy}}$ is $\overrightarrow{\mathbf{yx}}$, where you go back from $\mathbf{x}$ to $\mathbf{y}$.*

Remark. *These 8 conditions are the axioms of a vector space. Later in the course, we will generalize the notion of vectors in $\mathbb{R}^n$ to other spaces (playing fields).*

Proposition 1.8

Lines are translation invariant. That is, for every $\vec{x}, \vec{y}, \vec{v} \in \mathbb{R}^n$, then the line from $\vec{x}$ to $\vec{y}$ is the same as the line from $\vec{x}+\vec{v}$ to $\vec{y}+\vec{v}$.

Let us illustrate what this statement is trying to convey. We have two points $\vec{x}$, $\vec{y}$; now we translate each of these points by $\vec{v}$, and we want the line between the points to be preserved under this translation.



The proof is one line: $(\vec{y}+\vec{v})-(\vec{y}+\vec{v}) = \vec{y}-\vec{x}+\vec{v}-\vec{v} = \vec{y}-\vec{x}$. However, an immediate consequence of this is that we can "transport" vectors in space without distorting the vector. Colloquially, *5 miles South* to you describes the same direction and length as *5 miles South* to a person a few feet away. This justifies the way we visualize the correspondence between points and vectors - we "transport" the vectors to start from the origin $(0, ..., 0)$, and the end describes the point.

Remark. *Translation (and scaling) invariance is a property of Euclidean geometry. There are some exotic geometry systems that distort distance and direction through translation and scaling. One such example is the Poincaré metric.*

Another notion we can carry from Euclidean geometry is parallel lines. Here we not only define what it means for two vectors to be parallel (never touching), we also give a definition for two vectors to be parallel but point in opposite directions.

Definition 1.9 (Parallel and Antiparallel Vectors)

Let non-zero vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$. We say that $\vec{v}$ and $\vec{w}$ are **parallel** if there is some $c > 0$ such that $\vec{v} = c\vec{w}$. We say that $\vec{v}$ and $\vec{w}$ are **antiparallel** if there is some $c < 0$ such that $\vec{v} = c\vec{w}$.

Example 1.10

Find an expression for the points on the (infinite) line passing through $P(1, 1, 0)$ and $Q(0, 2, 2)$.

Let M be a point on the line $\overrightarrow{PQ}$, then $\overrightarrow{PM}$ is parallel to $\overrightarrow{PQ}$ (or $M = P$). Either way, we would have some $c \in \mathbb{R}$ such that

$$\overrightarrow{PM} = c\overrightarrow{PQ}.$$

Indeed, we can confirm that any point represented as $P + c\overrightarrow{PQ}$ will be on the line, as $c\overrightarrow{PQ}$ is parallel to $\overrightarrow{PQ}$. One exception is to check for when $c = 0$, but we already know P is on the line. We thus have an expression for the line to be

$$(1, 1, 0) + c(-1, 1, 2)$$

for all $c \in \mathbb{R}$. Another way to put it, the line is the **image** of the function $f(c) = (1, 1, 0) + c(-1, 1, 2)$ on $\mathbb{R}$.

Remark. *This form of representing a line is known as a* ***parametric representation*** *or a* ***parametrization*** *of the line. We will give a more concrete definition of parametrization when we move beyond straight lines to "curvy" curves and surfaces in* **??**.

Remark. *Using the parametric representation for a line $l(t) = (P_1, P_2, ..., P_n) + t(v_1, v_2, ..., v_n)$, we can take slices/cross sections across each of the j-th coordinates, and get*

$$x_j = P_j + tv_j \implies t = \frac{x_j - P_j}{v_j}$$

is the equation of a line in 2d. P_j, v_j are fixed, viewing in the variables $y = x_j, x = t$, this is the equation of a line with y-intercept x_j and slope v_j. Setting the values of t for all slices to be equal,

$$t = \frac{x_1 - P_1}{v_1} = \frac{x_2 - P_2}{v_2} = ... = \frac{x_j - P_j}{v_j} = ... = \frac{x_n - P_n}{v_n}$$

This form is known as the ***symmetric equations*** *of a line.*

Example 1.11

Find an expression for the points on the line connecting **between** $P(1, 1, 0)$ to $Q(0, 2, 2)$.

This will be a segment of the line from the previous example, so the answer would be the same expression, but we limit the domain of f to be an interval on $\mathbb{R}$. Let us examine what f does to a few values of c.

c	very negative	$c = 0$	$c = 1/2$	$c = 1$	very positive
f(c)	very off the segment	P	midpoint of $\overline{PQ}$	Q	very off the segment

As we move from very negative c to very positive c, we start very away from the line segment, reach P at $c = 0$, Q at $c = 1$, then move away from the line segment. Indeed the values $(1, 1, 0) + c(-1, 1, 2)$ will be on the line segment for $c \in [0, 1]$.

Finally, the nice algebraic properties for adding and scaling vectors gives us a natural way to understand these vectors as a sum of n component vectors, one for each dimension. These *special* vectors deserve a name.

Definition 1.12 (Standard Basis Vectors and Vector Decomposition)

We denote in $\mathbb{R}^n$, the **standard basis vectors**

$$\begin{aligned} \vec{e}_1 &\stackrel{\text{def}}{=} (1, 0, 0, ..., 0, 0, 0), \\ \vec{e}_2 &\stackrel{\text{def}}{=} (0, 1, 0, ..., 0, 0, 0), \\ \vec{e}_{n-1} &\stackrel{\text{def}}{=} (0, 0, 0, ..., 0, 1, 0), \\ \vec{e}_n &\stackrel{\text{def}}{=} (0, 0, 0, ..., 0, 0, 1), \end{aligned}$$

so that $\vec{v} = \sum_{i=1}^n v_i \vec{e}_i$. In $\mathbb{R}^3$, we sometimes write

$$\vec{i} \stackrel{\text{def}}{=} \vec{e}_1, \quad \vec{j} \stackrel{\text{def}}{=} \vec{e}_2, \quad \vec{k} \stackrel{\text{def}}{=} \vec{e}_3$$

to accommodate the physicists.

Exercises

1. Compute $5\vec{v} - 2\vec{w}$ and $-3\vec{w}$ for the following pairs of vectors:

   (a) $\vec{v} = 2\vec{i} + 3\vec{j}$, $\vec{w} = 4\vec{i} - 9\vec{j}$

   (b) $\vec{v} = (1, 2, -1)$, $\vec{w} = (2, -1, 0)$

   (c) $\vec{v} = -2\vec{e}_3 + 4\vec{e}_5$, $\vec{w} = \vec{e}_1 - 4\vec{e}_5$

   (d) $\vec{v} = (\cos t, \sin t)$, $\vec{w} = (\cos t)\vec{e}_2 - (\sin t)\vec{e}_1$

2. Find parametric equations that describe the line that

   (a) passes through $(1, 2, 3)$ and $(4, -1, 2)$

   (b) passes through $(0, -2, 3)$ and $(1, 2, 3)$

1.3 Length, angles and projections

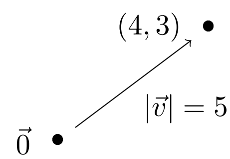
Definition 1.13 (Magnitude)

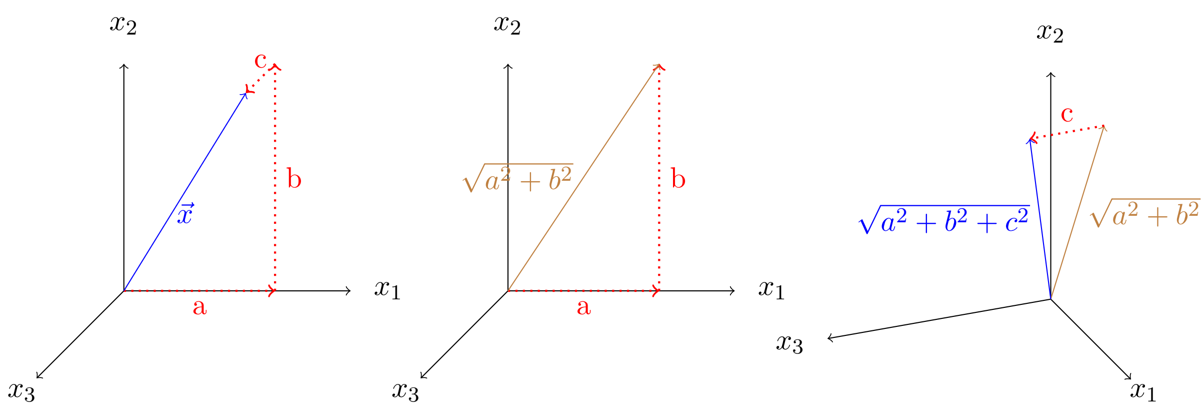
Let $\vec{v} \in \mathbb{R}^n$. The magnitude of $\vec{v}$ is denoted

$$|\vec{v}| \stackrel{\text{def}}{=} \sqrt{v_1^2 + v_2^2 + ...v_n^2}.$$

We build intuition through the lower dimensional cases. In $\mathbb{R}^2$, let us consider the point $(4, 3)$.

The magnitude of this vector is $\sqrt{3^2 + 4^2} = 5$. If this sounds very familiar, it is because this is indeed an application of Pythagorean theorem. In 3-dimensions, this still applies - take $\vec{x} = (a, b, c)$, we can traverse in each coordinate to apply Pythagorean theorem twice.





(a) $\vec{x} = (a, b, c)$. (b) Pyth. thm on a, b. (c) Reapply Pyth. thm.

Proposition 1.14

Let $\vec{v} \in \mathbb{R}^n, a \in \mathbb{R}$. Then

$$|a\vec{v}| = |a||\vec{v}|.$$

Definition 1.15 (Unit vector)

Let $\vec{v} \in \mathbb{R}^n$. We call $\vec{v}$ a **unit vector** if $|\vec{v}| = 1$.

Proposition 1.16

Let $\vec{v} \in \mathbb{R}^n$ be non-zero. Then there is a unique unit vector $|\vec{w}| = 1$ such that $\vec{v}$ is parallel to $\vec{w}$.

Proof. If $|\vec{w}| = 1, c > 0$ and $c\vec{w} = \vec{v}$, it must be that

$$c = |c||\vec{w}| = |c\vec{w}| = |\vec{v}|,$$

and

$$\vec{w} = \frac{1}{c}(c\vec{w}) = \frac{1}{|vecv|}\vec{v}.$$

This shows uniqueness - there is no other possible candidate except for this vector. Now we can confirm the vector $\frac{1}{|\vec{v}|}\vec{v}$ is a unit vector and is parallel to $\vec{v}$, so there exists a unique unit vector that satisfies the properties.

$\square$

We have now transported the notions of length and scaling into the coordinate system, and this allows us to make "measurements" such as angles and area.

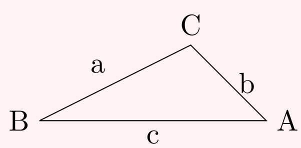
1.3.1 The Dot Product

Example 1.17

Let points A=$(5, 8)$, B=$(-2, 7)$, O=$(4, 5)$. Find the angle $\angle AOB$.

To solve this using only information about the lengths, recall the Law of Cosines:

Theorem 1.18 (Law of Cosines)

For any triangle 

,

$$c^2 = a^2 + b^2 - 2ab\cos(\angle BCA).$$

We now apply the Law of Cosines to $\triangle AOB$, so that

$$|\overrightarrow{AO}|^2 + |\overrightarrow{OB}|^2 - 2|\overrightarrow{AO}||\overrightarrow{OB}|\cos(\angle AOB) = |\overrightarrow{AB}|^2.$$

Plugging in the values, we solve

$$((4-5)^2 + (5-8)^2) + ((-2-4)^2 + (7-5)^2) - 2\sqrt{(4-5)^2 + (5-8)^2} \times \sqrt{(-2-4)^2 + (7-5)^2}\cos(\angle AOB) = ((-2-5)^2 + (7-8)^2).$$

Simplifying, we get

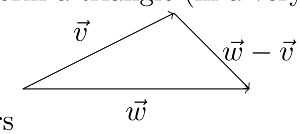
$$50 - 2\sqrt{10}\sqrt{40}\cos(\angle AOB) = 50 \implies \cos(\angle AOB) = 0.$$

So the angle is $\pi/2$.

Example 1.19

Find a closed form formula for the cosine of an angle formed by two vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$.

Let the angle formed be θ. Using the intuition from 2-D space, we can form a triangle (in a very complex n-dimensional space). We write the Law of Cosine in terms of vectors



$$|\vec{v}|^2 + |\vec{w}|^2 - 2|\vec{v}||\vec{w}| = |\vec{w} - \vec{v}|^2.$$

This expands to

$$\sum_{i=1}^{n} v_i^2 + \sum_{j=1}^{n} w_j^2 - 2\sqrt{\sum_{i=1}^{n} v_i^2}\sqrt{\sum_{j=1}^{n} w_j^2}\ \cos(\theta) = \sum_{i=1}^{n} (w_i - v_i)^2.$$

Rearranging $(a-b)^2 = a^2 + b^2 - 2ab$,

$$2\sqrt{\sum_{i=1}^{n} v_i^2}\sqrt{\sum_{j=1}^{n} w_j^2}\ \cos(\theta) = \sum_{i=1}^{n} 2w_i v_i$$

so

$$\cos(\theta) = \frac{\sum_{i=1}^{n} w_i v_i}{\sqrt{\sum_{i=1}^{n} v_i^2}\sqrt{\sum_{j=1}^{n} w_j^2}}.$$

Definition 1.20 (Dot Product)

Let $\vec{v}, \vec{w} \in \mathbb{R}^n$, the **dot product** between $\vec{v}$ and $\vec{w}$ is

$$\vec{v} \cdot \vec{w} \stackrel{\text{def}}{=} \sum_{i=1}^{n} v_i w_i.$$

Proposition 1.21

Let $\vec{v} \in \mathbb{R}^n$, then $|\vec{v}|^2 = \vec{v} \cdot \vec{v}$.

Proposition 1.22

Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$, $c \in \mathbb{R}$. The dot product satisfies the following properties:

- *(Symmetry)* $\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$.
- *(Linearity 1)* $(c\vec{v}) \cdot \vec{w} = c(\vec{v} \cdot \vec{w})(= \vec{v} \cdot (c\vec{w})$ by symmetry).
- *(Linearity 2)* $(\vec{v} + \vec{u}) \cdot \vec{w} = \vec{v} \cdot \vec{w} + \vec{u} \cdot \vec{w}$.
- *(Positive definiteness)* $\vec{v} \cdot \vec{v} \geq 0$, with equality if and only if $\vec{v} = \vec{0}$.

Proof. Here is the proof for the last property. The rest are left as an exercise We have $\vec{v} \cdot \vec{v} = |\vec{v}|^2 \geq= 0$, and if $\vec{v} \cdot \vec{v} = 0$, for all $1 \leq \jmath \leq n$, $v_j^2 \leq \sum_{i=1}^{n} v_i^2 = \vec{v} \cdot \vec{v} \leq 0$, so $v_j = 0$. This means $\vec{v} = \vec{0}$. $\square$

Corollary 1.23:

1. $|\vec{v}-\vec{w}|^2 = |\vec{v}|^2 + |\vec{w}|^2 - 2\vec{v}\cdot\vec{w}$.
2. In lower dimensions $\mathbb{R}^2, \mathbb{R}^3$ If $\vec{v}, \vec{w} \neq \vec{0}$, the angle between $\vec{v}$ and $\vec{w}$ is
$$\cos^{-1}\left(\frac{\vec{v}\cdot\vec{w}}{|\vec{v}||\vec{w}|}\right).$$
In particular, the angle is $\pi/2$ when $\vec{v}\cdot\vec{w} = 0$.
3. $|\vec{v}\cdot\vec{w}| \leq |\vec{v}||\vec{w}|$.

Theorem 1.24 (Cauchy-Schwarz inequality)

Let $\vec{v}, \vec{u} \in \mathbb{R}^n$. Then
$$|\vec{v}\cdot\vec{u}| \leq |\vec{u}||\vec{v}|.$$

The proof is in one of the exercises at the end of this chapter.
The dot product allows us to make sense of “angles” in higher dimensions, so we can generalize the notion of two vectors “perpendicular” to each other.

Definition 1.25 (Orthogonality)

Let $\vec{v}, \vec{w} \in \mathbb{R}^n$. We say $\vec{v}$ and $\vec{w}$ are **orthogonal** to each other if $\vec{v}\cdot\vec{w} = 0$.

Example 1.26

Find all vectors $\vec{v} \in \mathbb{R}^3$ such that $\vec{v}\cdot(-3\vec{i}+2\vec{j}+6\vec{k}) = 49$. Sketch the locus of corresponding points in $\mathbb{R}^3$.

The condition expands to

$$\begin{aligned} -3v_1 + 2v_2 + 6v_3 &= 49 \\ \implies v_3 &= \frac{49+3v_1-2v_2}{6} \end{aligned}$$

so that the set of all vectors is

$$\left\{ s\vec{i} + t\vec{j} + \left(\frac{49+3s-2t}{6}\right)\vec{k} | s,t \in \mathbb{R} \right\}.$$

To plot this in 3D space, notice that this describes the equation of a plane $z = (49+3x-2y)/6$. We pick any three points (of course we want those that are easy to calculate) $(0,0,49/6)$, $(0,49/12,0)$, $(-49/18,0,0)$.

```
def function_to_plot(X,Y):
return (49+3*X-2*Y)/6
#create figure
fig=plt.figure(figsize=(10,8))
ax = fig.add_subplot(projection='3d')
ax.view_init(elev=10, azim=120) #change view 1
```

```
#ax.view_init(elev=10, azim=0) #change view 2
X,Y=np.meshgrid(np.linspace(-10,10,10), np.linspace(-10,10,10))
Z=function_to_plot(X,Y)
surface=ax.plot_surface(X, Y, Z, alpha=0.5,
label='graph surface')
ax.set_xlabel('x')
ax.set_ylabel('y')
ax.set_zlabel('z')
#plot the curve and the cross section to integrate
ax.scatter(0,0,0,color='red',label='origin')
ax.text(0,0,0,'origin')
ax.scatter(-3,2,6,color='black')
ax.text(-3,2,6,'(-3,2,6)')
plt.legend()
plt.show()
```

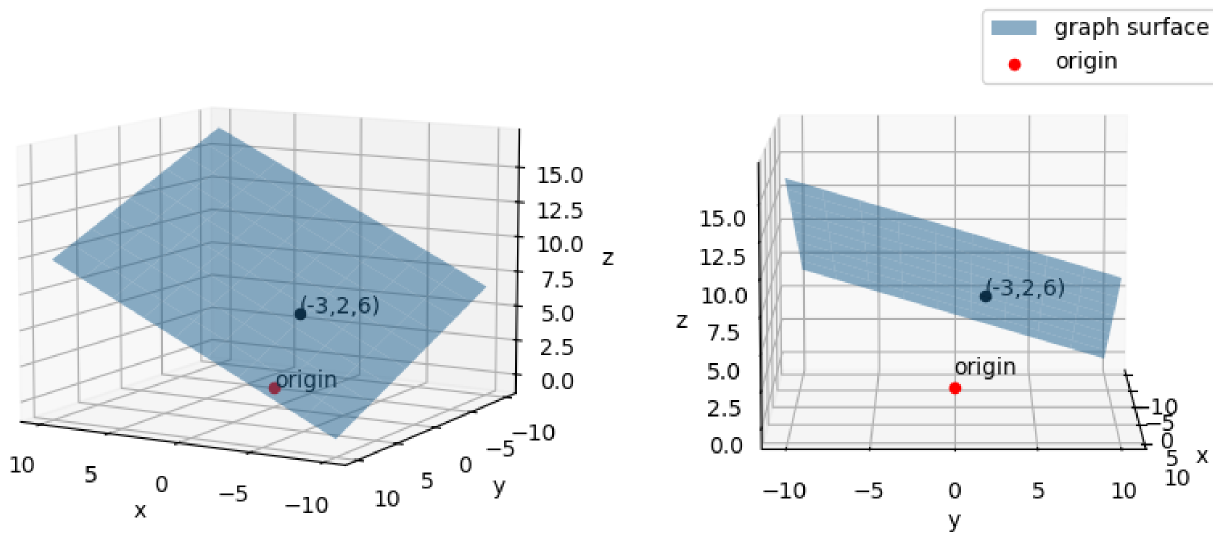


Figure 1.3: Two views of the plane

Interestingly, we see that $(-3, 2, 6)$ - the point corresponding to our vector - is a point on the plane!

Remark. *The form*

$$\left\{ s\vec{i} + t\vec{j} + \left(\frac{49 + 3s - 2t}{6} \right) \vec{k} | s, t \in \mathbb{R} \right\}$$

*is a **parametrization** of the plane described by* $-3x + 2y + 6z = 49$*, viewing this as a function* $\mathbb{R}^2 \to \mathbb{R}^3$.

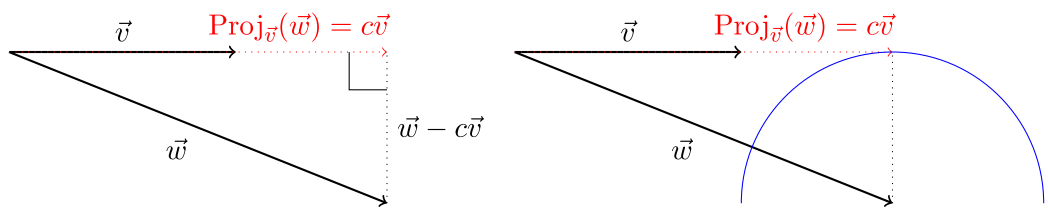
1.3.2 Projection

Definition 1.27 (Projection)

Let $\vec{v}, \vec{w} \in \mathbb{R}^n$, $\vec{v} \neq \vec{0}$, and $c \in \mathbb{R}$, such that $\vec{v} \cdot (\vec{w} - c\vec{v}) = 0$ (or equivalently, $\vec{w} - c\vec{v}$ orthogonal to $\vec{v}$).
We say that $c\vec{v} \overset{\text{def}}{=} \text{Proj}_{\vec{v}}(\vec{w})$ is the **vector projection of $\vec{w}$ on $\vec{v}$**.

To build intuition, it is always helpful to start with lower dimensions. We take a plane through $\vec{v}$ and $\vec{w}$. Looking at this slice, we can work in 2D.

We see that $c\vec{v}$ is the the point on the extension of $\vec{v}$ such that it is closest to the point $\vec{w}$. To geometrically show this idea, we draw a circle centered at $\vec{w}$ with radius $|\vec{w} - c\vec{v}|$. The line generated by $\vec{v}$ is tangent to this circle, so the contact point at $c\vec{v}$ is indeed the closest point to $\vec{w}$.



We also see from this geometric construction that 1) the projection $c\vec{v}$ is unique thus well defined, 2) $c|\vec{v}| = |\vec{w}| \cos\theta$

Proposition 1.28

The projection is unique and is given by

$$\text{Proj}_{\vec{v}}(\vec{w}) = \frac{\cos\theta|\vec{w}|}{|\vec{v}|} = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}|^2} = \left(\frac{\vec{v} \cdot \vec{w}}{\vec{v} \cdot \vec{v}}\right)\vec{v}.$$

This is the first time we show that something is unique. The standard argument goes as follows: Suppose another object a' satisfies all the properties you want for a. Then you can show $a = a'$, meaning every object that satisfies the properties is a, or equivalently, a is unique.

So what happens if we pick two points P, P' on $\vec{v}$ such that both make a right angle when connected to $\vec{w}$? We would have constructed a triangle $\triangle PP'\vec{w}$ with two right angles!

Proof. Let $c, c' \in \mathbb{R}$ such that $\vec{p}_1 = c\vec{v}$ and $\vec{p}_2 = c'\vec{v}$ are both satisfy the definition of $\text{Proj}_{\vec{v}}(\vec{w})$. We want to show that $c = c'$, so we consider $\vec{u} = (c - c')\vec{v}$, the line between the two projections.
By corollary 1.23, we have two equations

$$|\vec{w} - \vec{p}_1|^2 = |\vec{w} - \vec{p}_2|^2 + |\vec{u}|^2$$
$$|\vec{w} - \vec{p}_2|^2 = |\vec{w} - \vec{p}_1|^2 + |\vec{u}|^2$$

We can solve for $|\vec{u}| = 0$, and by positive definiteness, $\vec{u} = \vec{0}$ and $c - c' = |\vec{u}|/|\vec{v}| = 0$.
To show our formula for projection works, we can just compute that

$$\vec{v} \cdot \left(\vec{w} - \left(\frac{\vec{v} \cdot \vec{w}}{\vec{v} \cdot \vec{v}}\right)\vec{v}\right) = \vec{v} \cdot \vec{w} - \left(\frac{\vec{v} \cdot \vec{w}}{\vec{v} \cdot \vec{v}}\right)(\vec{v} \cdot \vec{v}) = 0.$$

□

Exercises

3. Determine if the following pairs of vectors are orthogonal:

   (a) $3\vec{e}_1 + 4\vec{e}_2$, $-4\vec{e}_1 + 3\vec{e}_2$

   (b) $(4, -1, 2)$, $(3, 0, -6)$

   (c) $3\vec{i} - 2\vec{j}$, $-2\vec{i} - 4\vec{k}$

   (d) $\vec{e}_1 + 3\vec{e}_3 + 5\vec{e}_5 + ... + (2k-1)\vec{e}_{2k-1}$, $2\vec{e}_2 + 4\vec{e}_4 + ... + (2k)\vec{e}_{2k}$

4. Refer to the plane in Example 1.26. Show that $\text{Proj}_{-3\vec{i}+2\vec{j}+6\vec{k}}(\vec{v})$ is the same for any vector $\vec{v}$ such that $\vec{v} \cdot (-3\vec{i} + 2\vec{j} + 6\vec{k}) = 49$, and compute this projection.

5. *(Geometry of a methane molecule)* Place four points $P(0,0,0)$, $Q(1,1,0)$, $R(1,0,1)$, $S(0,1,1)$ in $\mathbb{R}^3$.

   (a) Compute the distance between any two points of $PQRS$ and show that all 6 pairs are the same. This means that $PQRS$ forms a regular tetrahedron.

   (b) Verify that the geometric center of the tetrahedron $O(1/2, 1/2, 1/2)$ is equidistant to all of the vertices of the tetrahedron.

   (c) Compute the angle between two edges of the tetrahedron, rounded to 2 decimal places. By symmetry, you only have to compute one angle.

   (d) Compute the angle $\angle POQ$, rounded to 2 decimal places. Does this angle remind you of something from Chemistry?

1.4 Cross Product

Definition 1.29 (Cross Product)

Let $\vec{v}, \vec{w} \in \mathbb{R}^3$, we define the **cross product** of $\vec{v}$ and $\vec{w}$ to be

$$\vec{v} \times \vec{w} \stackrel{\text{def}}{=} \begin{bmatrix} v_2 w_3 - v_3 w_2 \\ v_3 w_1 - v_1 w_3 \\ v_1 w_2 - v_2 w_1 \end{bmatrix}.$$

Remark. *Later on, we will introduce determinants of a matrix. The cross product can be understood as the determinant of the 'matrix'*

$$\begin{bmatrix} \vec{i} & v_1 & w_1 \\ \vec{j} & v_2 & w_2 \\ \vec{k} & v_3 & w_3 \end{bmatrix}.$$

This definition seems a bit unmotivating, so let us work through some examples.

Example 1.30

Compute the 9 cross products for each pair of the standard basis vectors in $\mathbb{R}^3$.

With some (heavy) computation, we find

$$\begin{aligned}
&\vec{i}\times\vec{i}=\vec{0} && \vec{i}\times\vec{j}=\vec{k} && \vec{i}\times\vec{k}=-\vec{j}\\
&\vec{j}\times\vec{i}=-\vec{k} && \vec{j}\times\vec{j}=\vec{0} && \vec{j}\times\vec{k}=\vec{i}\\
&\vec{k}\times\vec{i}=\vec{j} && \vec{k}\times\vec{j}=-\vec{i} && \vec{k}\times\vec{k}=\vec{0}
\end{aligned}$$

Importantly, the cross product is **antisymmetric**, meaning $\vec{v}\times\vec{w}=-\vec{w}\times\vec{v}$. You can see special cases with the standard basis from above and confirm the general case in an exercise.

Example 1.31

Compute $\vec{v}\cdot(\vec{v}\times\vec{w})$ and $\vec{w}\cdot(\vec{v}\times\vec{w})$.

Again with some heavy computation,

$$\begin{aligned}
\vec{v}\cdot(\vec{v}\times\vec{w}) &= \begin{bmatrix} v_1\\ v_2\\ v_3\end{bmatrix}\cdot\begin{bmatrix} v_2w_3-v_3w_2\\ v_3w_1-v_1w_3\\ v_1w_2-v_2w_1\end{bmatrix}\\
&= {\color{red}v_1v_2w_3} - v_1v_3w_2 + {\color{blue}v_2v_3w_1} - {\color{red}v_2v_1w_3} + v_3v_1w_2 - {\color{blue}v_3v_2w_1}\\
&= 0,\\
\vec{w}\cdot(\vec{v}\times\vec{w}) &= \begin{bmatrix} w_1\\ w_2\\ w_3\end{bmatrix}\cdot\begin{bmatrix} v_2w_3-v_3w_2\\ v_3w_1-v_1w_3\\ v_1w_2-v_2w_1\end{bmatrix}\\
&= {\color{red}w_1v_2w_3} - w_1v_3w_2 + w_2v_3w_1 - {\color{blue}w_2v_1w_3} + {\color{blue}w_3v_1w_2} - {\color{red}w_3v_2w_1}\\
&= 0.
\end{aligned}$$

Which means $\vec{v}\times\vec{w}$ is orthogonal to $\vec{v}$ and $\vec{w}$!

Example 1.32

Compute $|\vec{v}\times\vec{w}|^2+(\vec{v}\cdot\vec{w})^2$.

We have

$$\begin{aligned}|\vec{v}\times\vec{w}|^2+(\vec{v}\cdot\vec{w})^2 =&(v_2w_3-v_3w_2)^2+(v_3w_1-v_1w_3)^2+(v_1w_2-v_2w_1)^2\\ &+(v_1w_1+v_2w_2+v_3w_3)^2\\ =&(v_1w_1)^2+(v_1w_2)^2+(v_1w_3)^2\\ &+(v_2w_1)^2+(v_2w_2)^2+(v_2w_3)^2\\ &+(v_3w_1)^2+(v_3w_2)^2+(v_3w_3)^2\\ =&(v_1^2+v_2^2+v_3^2)(w_1^2+w_2^2+w_3^3)\\ =&|\vec{v}|^2|\vec{w}|^2.\end{aligned}$$

Now we substitute $\vec{v}\cdot\vec{w}=|\vec{v}||\vec{w}|\cos\theta$, we get

$$|\vec{v}\times\vec{w}|=|\vec{v}||\vec{w}|\sqrt{1-\cos^2\theta}=|\vec{v}||\vec{w}|\sin\theta.$$

Where we know $\sin\theta\geq 0$ as θ is between 0 and π.

The value $|\vec{v}||\vec{w}|\sin\theta$ has a nice geometric meaning. It is the area spanned by the vectors $\vec{v}$ and $\vec{w}$.

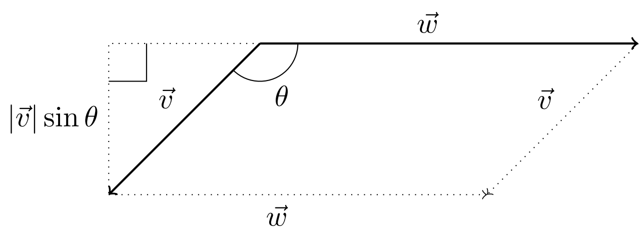


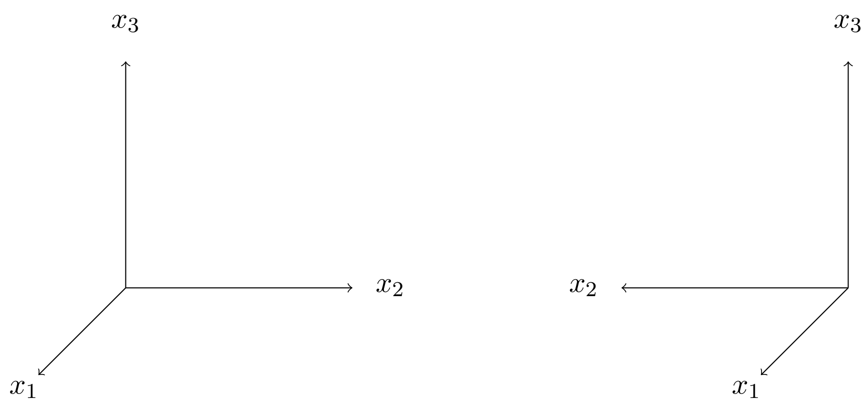
Figure 1.5: The area of a parallelogram formed by these vectors is the magnitude of the vector $\vec{v}\times|\vec{w}|$

Proposition 1.33

The direction of $\vec{v}\times\vec{w}$ is determined by the **right-hand rule** as follows:

> Using the right hand, align the index finger with the direction $\vec{v}$, and the middle finger with the direction of $\vec{w}$. Extend the thumb so that it is perpendicular to both the index finger and the middle finger. The thumb is pointing in the direction of $\vec{v}\times\vec{w}$.

This is a byproduct of the convention we use. In $\mathbb{R}^3$ we use what is known as a right-handed coordinate system - the vectors $\vec{i}$, $\vec{j}$, and $\vec{k}$ align with the first three fingers of the right hand respectively. If we used a left-handed coordinate system, the rule would be left-handed instead. We now have a few properties about the cross product from our computation, the first you will verify on your own:



(a) Right handed coordinate system (b) Left handed coordinate system

Proposition 1.34

Let $\vec{v}, \vec{w}, \vec{u} \in \mathbb{R}^n$, then

- *(distributivity)* $\vec{v} \times (\vec{u} + \vec{w}) = \vec{v} \times \vec{u} + \vec{v} \times \vec{w}$ and $(\vec{v} + \vec{u}) \times \vec{w} = \vec{v} \times \vec{w} + \vec{v} \times \vec{u}$.
- *(anti-symmetry)* $\vec{v} \times \vec{w} = -\vec{w} \times \vec{v}$.
- $\vec{v} \times \vec{w}$ is orthogonal (perpendicular in 3D space) to $\vec{v}$ and $\vec{w}$.
- The magnitude $|\vec{v} \times \vec{w}|$ is given by $|\vec{v}||\vec{w}| \sin\theta$, with θ being the angle between $\vec{v}$ and $\vec{w}$, so
  (i) The magnitude $|\vec{v} \times \vec{w}|$ also corresponds to the area of the parallelogram formed by $\vec{v}$ and $\vec{w}$.
  (ii) If $\vec{v}$ and $\vec{w}$ are parallel or antiparallel, $\vec{v} \times \vec{w} = \vec{0}$.

Remark. *Using distributivity of the cross product, you only need to memorize the cross product of the basis vectors, and write* $\vec{v} \times \vec{w} = \sum_{i=1}^{3} \sum_{j=1}^{3} v_i w_i (\vec{e}_i \times \vec{e}_j)$.

1.4.1 Triple products

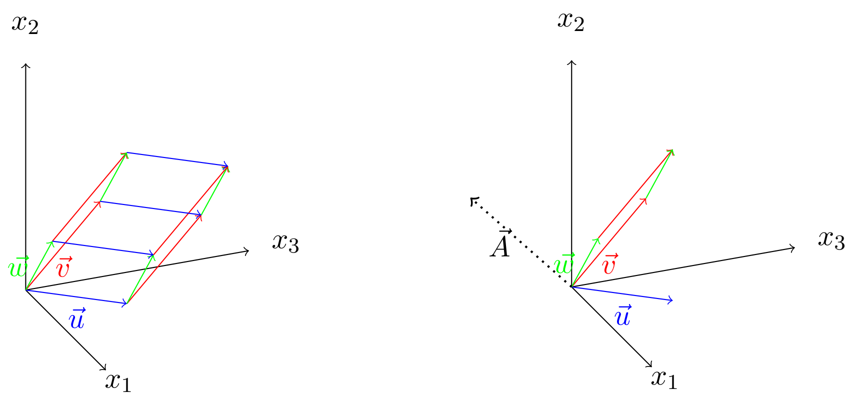
Example 1.35

Find the volume of the parallelepiped formed from $\vec{v} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \vec{w} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \vec{u} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

The parallelepiped is a generalization of a parallelogram to higher dimensions. Using combinations of $\vec{v}, \vec{w}, \vec{u}$, you can make the frame of a 3d solid. As in the figure on the left, edges of the same color correspond to the same vector (and thus parallel). The volume of this solid is still *base* $\times$ *height*, where the base is a 2D parallelogram formed by two vectors and the height is determined by third vector. We make an arbitrary decision and set the base to be $\vec{v}$ and $\vec{w}$. (setting any two vectors would give the same result in the end!) The area of this base is given by the cross product

$$\vec{A} = \vec{v} \times \vec{w} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}.$$

Now we can determine the height of the paralellepiped from $\vec{u}$. We want to isolate the component of $\vec{u}$ that is orthogonal to the base. Equivalently, we want to find the component of $\vec{u}$ that is pointing in the direction of $\vec{A}$, a vector that is orthogonal to both $\vec{v}$ and $\vec{w}$!



(a) The parallelepiped, draw in an optical illusion fashion.

(b) We want to get the projection of $\vec{u}$ on $\vec{A}$.

The volume is thus

$$|\mathrm{Proj}_{\vec{A}}(\vec{u})||\vec{v}| = \left|\frac{\vec{u} \cdot \vec{A}}{|\vec{A}|^2}\right| \times |\vec{A}| \times |\vec{A}| = |\vec{u} \cdot \vec{A}| = 2.$$

Proposition 1.36

The volume of the parallelpiped formed from $\vec{v}, \vec{w}, \vec{u}$ is

$$|(\vec{v} \times \vec{w}) \cdot \vec{u}|$$

Remark. *This is also the expression of the (absolute value of) determinant of*

$$\begin{bmatrix} \vec{v} & \vec{w} & \vec{u} \end{bmatrix}$$

where $\vec{v}, \vec{w}, \vec{u}$ are written as column vectors. Using properties of the determinant (later chapters), you can show cycling the three vectors does not change the volume. (i.e. you can calculate using what order of the three vectors you want)

Remark. *You may notice that the expression $(\vec{v} \times \vec{w}) \cdot \vec{u}$ can take on negative volumes. In this case, the three vectors (taken in order) do not follow the right-hand rule. For instance, in the last example, $\vec{u}$ points in the 'opposite' direction as $\vec{A}$.*

Exercises

6. Find $\vec{v} \times \vec{w}$ for the following:
   (a) $\vec{v} = (4, -2, 0)$, $\vec{w} = (2, 1, -1)$
   (b) $\vec{v} = (3, 3, 3)$, $\vec{w} = (4, -3, 2)$
   (c) $\vec{v} = 2\vec{i} + 3\vec{j} + 4\vec{k}$, $\vec{w} = \vec{i} - 3\vec{j} + 4\vec{k}$
7. Find the areas for the following shapes:
   (a) The parallelogram with vertices $P(0, 0, 0), Q(1, 1, 0), R(1, 2, 1), S(0, 1, 1)$.
   (b) The triangle with vertices $A(1, 9, 3), B(-2, 3, 0), C(3, -5, 3)$.
8. Find the volume of the parallelepiped formed from the vectors $\vec{v} = (1, 1, 0), \vec{w}(0, 2, -2), \vec{u} = (1, 0, 3)$.
9. A triangular kite has vertices $P(0, 0, 10), Q(2, 1, 10), S(0, 3, 12)$ and is displaced by the wind at a velocity of $(20\vec{i} + 6\vec{j} + 4\vec{k})/s$
   (a) Find the area of the kite.
   (b) After 1/2 seconds, find the volume of the space swept by the kite. (leave the answers in $[units]^3$)

1.5 Applications - Geometry of lines and planes

Definition 1.37 (Relations between lines)

For two (infinitely extending) lines in $\mathbb{R}^n$ parametrized in s and t respectively as $l_1 = P + t\vec{v}$, $l_2 = Q + s\vec{w}$, we say the lines are

- **Parallel**, if the $\vec{v}$ and $\vec{w}$ are parallel or antiparallel.
- **Intersecting**, if l_1 and l_2 exactly one point on both l_1 and l_2.
- **Skew**, if l_1 and l_2 are not parallel/antiparallel or intersecting.

Remark. *Lines do not have direction, so there usually is no need to distinguish between parallel and antiparallel lines. One may extend the definition of antiparallel to lines from $\vec{v}$ and $\vec{w}$.*

Proposition 1.38

Determination of parallel lines are independent of parametrization. Concretely,

> Let $P_1 + t_1\vec{v}_1$ and $P_2 + t_2\vec{v}_2$ be two parametrizations of l_1, $Q_1 + s_1\vec{w}_1$, $Q_2 + s_2\vec{w}_2$ be two parametrizations of l_2. If $\vec{v}_1 = c_1\vec{w}_1$ for some $c_1 \in \mathbb{R}$, then $\vec{v}_2 = c_2\vec{w}_2$ for some (possibly different)$c_2 \in \mathbb{R}$.

The proof is not very enlightening. However, the result of this guarantees that our definition of parallel lines is precise.

Proof. The idea is to show that $\vec{v}_1$ is a scalar multiple of $\vec{v}_2$, and by the same logic $\vec{w}_1$ is a scalar multiple of $\vec{w}_2$. Since all vectors are non-zero in the parametrization, we will get the result of $\vec{v}_2$ a scalar multiple of $\vec{w}_2$.

To show $\vec{v}_1 = k\vec{v}_2$ for some k, we can pick two distinct points A, B on l_1. From the parametriation we can get $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$ such that $P_1 + \alpha_1\vec{v}_1 = A = P_2 + \alpha_2\vec{v}_2$ and $P_1 + \beta_1\vec{v}_1 = B = P_2 + \beta_2\vec{v}_2$. Therefore we get the vector

$$\overrightarrow{AB} = (\beta_1 - \alpha_1)\vec{v}_1,$$
$$\overrightarrow{AB} = (\beta_2 - \alpha_2)\vec{v}_2.$$

As we picked distinct points A and B, we can conclude $\overrightarrow{AB} \neq \vec{0}$ and thus $\beta_1 - \alpha_1 \neq 0, \beta_2 - \alpha_2 \neq 0$, so that

$$\vec{v}_1 = \frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1}\vec{v}_2.$$

□

Example 1.39

Determine whether the lines parametrized by $l_1(t) = (1, 2, 1) + t(1, 3, -2)$ and $l_2(t) = (3, 1, 0) + t(-2, -6, 4)$ are parallel, intersecting, or skew. Confirm that l_1 and l_2 describe two different lines.

We notice that $-2 \times (1, 3, -2) = (-2, -6, 4)$, so these lines are parallel. To confirm that these two lines are not the same, we notice that $(1, 2, 1)$ is a point on l_1, but if we attempt to solve

$$(1, 2, 1) = l_2(t) = (3, 1, 0) + t(-2, -6, -4) \implies (-2, 1, 1) = t(-2, -6, -4)$$

which no t can solve! Specifically, the first coordinate forces $t = 1$ and the second coordinate forces $t = -6$.

Example 1.40

Determine whether the lines parametrized by $l_1(t) = (1, 2, 1) + t(1, 3, -2)$ and $l_2(t) = (0, 3, 9) + t(0, 2, 3)$ are parallel, intersecting, or skew.

$(1, 3, 2)$ is not a multiple of $(0, 2, 3)$, so the lines are not parallel. We might be tempted to solve for $l_1(t) = l_2(t)$ to check for intersection, but this misses a lot of cases! We need to compare all the points of l_1 with all the points of l_2, so we need two independent variables to describe where we are on each of the lines. That is, we solve for s, t in $l_1(t) = l_2(s)$,

$$\begin{aligned} &(1, 2, 1) + t(1, 3, -2) = (0, 3, 9) + s(0, 2, 3) \\ \implies &(t + 1, 3t + 2, -2t + 1) = (0, 2s + 3, 3s + 9) \\ &\implies t = -1 \text{ and } 3t + 2 = 2s + 3 \text{ and } -2t + 1 = 3s + 9 \end{aligned}$$

$t = -1, s = -2$ solves this system of equations. We can plug in t and s in our original parametrization to find $(0, 1, 3)$ is indeed a point on both l_1 and l_2. We would have missed this if we set $l_1(t) = l_2(t)$!

Example 1.41

Determine whether the lines parametrized by $l_1(t) = (1,2,1) + t(1,3,-2)$ and $l_2(t) = (0,3,8) + t(0,2,3)$ are parallel, intersecting, or skew.

We repeat the same process as above to see that the lines are not parallel and solve for

$$\begin{aligned} &(1,2,1) + t(1,3,-2) = (0,3,8) + s(0,2,3) \\ \implies &(t+1, 3t+2, -2t+1) = (0, 2s+3, 3s+8) \\ &\implies t = -1 \text{ and } 3t+2 = 2s+3 \text{ and } -2t+1 = 3s+8 \end{aligned}$$

This time, we do not have a solution - the first two equations forces $t = -1, s = -2$, and this does not solve the third. We therefore do not have a point of intersection, and the lines are skew.

Example 1.42

Determine if $P(5,6,9), Q(7,9,15), R(13,18,33)$ are **colinear** i.e. if they lie on the same line.

With the machinery we have built up, there are multiple ways to check if P, Q, R form a straight line. Here are a few ideas:

1. Check that $\overrightarrow{PQ}$ is parallel/antiparallel to $\overrightarrow{PR}$. Because the lines defined by $\overline{PQ}$ and $\overline{PR}$ are parallel and share a same point P, they are the same line.

2. Use the dot product to calculate the angle $\angle PQR = \pi$.

3. Use the cross product to calculate that $\overrightarrow{PQ} \times \overrightarrow{PR} = \vec{0}$. This means the triangle with vertices P, Q, R has no area and thus is a degenerate triangle.

Definition 1.43 (Characterization of Planes)

In euclidean geometry, planes can be characterized by any of the following ways:

- For **any three non-colinear points** P_1, P_2, P_3, there is a unique plane passing through P_1, P_2, P_3.
- For **any pair of intersecting lines**, there is a unique plane that contains both.
- For **a line** l **and a point** P, there is a unique plane that contains P and is perpendicular to l.
- For **a line** l **and a point** P **not on** l, there is a unique plane that contains both l and P.

Remark. *The third characterization is the hardest to visualize at first, but is also the easiest to describe with the analytical tools we have built towards. We can refer to Example 1.26. The plane sketched is the unique plane that contains* $(-3,2,6)$*, such that each vector in the plane is orthogonal to* $(-3,2,6)$*, so the line parametrized by* $l(t) = t(-3,2,6)$ *is perpendicular to the plane at the point of intersection* $(-3,2,6)$*.*

Example 1.44

In $\mathbb{R}^3$, determine the equation of the plane that contains $P_0(x_0, y_0, z_0)$ and is perpendicular to the line parametrized by $l(t) = Q_0 + t\vec{N}$, $\vec{N} = (a, b, c)$.

First we can exploit **translation invariance** of $\mathbb{R}^3$ and move the line to $\tilde{l} = P_0 + t\vec{N}$. Because l and $\tilde{l}$ are parallel, any line perpendicular to l will also be perpendicular to $\tilde{l}$.
Then by the characterization of a plane, any P on the plane satisfies the orthogonality relation $\overrightarrow{PP_0} \cdot \vec{N} = 0$. Expanding this, we get the equation

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0 \implies ax + by + cz = ax_0 + by_0 + cz_0.$$

Definition 1.45 (Normal form of a plane)

Let $a, b, c, d \in \mathbb{R}$, with at least one of $a, b, c \neq 0$. The equation of a plane in ^{3} written as

$$ax + by + cz = d$$

is called a **normal form** of the equation of a plane.

Remark. *As all planes can be characterized this method, all equations of planes can be put in normal form.*

Remark. *The normal form of a plane is not unique. Pick your favorite non-zero number α, the equation $\alpha ax + \alpha by + \alpha c_z = \alpha d$ describes the same plane.*

Theorem 1.46 (Normal vectors of planes)

When written in normal form, the plane is perpendicular to the vector $\vec{N} = (a, b, c)$. We call this vector $\vec{N}$ the **normal vector**.

We thus have a definition for two planes to be parallel.

Definition 1.47 (Parallel planes)

Two planes are **parallel** if the normal vectors are parallel.

A problem in the exercise will guide you through the proof of Theorem 1.42. The intuition behind the proof is the reverse direction of the equation $\overrightarrow{PP_0} \cdot \vec{N} = 0$ we derived from the last example.
We will now apply this theorem in a few examples.

Example 1.48

Find the equation of the plane that passes through the points $P(1, 0, 0), Q(0, 1, 0), R(0, 0, 1)$.

Method 1: One sees that (by coincidence) the sum of coordinates of each point are equal to 1, so immediately writes down $x + y + z = 1$. Despite being the fastest method, this is somewhat inconsistent.

Method 2: We set $ax + by + cz = d$ to be the equation, and plug in values for P, Q, R, giving the system of equations

$$\begin{aligned} a + 0b + 0c &= d \\ 0a + b + 0c &= d \\ 0a + 0b + c &= d \end{aligned}$$

This is an **underdetermined system**, meaning there are fewer equations than unknowns. The best we can say is that $a = b = c = d$. However, setting $a = b = c = d = k$ works for all $k \neq 0$, further confirming that the normal form is not unique. This method is reasonably fast when the system of equations are simple. When there are more non-zero coefficients, solving the system takes more time.

Method 3: We can compute the direction of the normal vector $\vec{N}$ of this plane. By the characterization of planes, $\vec{N}$ is orthogonal to $\overrightarrow{PQ}$ and $\overrightarrow{PR}$, two vectors in the plane. We can thus write $\vec{N}$ as a multiple of the cross product

$$\overrightarrow{PQ} \times \overrightarrow{PR} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \times \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

. Using this normal vector (or any multiple of it) and applying the theorem, we get $x + y + z = d$, and substituting P into the equation will give $x + y + z = 1$. This method is more general and consistent, as the number of operations in a cross product is constant.

Example 1.49

On the plane given by $ax + by + cz = d$, find the point on the plane that is closest to the origin.

Let r be the distance, we draw a sphere centered at the origin with radius r. This sphere is thus tangent to the surface at one point, and the vector corresponding to this point will be perpendicular to the plane. By the theorem, we can denote the point $P(\alpha a, \alpha b, \alpha c)$, with α a constant to be determined. Substituting P into the equation,

$$\alpha(a^2 + b^2 + c^2) = d \implies \alpha = \frac{d}{a^2 + b^2 + c^2}$$

The closest distance from the origin is

$$|(\alpha a, \alpha b, \alpha c)| = |\alpha|\sqrt{a^2 + b^2 + c^2} = \left|\frac{d}{\sqrt{a^2 + b^2 + c^2}}\right|$$

at the point

$$\left(\frac{da}{\sqrt{(a^2 + b^2 + c^2)}}, \frac{db}{\sqrt{(a^2 + b^2 + c^2)}}, \frac{dc}{\sqrt{(a^2 + b^2 + c^2)}}\right)$$

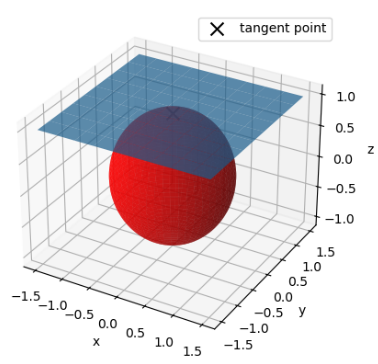


Figure 1.8: A toy example with the unit sphere of radius 1 and the plane described by $z = 1$.

1.5.1 Intersection of planes

Example 1.50

Find the intersection of the planes given by

$$6x + 2y - z = 2$$
$$\text{and } x - 2y + 3z = 5.$$

Method 1: The line lies on the first plane, so is perpendicular to its normal vector $(6, -2, -1)$. The line also lies on the second plane, so is perpendicular to its normal vector $(1, -2, 3)$. Using the cross product, the line should point in the direction of

$$\begin{bmatrix} 6 \\ 2 \\ -1 \end{bmatrix} \times \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ -19 \\ -14 \end{bmatrix}$$

Now we just need to find one point P on the line of intersection, so that the parametric equation (in t) is $P + t(4, -19, -14)$. We can impose an additional restriction that $x = 0$ (or $y = 0$ or $z = 0$) and solve the simultaneous equations

$$2y - z = 2$$
$$-2y + 3z = 5.$$

and get $(0, 11/4, 7/2)$ is a point on the intersection. The final parametric representation is $l(t) = (0, 11/4, 7/2) + t(4, -19, -14)$.

Method 2: We can solve the system of equations to find the line of intersection, and arrive at the set of symmetric equations. A similar method would be to solve for two points on the line of intersection, then get the parametric equation. This method will be introduced later

Exercises

10. Find parametric and symmetric equations for the line that
    (a) passes through $(0, 0, 0)$ and is parallel to $\vec{v} = 3\vec{i} + 4\vec{j} + 5\vec{k}$
    (b) passes through $(1, 1)$ and is perpendicular to the line described by $y = 3x + 3$
    (c) passes through $(0, 1, 2)$ and is perpendicular to plane with equation $3x + 4y = 2 + z$
    (d) passes through $(0, 1, 2)$ and is perpendicular to the yz-plane
    (e) passes through $(1, 3, 0)$ and is parallel to the line with symmetric equations $x = y - 1 = (z - 3)/2$

11. Determine if these pair of lines are parallel, intersecting, or skew:
    (a) (As described by parametric equations) $l_1(t) = (1, -1, 2) + t(2, 1, 1), l_2(t) = t(1, 0, -1)$
    (b) The line passing through $(3, 1, 3)$ and $(-2, -4, -5)$ and the line passing through $(1, 3, 5)$ and $(11, 13, 21)$.

12. Determine if $P(3,1,2), Q(-1,0,2), R(11,3,2)$ are colinear.

13. Find an equation for the plane

    (a) Containing $P(1,2,0)$ with normal vector $\vec{N} = 2\vec{i} - \vec{j} + 3\vec{k}$
    (b) containing $P(2,4,5)$ with normal vector $\vec{N} = \vec{i} + 3\vec{k}$
    (c) containing $P(1,4,3)$ and perpendicular to the line given by the parametrization $r(t) = (1+t, 2+4t, t)$
    (d) containing the origin and parallel to the plane described by $3x + 4y - 6z = -1$
    (e) containing the points $P(0,0,0), Q(1,-2,8), R(-2,-1,3)$

14. Find the intersection of the planes with equations $2x + 3y - z = 1$ and $x - y - z = 0$

15. Find the angle between the normal vectors of the following planes:

    (a) the planes with equations $x + 2y - z = 2$ and $2x - y + 3z = 1$
    (b) the plane with equation $2x+3y-6z = 0$ and the plane containing the points $P(1,3,-2), Q(5,1,3), R(1,0,1$

1.6 End of Chapter Exercises

16. Let $PQRS$ be a parallelogram. Show that $\overline{PR}$ and $\overline{QS}$ bisect each other i.e. intersect at their midpoints.

17. Complete the proof to Proposition 1.19 - Show that the dot product is symmetric and linear.

18. Prove the Cauchy-Schwarz inequality. Here is a hint: Write $\vec{w} = \text{Proj}_{\vec{v}}\vec{w} + (\vec{w} - \text{Proj}_{\vec{v}}\vec{w})$, and take the squared-magnitude on each side using the dot product.

19. Prove or give a counterexample: The cross product is associative.

20. Prove the following identity: $(\vec{a} \times \vec{b}) \times \vec{c} = -(\vec{b} \cdot \vec{c})\vec{a} + (\vec{a} \cdot \vec{c})\vec{b}$.

21. Show that the cross product satisfies the Jacobi identity: For any $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^3$, $(\vec{a} \times \vec{b}) \times \vec{c} + (\vec{b} \times \vec{c}) \times \vec{a} + (\vec{c} \times \vec{a}) \times \vec{b} = \vec{0}$.

22. On the plane given by $ax + by + cz = d$, show that the closest distance from $P(x_1, y_1, z_1)$ to the plane is given by $|ax_1 + by_1 + cz_1 - d|/\sqrt{a^2 + b^2 + c^2}$. (Hint: Refer to Example 1.44)

23. Define two lines in $\mathbb{R}^3$ by the parametric equations

$$l_1(t) = \vec{r}_1 + t\vec{v}_1$$
$$l_2(t) = \vec{r}_2 + t\vec{v}_2$$

    We will derive a formula for the shortest distance between the two lines.

    (a) Consider the special case where the two lines are parallel. Construct two points A and B on l_1 and l_2 respectively. Find B' Such that $\overrightarrow{AB'} = \text{Proj}_{\vec{v_1}}(\overrightarrow{AB})$. Argue that B' is on the line l_1, and that the line $\overline{BB'}$ is perpendicular to l_1 and l_2. Conclude that the shortest distance between l_1 and l_2 is given by

$$\sqrt{|\vec{r}_2 - \vec{r}_1|^2 - \left(\frac{(\vec{r}_2 - \vec{r_1}) \cdot \vec{v}_1}{|\vec{v}_1|}\right)^2}$$

(b) Now we can assume the lines are not parallel. Find a vector $\vec{N}$ that is perpendicular to both lines.

(c) (In your mind,) look in the direction of $\vec{N}$. $\vec{N}$ will now look like a dot. The lines will seem to overlap at points A' and B' respectively. Argue that $\overline{A'B'}$ is the shortest line from l_1 to l_2, and it is given by

$$\overrightarrow{A'B'} = \mathrm{Proj}_{\vec{N}}(\overrightarrow{AB})$$

, for any A on l_1 and B on l_2.

(d) Conclude that the shortest distance in this case is given by

$$\left|\frac{(\vec{r_2} - \vec{r_1}) \cdot (\vec{v}_1 \times \vec{v}_2)}{|\vec{v}_1 \times \vec{v}_2|}\right|.$$

(e) Suppose l_1 and l_2 intersect at P. Verify that the formula evaluates to 0, thus it gives a criterion to check if two lines intersect.

can this be compiled?

Chapter 2

Linear Algebra Basics

The foundational abstraction of linear algebra is the vector space. A *vector space* is essentially a collection of objects that it makes sense to take linear combinations of. Two operations must be defined: addition and scalar multiplication. A well defined vector space meets all of the vector space axioms, which will be listed shortly. Many consequences can be drawn from these axioms, and we can build up linear algebra to solve any linear problem.

Remark. *One has to specify the field of scalars (field just means number system) related to the vector space. In most cases here, we will be talking about real vector spaces (the field of scalars is* $\mathbb{R}$*). However, vector spaces can be defined with many other fields of scalars. Examples will follow the definition.*

Definition 2.1 (Informal definition of a field)

A field $\mathbb{K}$ is a set that supports addition, subtraction, multiplication, division with the similar properties that you expect from the real numbers.

Where is this going?

In the previous chapter, we have seen vector manipulation in $\mathbb{R}^n$. This chapter tries to generalize some observations we have about $\mathbb{R}^n$ to other mathematical objects. For now, we call these sets that have '$\mathbb{R}^n$'-vector-like properties **vector spaces**. This notion is a bit abstract, and we usually default to ^{n} to gain intuition about vector spaces and motivate new definitions and techniques. It just so happens that *every finite dimensional real vector space is 'almost the same' as* $\mathbb{R}^n$, so many conclusions we get from $\mathbb{R}^n$ naturally extend to other vector spaces.

In the previous chapter, one proposition lists 8 characteristics of $\mathbb{R}^n$. Let us extract all these 8 statements, and define a vector space to be a set that satisfies these properties.

Definition 2.2 (Vector Space)

Let $\mathbb{K}$ be a field, and V be a set closed under operations addition $+ : V \times V \to V$ and multiplication $\cdot : \mathbb{K} \times V \to V$. We call V a **vector space**, or a $\mathbb{K}$**-vector space** to specify the field if the following axioms hold.

1. *(Associativity)*$x + (y + z) = (x + y) + z \ \forall x, y, z \in V$.
2. *(Commutativity)* $x + y = y + x \ \forall x, y \in V$.
3. *(Identity)* There exists some vector 0_V s.t. $x + 0_V = x \ \forall x \in V$.
4. *(Inverse)* $\forall x \in V$, $\exists y \in V$ s.t. $x + y = 0_v$.
5. *(Scalar multiplication)* $a \cdot (b \cdot x) = (a \cdot b) \cdot x \ \forall a, b \in \mathbb{K}, x \in V$.
6. *(Scalar Identity)* $1 \cdot x = x \ \forall x \in V$.
7. *(Distributivity 1)* $a \cdot (x + y) = a \cdot x + a \cdot y \ \forall a \in \mathbb{K}, x, y \in V$.
8. *(Distributivity 2)* $(a + b) \cdot x = a \cdot x + b \cdot x \ \forall a, b \in \mathbb{K}, x \in V$.

Example 2.3

The following are examples of vector spaces.

- $\mathbb{R}^n$, with addition and multiplication as we have defined so far.
- $\{0_V\}$, the set containing just the zero vector, is a vector space over any field.
- The set of polynomials with degree 3 or less. Addition and multiplication are defined as $(f + g)(x) = f(x) + g(x)$ and $cf(x) = c \times f(x)$.
- $\mathbb{R}$ is a $\mathbb{Q}$-vector space, where $\mathbb{Q}$ is the set of all rationals (fractions).

The following are non-examples of vector spaces.

- ϕ, the empty set is not a vector space over any field as it does not have a zero vector.
- The set of polynomials with degree 3 or more (using the addition and multiplication rules defined above).
- $\mathbb{Q}$ is not an $\mathbb{R}$-vector space. This is because $\pi(1) = \pi$ which is not in $\mathbb{Q}$.

More frequently, we make new vector spaces from taking subsets of existing vector spaces. For instance, there might be a subset of $\mathbb{R}^n$ that also satisfies all the axioms of being a vector space.

Definition 2.4 (Subspace)

Let V be a $\mathbb{K}$ vector space. We call $W \subseteq V$ a **subspace** of V if W forms a vector space using the inherited operations $+ : W \times W \to W$ and $\cdot : \mathbb{K} \times W \to W$.

Remark. *Because the inherited operations will automatically satisfy the axioms of a vector space,* *it suffices to show that (1) W is nonempty and (2) W is closed under vector addition and scalar multiplication to confirm that W is a subspace.*

The condition W is nonempty is required because the identity axiom requires a zero vector, which one obtains by multiplying 0 to an arbitrary vector in the subset.

Example 2.5

1. In $\mathbb{R}$-vector space $\mathbb{R}^3$, the set of all vectors in the form of $k\vec{i}$ forms a subspace.
2. Take $\mathbb{R}$ as a $\mathbb{Q}$-vector space. $\mathbb{Q} \subset \mathbb{R}$ is a subspace.
3. Take $\mathbb{R}$ as a $\mathbb{R}$-vector space. $\mathbb{Q} \subset \mathbb{R}$ is **not** a subspace because it is not closed under multiplication.

Proposition 2.6

Let $W, U \subseteq V$ be subspaces. The intersection $W \cap U$ is a subspace of V.

Proof. Both W and U contain 0_V, so the intersection is non-empty. We also have for $v_1, v_2 \in W \cap U$,

$$v_1 + v_2 \in U, \quad v_1 + v_2 \in W,$$

so addition is closed in $W \cap U$. Similarly, scalar multiplication closed under W and U, so is closed in the intersection. □

Exercises

1. Determine if the following sets are real vector spaces:
   (a) The set of real polynomials $f(x) = ax^3 + bx^2 + cx + d$ for all real constants a, b, c, d, such that $f\prime(x) = 0$.
   (b) The set of continuous functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(0) = 1$.
2. Give an example of a vector space that only has one subspace.
3. Give a subspace of $\mathbb{R}^4$ (that is not $\mathbb{R}^4$) that contains the vector $(-1, 3, 0, 7)$.

2.1 Span, Linear Independence

2.1.1 Span

Throughout this section, let $\mathbb{K}$ be a field. We constrain ourselves to work in V, a $\mathbb{K}$-vector space.

Definition 2.7 (Span)

Let k be some positive integer. Let $\{v_1, v_2, ..., v_k\} \subseteq V$. The **span** of $\{v_1, v_2, ..., v_k\}$ is denoted as

$$\text{span}(v_1, v_2, ..., v_k)$$

and is the *smallest* subspace of V containing $\{v_1, v_2, ..., v_k\}$. If $\text{span}(v_1, v_2, ..., v_k) = V$, we say that $\{v_1, v_2, ..., v_k\}$ is a spanning set of V, or $\{v_1, v_2, ..., v_k\}$ spans V.

The *smallest* here means that if another subspace W contains $\{v_1, ..., v_k\}$, W cannot be a subset of $\text{span}(v_1, ..., v_k)$. How do we know such a subspace exists? We can take the intersection of all the subspaces containing $\{v_1, ..., v_k\}$

$$\text{span}(v_1, ..., v_k) = \bigcap_{W \text{ subspace containing } \{v_1,...,v_k\}} W$$

which is a subspace containing $\{v_1, ..., v_k\}$ and is a subset of all other subspaces containing $\{v_1, ..., v_k\}$. We know $V \subseteq V$ is a subspace containing $\{v_1, ..., v_k\}$, so the intersection is between at least one set and is thus well-defined.

Proposition 2.8

The span of $\{v_1, v_2, \ ..., v_k\} \subseteq V$ is all linear combinations of the vectors in the set:

$$\text{span}\{v_1, v_2, \ ..., v_k\} = \{c_1v_1 + c_2v_2 + ... + c_kv_k | c_1, c_2, ...c_k \in \mathbb{K}\}$$

Proof. We first show $\text{span}\{v_1, v_2, ..., v_k\} \supseteq \{c_1v_1 + c_2v_2 + ... + c_kv_k | c_1, c_2, ...c_k \in \mathbb{K}\}$. That is, for every W subset containing $v_1, ..., v_k$, W must also contain $c_1v_1 + ... + c_kv_k$.
We now show $\text{span}\{v_1, v_2, ..., v_k\} \subseteq \{c_1v_1 + c_2v_2 + ... + c_kv_k | c_1, c_2, ...c_k \in \mathbb{K}\}$. The right side is a set that nonempty, is closed under addition and scalar multiplication, and contains $v_1 = 1v_1 + 0v_2 + ... + 0v_k$,...,$v_k = 0v_1 + ... + 0v_{k-1} + 1v_k$. Therefore, it is one of the W subspaces whose intersection is used to construct the span. $\square$

It can be difficult to how to think about what the span of a set of vectors looks like, though it is also important to develop an intuition for it as more complex techniques are developed. It is also important to consider what vectors span a given subspace.

Example 2.9

Sketch, in $\mathbb{R}^3$, the following spans: $\text{span}(\{\vec{i}\})$, $\text{span}(\{\vec{i}, \vec{j}\})$, $\text{span}(\{\vec{i}, \vec{j}, \vec{k}\})$.

TODO: write something here

2.1.2 Linear Independence

Thinking geometrically in the previous example, the span of one vector is one-dimensional (a line), the span of two vectors is two-dimensional (a plane), and the span of three vectors is three-dimensional (the whole 3D space). Specifically, we can give a correspondence between the linear combination of k vectors $\{c_1v_1, ..., c_kv_k\}$and a point in $\mathbb{R}^k$ as

$$c_1v_1 + c_2v_2 + ... + c_kv_k \sim \begin{bmatrix} c_1 \\ c_2 \\ ... \\ c_k \end{bmatrix}$$

Assuming this correspondence works, the geomtry of the span should resemble $\mathbb{R}^k$. Unfortunately, this is not always true, for instance, consider the following counterexample:

Example 2.10

Sketch, in $\mathbb{R}^3$, $\text{span}(\vec{i}, \vec{k}, \vec{i} + 2\vec{k})$.

TODO:

What went wrong here is that the third vector $\vec{i} + 2\vec{k}$ is already a linear combination of the first two, and so this vector is 'redundant' in the set. More precisely, $(1, 2, 0)$ and $(0, 0, 1)$ give the same linear combination of the three vectors, so the correspondence fails. We thus want to answer the following question:

Given a set of k vectors $\{v_1, ..., v_k\}$, when do these vectors span the full k-dimensions? If they do not span the full k dimensions, how many dimensions do they span?

We already have one candidate criterion for spanning the full dimensions.

Definition 2.11 (Linear Independence)

Let $\{v_1, ..., v_k\} \subseteq V$. We say that $v_1, ..., v_k$ are **linearly independent** if every linear combination of $v_1, ..., v_k$ is unique. Precisely, if there are $c_1, ..., c_k, d_1, ..., d_k \in \mathbb{K}$ such that

$$c_1v_1 + c_2v_2 + ... + c_kv_k = d_1v_1 + d_2v_2 + ... + d_kv_k,$$

then

$$c_1 = d_1, c_2 = d_2, ..., c_k = d_k.$$

If $\{v_1, ..., v_k\}$ is not linearly independent, we say that it is **linearly dependent**.

This might be very tricky to verify, so we would like some simplier definitions for linear independence.

Proposition 2.12

The following are equivalent definitions for linear independence:

- (*) There is only one way to make 0_V. i.e. If $c_1v_1 + ... + c_kv_k = 0_V$, then $c_1 = c_2 = ... = c_k = 0$.
- (**) If $k \geq 2$, there is no way to express one vector as a linear combination as the others. i.e. For all $1 \leq j \leq k$, $v_j \notin \text{span}(v_1, ..., v_{j-1}, v_{j+1}, ..., v_k)$.

Proof. We first show the original definition of linear independence implies (*). We already have one way to create the zero vector as a linear combination of the set, namely

$$0v_1 + 0v_2 + ... + 0v_k = 0_V.$$

By definition of linear independence, this is the only way to create the zero vector.
We now show (*) implies the original definition of linear independence. Let (*) hold. Then if

$$c_1v_1 + ... + c_kv_k = d_1v_1 + ... + d_kv_k,$$

then rearranging the terms we will get

$$(c_1 - d_1)v_1 + ... + (c_k - d_k)v_k = 0_V.$$

Since there is only one way to make 0_V, it must be that $c_1 - d_1 = c_2 - d_2 = ... = c_k - d_k = 0$, or $c_1 = d_1, ...c_k = d_k$. This matches our original definition of linear independence.
Now we set $k \geq 2$ and show the equivelence of (*) and (**). We now show (*) implies (**). Let (*) hold, and $v \in \text{span}(v_1, ..., v_{j-1}, v_{j+1}, ..., v_k)$. Then $v - v_j$ is a linear combination of $v_1, ..., v_k$ that is not $0v_1 + ... + 0v_k$ as the coefficient before v_j is -1. So that

$$v - v_j \neq 0_V \implies v \neq v_j.$$

What we have shown is that everything in $\text{span}(v_1, ..., v_{j-1}, v_{j+1}, ..., v_k)$ is not v_j. Which means $v_j \notin \text{span}(v_1, ..., v_{j-1}, v_{j+1}, ..., v_k)$.
One end of chapter exercise will guide you through the direction (**) implies (*). □

TODO: introduce contraposition and contradiction?

We now have an informal statement: *If the set is not linearly independent, then the dimension of the span must be less than k.* This is because from (**), at least one of the vectors is redundant, so that we can remove that vector and still produce the same span with $k - 1$ vectors.

2.1.3 Basis

Finally, we have a special notion when the set $\{v_1, \ldots, v_k\}$ span V and are linearly independent. This means that every $v \in V$ is *a unique combination* of $\{v_1, \ldots, v_k\}$.

Definition 2.13 (Basis)

Let $\{v_1, ..., v_k\} \subseteq V$. We say that $v_1, ..., v_k$ form a **basis** for V if they span V and are linearly independent.

Example 2.14

In $\mathbb{R}^n$, the standard basis vectors $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ are a basis. This is why we call them the standard basis.

Definition 2.15 (Dimension)

If a vector space V has a finite basis, we call V finite dimensional. Else we call V infinite dimensional.

Remark. *It is true that every basis of a vector space has the same number of elements, so we can assign a number (or infinity) to the dimension of a vector space. However, we will have to prove that later when we develop more technology.*

Theorem 2.16 (Every Vector Space has a Basis)

Let V be a $\mathbb{K}$ vector space. Then V has a basis.

Understanding this proof is optional. Our main hurdle here is that vector spaces are not necessarily 'small' to work with. For instance, $\mathbb{R}^n$ is finite dimensional, so is in some sense small. But now consider the following real vector space:

$$\tilde{V} = \{(a_1, a_2, a_3, \ldots) \mid a_j \in \mathbb{R}\}$$

Each element in this set is an infinite sequence of real numbers, and addition/scaling is entry-wise. You may argue that the infinite set $S = \{(1,0,0,\ldots),(0,1,0,\ldots),(0,0,1,\ldots),\ldots\}$ could be a basis. However, they do not span $\tilde{V}$. To show that, I claim that the subset

$$U = \{(a_1, a_2, a_3, \ldots) \in \tilde{V} \mid \text{only finitely many } a_j\text{'s are non zero}\}$$

is a subspace of $\tilde{V}$. It is a strict subset of $\tilde{V}$, contains the zero vector $(0,0,0,\ldots)$ and is closed under addition and $\mathbb{R}$-scaling. It also contains S as each element in S has 1 (thus finitely many) non-zero element(s). Thus $\text{span}(S) \subseteq U \subset \tilde{V}$.

To circumvent this, this proof requires the use of the axiom of choice, which we will give a statement below.

Definition 2.17 (Axiom of Choice)

The **Axiom of Choice** is one of the axioms in Zermelo-Frenkel (ZFC) Set theory. Its statement is as follows:

> Let I be a set, $(S_i)_{i\in I}$ be an indexed family of non-empty sets. Then there exists an indexed set $(s_i)_{i\in I}$ such that $s_i \in S_i$ for each $i \in I$.

Colloquially, it means there is a "choice function" that allows you to pick an element from each of the non-empty S_i's.

Remark. *What we lose from the Axiom of Choice is that the choice function is non-constructive. This means that we know that something exists without knowing exactly how to construct it. The Axiom of Choice also leads to weird consequences, for example the Banach-Tarski paradox states that you can rearrange the points of a solid 3D ball to make two solid 3D balls, each having the same size as the original. (!) While some mathematicians do not accept the Axiom of Choice and adopt ZF (ZFC without Choice), it generally makes our lives easier.*

From the Axiom of Choice, one can derive Zorn's Lemma. This is the optional part.

Theorem 2.18 (Zorn's Lemma)

Assume the Axiom of Choice. Let S be a partially ordered set, that is, there is a comparison function $\leq$ for some pairs of elements. Then if for every ascending chain in S

$$s_1, s_2, \ldots \in S \text{ such that } s_1 \leq s_2 \leq s_3 \leq \ldots,$$

There is some $t \in S$ such that $s_j \leq t$, then S has a maximal element (nothing is bigger than S).

Remark. *The example of Zorn's Lemma is as follows: Let S contain subsets of V. We have a partial ordering on S using inclusion.*

$$A \leq B \textit{ if } A \subseteq B.$$

If for every chain $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ we can find $B \in S$ that contains all the A_j's, then S has a maximal element.

Proof of Every Vector Space has a Basis. Let $S = \{L \subseteq V \mid \text{L is linearly independent}\}$. We give S the partial ordering with respect to set inclusion. We first show S has a maximal element using Zorn's Lemma.

For each ascending chain $L_1 \subseteq L_2 \subseteq \ldots$, we claim $U = \bigcup_i L_i$ is a linearly independent set. We verify it directly: Let $v_1, \ldots, v_k \in U, c_1, \ldots, c_k \in \mathbb{K}$, and $\sum_i c_i v_i = 0_V$. We know that for each $1 \leq i \leq k$ there L_{a_i} such that $v_i \in L_{a_i}$. Take $a = \max_i a_i$. Then $v_i \in L_a$. Because L_a is a linearly independent set, it must be that $c_1 = \ldots = c_k = 0$.

Now we apply Zorn's Lemma to get that there is a maximal element $M \in S$. We claim Span(M). If not, we can find $v \in V$ such that $v \notin \text{Span}(M)$, but then this means $M \cup \{v\}$ is a linearly independent set. This contradicts that M is a maximal element. ☐

Exercises

4. *(**) $\implies$ (*) in linear independence* Let $c_1 v_1 + \ldots + c_k v_k = 0_V$. If $c_1 \neq 0$, argue that $v_1 \in \text{span}(v_2, ..., v_k)$, thus this case cannot happen and $c_1 = 0$. Then show that $c_j = 0$ for all j, and thus the v_j's are linearly independent.

5. Write down a basis for the vector space of all polynomials of degree k or less.

2.2 Systems of Linear Equations

As we alluded to earlier, many types of vector spaces are in some way very similar to $\mathbb{R}^n$. For instance, the span of a set of k linearly independent real-vectors has a natural correspondence to $\mathbb{R}^k$. With the blind faith that everything here can generalize nicely back to abstract vector spaces, we limit ourselves again back to talking about $\mathbb{R}^n$.
Here, we introduce a new notation to write linear combiantions, as $c_1v_1 + ... + c_kv_k$ is very cumbersome.

Definition 2.19 (Matrix)

Let m and n be positive integers. An $m \times n$ **matrix** is a rectangular array of m rows and n columns in the form

$$\begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,n} \\ \vdots & & \ddots & \\ c_{m,1} & c_{m,2} & \cdots & c_{m,n} \end{bmatrix}$$

where each $c_{i,j}$ is an **entry** of the matrix. We denote the set of all $m \times n$ matrices with real entries $M_{m\times n}(\mathbb{R})$. In general, we have $M_{m\times n}(\mathbb{K})$ for entries in the field $\mathbb{K}$.

Notation. *Suggestively, we can write an $m \times n$ matrix as*

$$\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & ... & \vec{v}_n \end{bmatrix}$$

where each $\vec{v}_j \in \mathbb{R}^m$ is written as a column vector

$$\begin{bmatrix} c_{j,1} \\ c_{j,2} \\ \vdots \\ c_{j,m} \end{bmatrix}.$$

If we want to think of the matrix by its entries, we can also write the matrix as

$$\{c_{i,j}\}$$

Definition 2.20 (Matrix-vector Product)

Let

$$A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & ... & \vec{v}_n \end{bmatrix}$$

be an $m \times n$ matrix, and

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \in \mathbb{R}^n.$$

The product of A and $\vec{b}$ is evaluated as

$$A\vec{b} = b_1\vec{v}_1 + b_2\vec{v}_2 + ... + b_n\vec{v}_n.$$

Example 2.21

Let $M = \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \dots & \vec{e}_n \end{bmatrix}$ be an $n \times n$ matrix. Compute $M\vec{x}$ for any $\vec{x} \in \mathbb{R}^n$.

The computation is not too difficult. We can follow the definition to get

$$M\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \dots + x_n\vec{e}_n = \vec{x}.$$

Basically, multiplying a vector with the matrix with columns $\vec{e}_1, \dots, \vec{e}_n$ does not change the vector. This should not be surprising, since the corresponding system of equations for $M\vec{x} = \vec{b}$ is

$$\begin{aligned} x_1 &= b_1 \\ x_2 &= b_2 \\ &\vdots \\ x_n &= b_n \end{aligned}$$

Definition 2.22 (Identity Matrix)

We denote the $n \times n$ **identity matrix** by

$$I_n \stackrel{\text{def}}{=} \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \dots & \vec{e}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

Proposition 2.23

The matrix-vector product satisfies the following properties. Let $c \in \mathbb{R}, \vec{x}_1, \vec{x}_2 \in \mathbb{R}^n, A \in M_{m\times n}(\mathbb{R})$, then

- $A(\vec{x}_1 + \vec{x}_2) = A\vec{x}_1 + A\vec{x}_2$,
- $A(c\vec{x}_1) = c(A\vec{x}_1)$.

This means that the matrix-vector product preserves linear combinations. $A(c_1\vec{x}_1 + ... + c_k\vec{v}_k) = c_1A\vec{x}_1 + ... + c_kA\vec{x}_k$.

Example 2.24

Determine if the vectors

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

are linearly independent.

Recalling the definition for linear independence, we solve for

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

This is a compact way to express the system

$$\begin{aligned} c_1 + c_2 + 0 &= 0 \\ 0 + c_2 + c_3 &= 0 \\ c_1 + 0 + c_3 &= 0 \end{aligned}$$

There is a very slick way to find the solution (which involves adding all three equations together), but let us solve this methodically. This method is known as elimination, which involves isolating

variables, and thus reducing the complexity of the system one by one.
First, we look at the first equation and isolate $c_1 = -c_2$. Now we can look replace all instances of c_1 in the second and third equation with $-c_2$, giving us

$$\begin{aligned} c_1 + c_2 + 0 &= 0 \\ 0 + c_2 + c_3 &= 0 \\ 0 - c_2 + c_3 &= 0 \end{aligned}$$

If we look at just the second and third equations, these only have two variables, so we have effectively decreased the complexity of the system by 1. Whatever we get from the second and third equations, we can substitute back into the first to get c_1. Now, we repeat for the second equation, $c_2 = -c_3$, and replacing all instances of c_2 we have

$$\begin{aligned} c_1 + 0 - c_3 &= 0 \\ 0 + c_2 + c_3 &= 0 \\ 0 + 0 + 2c_3 &= 0 \end{aligned}$$

The third equation now is just an equation in one variable, so we can go ahead and solve

$$\begin{aligned} c_1 + 0 - c_3 &= 0 \\ 0 + c_2 + c_3 &= 0 \\ 0 + 0 + c_3 &= 0 \end{aligned}$$

and replace all instances of c_3 with 0 to get

$$\begin{aligned} c_1 + 0 + 0 &= 0 \\ 0 + c_2 + 0 &= 0 \\ 0 + 0 + c_3 &= 0 \end{aligned}$$

or $c_1 = c_2 = c_3 = 0$. This is indeed a solution to the system, so we can now conclude that these vectors are indeed linearly independent. If we condense the systems back to matrices (concatenating the 3×3 matrix and the vector on the right) we get

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 2 & 0 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

Example 2.25

Determine the solutions to the system of equations

$$\begin{aligned} 0 + 2x_2 - x_3 &= 1 \\ x_1 - x_2 + x_3 &= 0 \\ x_1 + x_2 + 2x_3 &= 1 \end{aligned}$$

The first equation here does not let us isolate x_1, but we can simply exchange the first two equations to get

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 2 & -1 & 1 \\ 1 & 1 & 2 & 1 \end{array}\right]$$

We want to remove any dependence of x_1 for the other two equations, so we can subtract the first equation from the third to get

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 2 & -1 & 1 \\ 0 & 2 & 1 & 1 \end{array}\right]$$

We divide the second equation by 2, add this equation to the first, and subtract from the third.

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 1 & \frac{-1}{2} & \frac{1}{2} \\ 0 & 2 & 1 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{-1}{2} & \frac{1}{2} \\ 0 & 2 & 1 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{-1}{2} & \frac{1}{2} \\ 0 & 0 & 2 & 0 \end{array}\right]$$

Finally, we divide the third equation by 2, and remove all other entries of the third column.

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \end{array}\right]$$

So the solution is $x_1 = 1/2, x_2 = 1/2, x_3 = 0$.

2.2.1 Elementary Row Operations and Row Echelon Form

By how we manipulated the equations in the previous examples, we motivate these operations on matrices.

Definition 2.26 (Elementary Row Operations and row equivalence)

The following operations on a matrix $A \in M_{m\times n}$ are known as **elementary row operations**:

1. *(Row Swap)* Exchange any two rows.
2. *(Scaling)* Multiply a row by a **non-zero** constant.
3. *(Sum)* Add a multiple of a row to another row.

Let $B \in M_{m\times n}$. We say that A is **row equivalent** to B if A can be transformed into B By applying a sequence elementary row operations. We denote this equivalence by $A \sim B$.

Remark. *We can group the big set of $M_{m\times n}(\mathbb{R}^n)$ into classes, where each class contains matrices that are row equivalent to each other.*

Proposition 2.27

Elementary row operations are invertible. That is, if $A \sim A'$ after applying an elementary row operation, $A' \sim A$ through applying a (possibly different) row operation.

Proposition 2.28

An $m \times n$ matrix can be viewed as a system of m equations in $n-1$ variables using the representation in the previous two examples. In this representation, row equivalent matrices have the same solution sets.

The proof is not very enlightening. It boils down to checking that each of the three elementary row operations do not change the solution set of the corresponding linear system, so a sequence of them will not change the solution sets. The key takeaway from this is that we can reduce the complexity of the matrix through elementary row operations. Let us define the 'simple' forms of a matrix you can get through row operations.

Definition 2.29 (Row Echelon Form and Pivots)

A matrix is in **row echelon form** if

- Rows with all zero entries are on the bottom.
- For each row having non-zero entries, the first non-zero entry is on the right of the first non-zero entry of the row above.

In row echelon form, the first non-zero entry of a row is called a **pivot**. Colloquially, the non-zero entries form a "staircase" like shape.

Example 2.30

The following matrix is in row echelon form:

$$\begin{bmatrix} 1 & * & * & * & * \\ 0 & 0 & 2 & * & * \\ 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The pivots of this matrix, from the first row, are 1, 2 and 1.

In most cases, reducing a matrix into the row echelon form is "good enough" to find solutions. Start from the bottom row, and substitute variables go up. However, one problem with working with row echelon form is that a matrix is row equivalent to many matrices in row echelon form, so we want some 'super' row echelon form that is unique to each matrix.

Definition 2.31 (Reduced Row Echelon Form)

A matrix is in **Reduced Row Echelon Form**(rref) if

- It is in row echelon form.
- All pivots are equal to 1.
- Every pivot is the only non-zero entry in its column.

Example 2.32

The following matrix is in reduced row echelon form:

$$\begin{bmatrix} 1 & * & 0 & 0 & * \\ 0 & 0 & 1 & 0 & * \\ 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

By how we defined the reduced row echelon form, we cannot find two distinct matrices in reduced row echelon form that are also row equivalent. This guarantees that a matrix is row equivalent to at most one reduced row echelon form matrix.

Theorem 2.33 (Existence and Uniqueness of Reduced Row Echelon Form)

Let $A \in M_{m\times n}(\mathbb{R})$. Then there exists a unique $A' \in M_{m\times n}(\mathbb{R})$ such that A' is in reduced row echelon form and $A \sim A'$. We define A' to be the **reduced row echelon form** of A, and denote

$$A' = \mathrm{rref}(A).$$

Proof. Formalizing our steps in the previous examples, we have an algorithm for reducing a matrix into rref.

TODO: Gaussian-Jordan Reduction □

Example 2.34

Compute

$$\mathrm{rref}\left(\begin{bmatrix} somethinghere lol \end{bmatrix}\right)$$

Algorithm 1: Gaussian-Jordan Reduction of $A \in M_{m\times n}(\mathbb{R}^n)$

```
i ← 1, j ← 1 while j ≤ n do
    for k in i...m do
        /* Find a row with non-zero entry in that column */
        if a_{k,j} ≠ 0 then
            swap(row i, row k)                      /* Move row k to the top row */
            row i ← 1/a_{i,j} × row i               /* Normalize pivot to 1 */
            for l in 1...m, l ≠ i do
                row l ← row l - (a_{l,j} × row i)   /* all other entries in the same column = 0 */
            end
            i ← i + 1                               /* Next row with pivot goes to the row below */
            Break
        end
    end
    j ← j + 1                                       /* check pivot in next column */
end
```

2.2.2 Recovering solutions from rref

Producing one solution

Let us consider a general matrix in rref:

$$\left[\begin{array}{ccccc|c} 1 & * & * & * & * & a \\ 0 & 0 & 1 & * & * & b \\ 0 & 0 & 0 & 0 & 1 & c \\ 0 & 0 & 0 & 0 & 0 & d \end{array}\right]$$

If $d \neq 0$, the rref form will imply d is a pivot and thus equal to 1. So the last equation requires $0 = 1$, which means it has no solutions!

Now we set $d = 0$, so a, b, c can be anything. Regardless of the * entries, we can set for the columns with pivots, $x_1 = a, x_3 = b, x_5 = c$, and for the columns corresponding to non-pivots, $x_2 = x_4 = 0$. We thus have

Theorem 2.35 (Solutions to linear systems)

Let $A \in M_{m\times n}(\mathbb{R}), \vec{b} \in \mathbb{R}^m$, the system of equations described by $A\vec{x} = \vec{b}$ has

$$\begin{cases} \text{No solutions, if rref}\left(\left[A|\vec{b}\right]\right) \text{ has a pivot in the last column,} \\ \text{At least one solution, if the last column does not have a pivot} \end{cases}$$

Moreover, if the latter case holds, one solution (a particular solution) can be constructed from the rref by just solving the columns of rref$\left(\left[A|\vec{b}\right]\right)$ with pivots, and setting the variables corresponding to non-pivot columns to 0.

Producing all solutions

Now, suppose $A\vec{x} = \vec{b}$ has at least one solution. It can have infinitely many solutions! We want to describe this set of solutions without actually writing infinitely many vectors. One starting point is to see what properties this set satisfies. One starting point would be to guess that the set $\{\vec{x} \in \mathbb{R}^n | A\vec{x} = \vec{b}\}$ is a subspace of $\mathbb{R}^n$. This is generally not true, as every subspace contains $\vec{0}_{\mathbb{R}^n}$, and since $A\vec{0}_{\mathbb{R}^n} = \vec{0}_{\mathbb{R}^m}$, the only possible way this could be a subspace is when $\vec{b} = \vec{0}_{\mathbb{R}^m}$.

We thus start with the solution set to $A\vec{x} = \vec{0}$.

Definition 2.36 (Nullspace)

Let $A \in M_{m\times n}(\mathbb{R})$. The **nullspace** of the A, denoted $N(A)$, is the set of all solutions $\vec{x}$ for $A\vec{x} = \vec{0}$. i.e.

$$N(A) = \{\vec{x} \in \mathbb{R}^n | A\vec{x} = \vec{0}_{\mathbb{R}^m}\}.$$

Proposition 2.37

Let $A \in M_{m\times n}(\mathbb{R})$.

$$N(A) = N(\mathrm{rref}(A)).$$

Proof. **TODO:** □

This allows us to limit the discussion of nullspaces to when A is in rref.

It turns out that the nullspace is a subspace. This is why we tend to like systems of linear equations where the right hand side is all zero. We call these systems **homogeneous**, and the solutions to homogeneous linear systems form a subspace.

Theorem 2.38 (Nullspace is a Subspace)

Let $A \in M_{m\times n}(\mathbb{R})$. $N(A)$ is a subspace of $\mathbb{R}^n$.

Proof. **TODO:** zero is in here, closed under addition and multiplication □

Example 2.39

Find the solutions to the system represented by

$$\left[\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -3 & 0 \\ 0 & 1 & 1 & 0 & 4 & 0 \\ 0 & 0 & 0 & 1 & -2 & 0 \end{array}\right].$$

Express the answer in the form of the span of a set of linearly independent vectors.

Let us start with the last variable x_5 and work our way towards x_1. There is no equation that fixes x_5, so let $x_5 = s$ for some $s \in \mathbb{R}$. Then the last equations fixes $x_4 = 2s$.

Again, we do not have an equation that fixes x_3, so let $x_3 = t$ for some $t \in \mathbb{R}$. This will force (from the second equation) $x_2 = -t - 4s$ and (from the first equation) $x_1 = -t + 3s$.

Our solution vector $(x_1, x_2, x_3, x_4, x_5)$ thus has the form

$$\begin{bmatrix} -t+3s \\ -t-4s \\ t \\ 2s \\ s \end{bmatrix} = s \begin{bmatrix} 3 \\ -4 \\ 0 \\ 2 \\ 1 \end{bmatrix} + t \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

This describes exactly the span of the vectors

$$\begin{bmatrix} 3 \\ -4 \\ 0 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

To confirm these vectors are linearly independent, notice the third entry of the first vector and the fifth entry of the second vector is 0.

From this example, we have an algorithm to solve for the nullspace as a span of vectors.

Algorithm 2: Generating the Nullspace of Matrix A

```
1  A ← rref(A)
2  S ← empty set
3  for each column j without a pivot do
4  |   v⃗ ← e⃗_j
5  |   for each row i with a pivot do
6  |   |   Locate pivot of row i in column k
7  |   |   v⃗ ← v⃗ − a_{i,j} e⃗_k
8  |   end
9  |   Add v⃗ to S
10 end
11 return span(S)
```

This is what we did in the previous example, except written in pseudocode. The idea is that each column without a pivot denotes a free variable, while each column with a pivot is a fixed by the free variables. This is what we do in lines 5 to 8, after fixing our free variable to be 1, we subtract from the fixed variables. In particular, *there is at most one solution if every column has a pivot.* Now we return back to solving $A\vec{x} = \vec{b}$. We showed that the solutions do not form a subspace. However, these solutions are related to the nullspace.

Example 2.40

Find the solutions to the system represented by

$$\left[\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -3 & 1 \\ 0 & 1 & 1 & 0 & 4 & 2 \\ 0 & 0 & 0 & 1 & -2 & 3 \end{array}\right].$$

Similar to computing nullspace, let us start with the last variable x_5 and work our way towards x_1. There is no equation that fixes x_5, so let $x_5 = s$ for some $s \in \mathbb{R}$. Then the last equations fixes $x_4 = 3 + 2s$.

Again, we do not have an equation that fixes x_3, so let $x_3 = t$ for some $t \in \mathbb{R}$. This will force (from the second equation) $x_2 = 2 - t - 4s$ and (from the first equation) $x_1 = 1 - t + 3s$.

Our solution vector $(x_1, x_2, x_3, x_4, x_5)$ thus has the form

$$\begin{bmatrix} 1-t+3s \\ 2-t-4s \\ t \\ 3+2s \\ s \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \\ 0 \end{bmatrix} + s \begin{bmatrix} 3 \\ -4 \\ 0 \\ 2 \\ 1 \end{bmatrix} + t \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

This is the nullspace, but shifted by the vector $(1, 2, 0, 3, 0)$.

Proposition 2.41

Let $\vec{x}_1, \vec{x}_2 \in \mathbb{R}^n$ be solutions to $A\vec{x} = \vec{b}$.Then

$$(\vec{x}_1 - \vec{x}_2) \in N(A).$$

Proof. We compute directly that $A(\vec{x}_1 - \vec{x}_2) = \vec{b} - \vec{b} = \vec{0}$. □

Corollary 2.42: Let $\vec{v}$ be *one* solution to $A\vec{x} = \vec{b}$. Then all the solutions of $A\vec{x} = \vec{b}$ is exactly

$$\{\vec{v} + \vec{n} | \vec{n} \in N(A)\}.$$

This means we only need one solution (particular solution), and the nullspace (homogeneous) solution to solve any matrix. This one solution can be obtained by setting all the 'free variables' in the solution to zero, and solving for all the pivots.

When producing solutions, the number of pivots of rref(A) is important. Let us define the number of pivots to be the rank of A.

Definition 2.43 (Rank)

Let $A \in M_{m\times n}(\mathbb{R})$. We define the **rank** of A

$$\operatorname{rank}(A) = \text{the number of pivots in } \operatorname{rref}(A).$$

We can thus formulate having zero, one or infinitely many solutions in terms of the number of pivots.

Proposition 2.44

Let $A \in M_{m\times n}(\mathbb{R}), \vec{b} \in \mathbb{R}^m$, the system of equations described by $A\vec{x} = \vec{b}$ has

$$\begin{cases} \text{At least one solution, if } \operatorname{rank}\left(\left[A|\vec{b}\right]\right) = \operatorname{rank}(A) \\ \text{At most one solution, if } \operatorname{rank}(A) = n. \end{cases}$$

Going back to why we started solving linear systems, we want to find a test to see if the vectors are linearly independent, or if some vector is in the span of a set of vectors. Writing $A = [\vec{v}_1 \ \dots \ \vec{v}_n]$, $A\vec{x} = \vec{b}$ having at least one solution means $\vec{b} \in \operatorname{span}(\vec{v}_1, \dots, \vec{v}_n)$, and if $A\vec{x} = \vec{0}$ has exactly one solution, the $\vec{v}_j$'s are linearly independent.

Theorem 2.45 (Rank and span/linear independence correspondence)

Let $A = [\vec{v}_1 \ \dots \ \vec{v}_n] \in M_{m\times n}(\mathbb{R})$, $\vec{b} \in \mathbb{R}^m$. Then

- $\operatorname{rank}\left(\left[A|\vec{b}\right]\right) = \operatorname{rank}(A) \iff \vec{b} \in \operatorname{span}(\vec{v}_1, \dots, \vec{v}_n)$.
- $\operatorname{rank}(A) = m \iff \operatorname{span}(\vec{v}_1, \dots, \vec{v}_n) = \mathbb{R}^m$.
- $\operatorname{rank}(A) = n \iff \{\vec{v}_1, \dots, \vec{v}_n\}$ is linearly independent.

Remark. *This is the first time we used the $\iff$ symbol. This means 'if and only if', or 'iff' for short. '$A \iff B$' means that if A holds, then B holds. And if B holds, A holds. For instance, 'more than 99% guarantees an A' means*

You have more than 99% in this course $\implies$ you get an A

But you can still get an A without reaching this score, for instance a 95% could get an A too. An if and only if statement (for a strict grading class) would look more like

The threshold for A in this class is a 90%.

so {You get A} $\iff$ {You get 90+%}.

Proof. The first and third statements are evident from the definitions of span and linear independence.

To show the second statement, we first prove the $\implies$ direction. Let $\vec{b} \in \mathbb{R}^m$, then reducing $[A|\vec{b}]$ would give m pivots in the first n columns, so there cannot be a pivot in the final column. (or else, we would have $m+1$ pivots in a matrix with only m rows). Therefore, the system has a solution.

For the $\impliedby$ direction, we can reverse all the elementary operations from A to $\operatorname{rref}(A)$ to get from $[\operatorname{rref}(A)|\vec{e}_m]$ to $[A|\vec{w}]$ for some $\vec{w} \in \mathbb{R}^m$. This system of equations has no solution. $\square$

Something amazing also happens when the vectors are a basis (recall it means both linearly independent and spans $\mathbb{R}^m$).

The second and third statements forces $m = \text{rank}(A) = n$, so A is a square matrix. The vectors span $\mathbb{R}^m$ iff they are linearly independent. Moreover, since $\text{rank}A = n$ for this square matrix, $\text{rref}A = I_n$, the identity matrix.

Theorem 2.46 (This One/ That One)

Let $A \in M_{n\times n}(\mathbb{R})$, and write $A = [\vec{v}_1 \dots \vec{v}_n]$. Then the following are equivalent.

1. $\{\vec{v}_n, \dots, \vec{v}_n\}$ form a basis.
2. $\{\vec{v}_n, \dots, \vec{v}_n\}$ are linearly independent.
3. $\{\vec{v}_n, \dots, \vec{v}_n\}$ span $\mathbb{R}^n$.
4. The system of equations $A\vec{x} = \vec{b}$ has exactly one solution for each $\vec{b} \in \mathbb{R}^n$.
5. The system of equations $A\vec{x} = \vec{b}$ has at least one solution for each $\vec{b} \in \mathbb{R}^n$.
6. The system of equations $A\vec{x} = \vec{b}$ has at most one solution for each $\vec{b} \in \mathbb{R}^n$.
7. $\text{Rank}(A) = n$.
8. $\text{rref}(A)$ has a pivot in every column.
9. $\text{rref}(A)$ has a pivot in every row.
10. $\text{rref}(A) = I_n$
11. $N(A) = \{\vec{0}\}$.

Remark. *This One Theorem goes by many names: Invertibility Theorem, Inverse Matrix Theorem, Inverse Function Theorem, and (my favorite) Amazingly Awesome Theorem by Prof. Cañez. Despite the lack of standardization, it is covered in every linera algebra course in some form or another. I'm putting it as 'That One Theorem', so when you quote 'this follows from That One Theorem' everyone will understand which one you are talking about.*

Remark. *This One Theorem is still incomplete. We will build towards more equivalences.*

Exercises

6. Determine if the following sets of vectors are linearly dependent or independent. If they are not linearly independent, generate the set of homogenous solutions to the corresponding linear system.
   (a) $\{(1, 2, 3), (0, 1, 1), (1, 1, 2)\}$
   (b) $\{\vec{e}_1, \vec{e}_2, \vec{e}_5\} \subset \mathbb{R}^7$
   (c)
7. Let $m > n$. Show that every set of m vectors in $\mathbb{R}^n$ is not linearly independent. Show that every set of n vectors in $\mathbb{R}^m$ does not span m. Thus show that every basis for $\mathbb{R}^n$ has n vectors.

8. Let $\{v_1, ..., v_k\} \subset V$. Show that exactly one of the following statements hold:

   (a) For every $v \in \text{span}(\{v_1, ..., v_k\})$, there is exactly one way to write v as a linear combination of $v_1, ..., v_k$.

   (b) For every $v \in \text{span}(\{v_1, ..., v_k\})$, there is more than one way to write v as a linear combination of $v_1, ..., v_k$.

9. *Every vector space has a basis (weaker version)* This is a weaker statement of Every Vector Space has a Basis. To get around this, we use a finiteness assumption that our vector space V is a subspace of $\mathbb{R}^n$.

   (a) Suppose that our subspace is $\{\vec{0}\}$. Show that this is spanned by the empty set. The empty set is (vacuously) linearly independent*, so we have a basis. ** all elements of an empty set satisfy any condition*

   (b) Now suppose $S = \{\vec{v}_1, \ldots, \vec{v}_k\} \subset V \subseteq \mathbb{R}^n$. Show that this is a linearly independent set in V if and only if this is a linearly independent set $\mathbb{R}^n$, so we can call it linearly independent without worrying about which vector space we work in.

   (c) Let S be linearly independent. If $\text{span}(S) = V$ this will be a basis for V. Else show that there is some $\vec{v}_{k+1} \in V$ such that the set $S + \{\vec{v}_{k+1}\}$, so you can extend S to a bigger linearly independent set.

   (d) Show that the previous process terminates, i.e. you cannot add in infinitely many vectors and still make S linearly independent. Therefore, it stops at some k where $\text{span}(\{\vec{v}_1, \ldots, \vec{v}_k\}) = V$.

10. Show that every subset of a linearly independent set is linearly independent.

11.

2.3 Linear Transformations

Matrix-vector multiplication preserve linear combinations. Let us name this property.

Definition 2.47 (Linear Transformation)

Let V and W be $\mathbb{K}$-vector spaces. A function $f : V \to W$ is a **linear transformation** if it preserves linear combinations. Namely, for every $c \in \mathbb{K}, \vec{v}_1, \vec{v}_2 \in V$,

- $f(\vec{v}_1 + \vec{v}_2) = f(\vec{v}_1) + f(\vec{v}_2)$,
- $f(c\vec{v}_1) = cf(\vec{v}_1)$.

Many functions in the wild are linear transformations. This includes the functions we have introduced so far.

Example 2.48

The following are examples linear transformations:

1. A matrix $A \in M_{m\times n}(\mathbb{R})$ admits a linear transformation $F : \mathbb{R}^n \to \mathbb{R}^m$ by
$$F(\vec{x}) = A\vec{x}.$$
2. Let $\vec{v} \in \mathbb{R}^n$. $f(\vec{x}) = \vec{v} \cdot \vec{x}$ is a linear transformation from $\mathbb{R}^n \to \mathbb{R}$. The projection $\mathrm{Proj}_{\vec{v}}$ is a linear transformation from $\mathbb{R}^n \to \mathbb{R}^n$.
3. Let $\vec{v} \in \mathbb{R}^3$. $f(\vec{x}) = \vec{v} \times \vec{x}$ is a linear transformation from $\mathbb{R}^3 \to \mathbb{R}^3$.
4. Recall that the set of polynomials $f(x)$ is a real vector space. The differentiation operator $\frac{d}{dx}$ is a linear transformation from the set of polyomials to itself.
5. Integration over a bounded interval $\int_a^b$ is a linear transformation from the space of integrable functions to $\mathbb{R}$.

Proposition 2.49

Let U, V, W be $\mathbb{K}$-vector spaces, and $T_1 : U \to V, T_2 : U \to V, T_3 : V \to W$ be linear transformations, then the following are also linear transformations:

- $f_1 : U \to V,\ f_1(\vec{x}) = T_1(\vec{x}) + T_2(\vec{x})$
- $f_2 : U \to V,\ f_2(\vec{x}) = cT_1(\vec{x})$ for any $c \in \mathbb{K}$
- $f_3 : U \to W,\ f_3(\vec{x}) = T_3(T_1(\vec{x}))$.

The first example of linear transformation given $A\vec{x}$ is a very general type of linear transformation. The punchline is that it is **the** example of a linear transformation.

Theorem 2.50 (Matrix Representation of Linear Transformation)

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there is a matrix $A \in M_{m\times n}$ such that for all $\vec{x} \in \mathbb{R}^n$,
$$T(\vec{x}) = A\vec{x}.$$
We denote this matrix the **matrix representation** of T.

Remark. *Importantly, this theorem is only applicable for $\mathbb{R}^n \to \mathbb{R}^m$, both are* ***finite dimensional****. Some linear functions work with infinite dimensional (we will give a definition later) vector spaces. Intuitively, this means the analogous 'standard basis vectors' is an infinite set $\{\vec{e}_1, \vec{e}_2, \ldots\}$.*

Before we head to the details of the proof, here is the idea behind our constuction. Let us suppose that this matrix exists; how do we compute each entry of the matrix? The trick is to think of the matrix as $\begin{bmatrix}\vec{w}_1 & \vec{w}_2 & \ldots & \vec{w}_n\end{bmatrix}$. We can get a linear combination of these to equal any of the $\vec{w}_j$'s. Namely, using the standard basis vectors, $A\vec{e}_j = \vec{w}_j$. Since $A\vec{e}_j = T(\vec{e}_j)$That means our matrix must be in the form
$$\begin{bmatrix}T(\vec{e}_1)\ T(\vec{e}_2)\ \ldots\ T(\vec{e}_n)\end{bmatrix}.$$

Now all we need to do check that this matrix works.

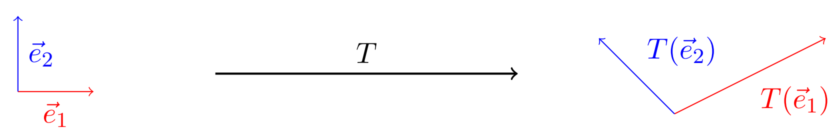
Proof. We claim that the matrix $A = \begin{bmatrix} T(\vec{e}_1) & T(\vec{e}_2) & \dots & T(\vec{e}_n) \end{bmatrix}$ satisfies $T(\vec{x}) = A\vec{x}$. Let $\vec{x} \in \mathbb{R}^n$. Then $\vec{x} = \sum_{i=1}^{n} x_i \vec{e}_i$. So that by linearity,

$$T(\vec{x}) = \sum_{i=1}^{n} x_i T(\vec{e}_i) = A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\vec{x}.$$

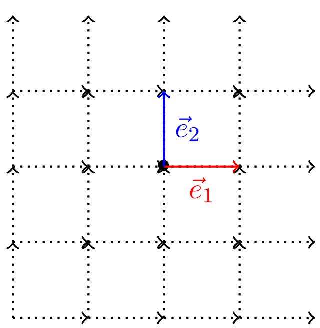
$\square$

More intuition behind the proof

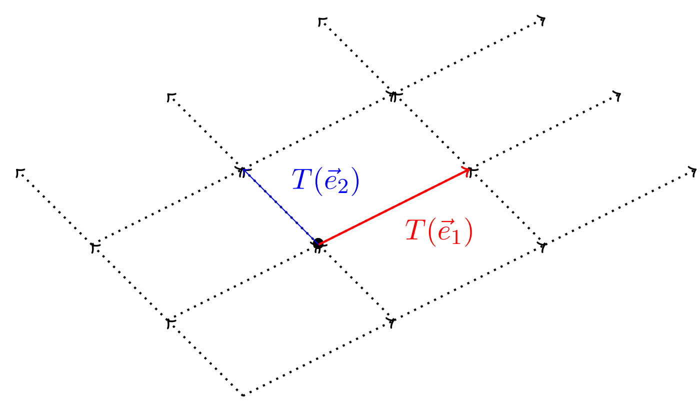
We can also think of a linear transformation geometrically. Let us track the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ that sends $T(\vec{e}_1) = (2, 1)$ and $T(\vec{e}_2) = (-1, 1)$. Let us only consider the integer lattice



points for now. If we start at (a, b), meaning from the origin, we take a steps right and b steps up. In our transformed plane, we need to take a steps in $(2, 1)$ and b steps $(-1, 1)$ by linearity. We can geometrically track the place of these lattic points using a grid. This is just a change of our



(a) Original grid



(b) Transformed grid

coordinate system. As long as we know where the standard basis vectors go (or any basis goes), it is enough to know the whole linear transformation. The construction is you 1. decompose any vector into the linear combination of the basis vectors, then 2. add up the corresponding linear combination of the transformed basis vectors. **This intuition of linear transformations being a shift in coordinates is very useful**, and helps transform abstract problems in linear algebra back to concrete geometry.

Example 2.51

Express the linear transformations of the dot product, cross product, and projection as matrix-vector multiplications.

Dot Product: Let $\vec{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$. Then $\vec{v} \cdot \vec{e}_j = v_j$. So

$$\vec{v} \cdot \vec{x} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \vec{x}$$

.

Cross Product Let $\vec{v} = (v_1, v_2, v_3) \in \mathbb{R}^n$. Then

$$\vec{v} \times \vec{x} = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix} \vec{x}$$

Projection Let $\vec{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$. Then

$$\mathrm{Proj}_{\vec{v}}(\vec{x}) = \frac{1}{|\vec{v}|^2} \begin{bmatrix} v_1 v_1 & v_1 v_2 & \dots & v_1 v_n \\ v_2 v_1 & v_2 v_2 & \dots & v_2 v_n \\ \vdots & & \ddots & \\ v_n v_1 & v_n v - 2 & \dots & v_n v_n \end{bmatrix} \vec{x}$$

where multiplication of matrix by scalar is entry wise.

Example 2.52

Let $T : \mathbb{R}^m \to \mathbb{R}^n$, $S : \mathbb{R}^n \to \mathbb{R}^p$. Let $A \in M_{m\times n}, B \in M_{n\times p}$ be the matrix representations of S and T respectively. We know that the function $F : \mathbb{R}^m \to \mathbb{R}^p$ defined by $F(\vec{x}) = S(T(\vec{x}))$ has a matrix representation. Compute this matrix.

The computation is not very interesting. However, you will do this until the day you drop ISP. If we write $A = \begin{bmatrix} \vec{x}_1 & \dots & \vec{x}_m \end{bmatrix}, \vec{b}_j \in \mathbb{R}^n$, then for each $\vec{e}_j \in \mathbb{R}^m$ we have $A\vec{e}_j = \vec{x}_j$. So that

$$S(T(\vec{e}_j)) = S(A\vec{e}_j) = S(\vec{x}_j) = B(\vec{x}_j).$$

The matrix is

$$\begin{bmatrix} B\vec{x}_1 & B\vec{x}_2 & \dots & B\vec{x}_m \end{bmatrix}.$$

Now, if we forcefully write the linear transformations in matrices, we get $S(T(\vec{x})) = BA\vec{x}$. This allows us to define the product of matrices.

Definition 2.53 (Matrix Operations)

Let $\alpha \in \mathbb{R}$, $A = \{a_{i,j}\}, B = \{b_i, j\} \in M_{m\times n}(\mathbb{R})$, $C = \{c_{i,j}\} \in M_{n\times p}(\mathbb{R})$. We define the addition of matrices

$$A + B = \{a_{i,j} + b_{i,j}\},$$

the pointwise addition of each entry. We define the scaling of a matrix

$$\alpha A = \{\alpha a_{i,j}\},$$

the pointise scaling of each entry. We also define the product of matrices

$$AC = \left\{\sum_{k=1}^{n} a_{i,k} c_{k_j}\right\}.$$

Remark. *The definition of matrix product is equivalent to what we got from getting the matrix from the composition of functions. The only difference is that we are composing a function the other way from $\mathbb{R}^p \to \mathbb{R}^n \to \mathbb{R}^m$.*

Remark. *There is a mnenomic to memorize the formula for matrix product AB. It involves putting A on the bottom left and B on the top right, like so:*

$$\begin{array}{cc} & \begin{bmatrix} b_{1,1} & \dots & b_{1,p} \\ \vdots & \ddots & \\ b_{n,1} & \dots & b_{n,p} \end{bmatrix} \\ \begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \ddots & \\ a_{m,1} & \dots & a_{m,n} \end{bmatrix} & \begin{bmatrix} \\ \textit{product goes here} \\ \\ \end{bmatrix} \end{array}$$

The product of an $m \times n$ and an $n \times p$ matrix is an $m \times p$ matrix. To find the i, j'th entry of the product, locate the i-th row on matrix A (left) and the j-th column on matrix B (top). Treating the row and column as vectors with n entries, the dot product is the i, j-th entry of the product.

Remark. *Viewing a column vector as a $n \times 1$ matrix, you can verify that the matrix-vector product is compatible with the products of two matrices.*

Proposition 2.54

The matrix operations satisfy the following properties:
Addition: Let $A, B, C \in M_{m\times n}(\mathbb{R})$, then:

- *(Associativity)* $A + (B + C) = (A + B) + C$.
- *(Commutativity)* $A + B = B + A$.
- *(Identity)* The zero matrix $0_{m\times n}$ satisfies $A + 0_{m\times n} = A$.

Scaling: Let $\alpha, \beta \in \mathbb{R}, A, B \in M_{m\times n}(\mathbb{R})$, then:

- *(Associativity)* $\alpha(\beta)A = (\alpha\beta)A$.
- *(Distributivity)* $(\alpha + \beta)(A + B) = \alpha A + \alpha B + \beta A + \beta B$.

Multiplication: Let $A, B \in M_{m\times n}(\mathbb{R}), C, D \in M_{n\times p}(\mathbb{R}), E \in M_{p\times q}(\mathbb{R})$.

$$\begin{tikzcd}
\mathbb{R}^q \arrow[r, "E"] & \mathbb{R}^p \arrow[r, bend left, "C"] \arrow[r, bend right, "D"'] & \mathbb{R}^n \arrow[loop below, "I_n"] \arrow[r, bend left, "A"] \arrow[r, bend right, "B"'] & \mathbb{R}^m
\end{tikzcd}$$

Then

- *(Associativity)* $A(CE) = (AC)E$.
- *Compatibility with Scaling* $A(\alpha C) = \alpha(AC) = (\alpha A)(C)$.
- *(Distributivity)* $(A + B)(C + D) = AC + AD + BC + BD$.
- *(Identity)* The identity matrix $I_n \overset{\text{def}}{=} \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \dots & \vec{e}_n \end{bmatrix}$ satisfies $I_n C = C$ and $BI_n = B$.

Remark. *The matrix product is generally not commutative. In fact, the product DB does not make sense, because the dimensions of the matrices are not compatible!*

Exercises

12. Find the corresponding matrices for the following transformations:

    (a)

13. Find an example where

    (a) Two matrices $A, B \in M_{n\times n}(\mathbb{R})$ are not commutative. i.e. $AB \neq BA$.
    (b) A matrix $0_{n\times n} \neq A \in M_{n\times n}(\mathbb{R})$ such that $A^k = 0_{n\times n}$ for some $k > 1$.

14. Let $\theta, \phi \in [0, 2\pi)$. Sketch how the matrix

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

transforms the grid. Show that

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{bmatrix}$$

and thus the set of matrix in this form are commutative. Is there a geometric argument for why these matrices must be commutative?

15. Run the following code in python multiple times and pinky promise that you got the correct answer.

```python
import numpy as np
matrix1=np.random.randint(-10,11,size=(3,3))
matrix2=np.random.randint(-10,11,size=(3,3))
print("Multiply these two matrices:")
print(matrix1,"\n",matrix2)
print("Confirm that the answer is the following:")
print(matrix1.dot(matrix2))
```

2.4 Surjective and Injective Linear Transformation

Definition 2.55 (Types of functions)

Let $f : U \to W$ be a function. We say f is:

- **Surjective**, if the image of U under f, $f(U) = W$. This means for every $w \in W$, there is some $u \in U$ such that $f(u) = w$.
- **Injective**, if each $u \in U$ maps to a unique element in W. This means for every $u_1, u_2 \in U$. If $f(u_1) = f(u_2)$, $u_1 = u_2$.
- **Bijective**, if f is surjective and injective.

Example 2.56

The following functions are surjective:

- The 'extract' function $\Pi_j : \mathbb{R}^n \to \mathbb{R}$, defined by $\Pi_j(x_1, x_2, ..., x_n) = x_j$ and extracts the j-th coordinate.
- The differentiation operator $\frac{d}{dx}$ from the space of **differentiable** functions to the space of **continuous** functions. This is because all continuous functions have antiderivatives. Moreover, as all derivatives are continuous, the differentiation operator is well defined.
- Integration over the interval $[0, 1]$ sends the integrable function g to the real number $\int_0^1 g(x)\ dx$. This is surjective because the integral of $g(x) = a$ over $[0, 1]$ is a for all $a \in \mathbb{R}$.

The following functions are injective:

- The 'pad zeros' function $\iota : \mathbb{R}^n \to \mathbb{R}^{n+1}$ defined by $\iota(x_1, ..., x_n) = (x_1, ..., x_n, 0)$ and adds a 0 to the final coordinate.
- The linear function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2x$.

The following functions are surjective:

- The identity map $Id : \mathbb{R}^n \to \mathbb{R}^n$ defined by $Id(\vec{x}) = I_n\vec{x} = \vec{x}$.

We can phrase these characterizations in of the number of solutions in u that solve $f(u) = w$.

Proposition 2.57

- f is surjective $\iff$ $f(u) = w$ has *at least one solution* in u for each $w \in W$.
- f is injective $\iff$ $f(u) = w$ has *at most one solution* in u for each $w \in W$.
- f is bijective $\iff$ $f(u) = w$ has *exactly one solution* in u for each $w \in W$.

When f is a linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$, it has a matrix representation $A \in M_{m\times n}(\mathbb{R})$. To determine if f is surjective, injective, or bijective, we look at the number of solutions in $\vec{x}$ for the equation $A\vec{x} = \vec{b}$ for each $\vec{b} \in \mathbb{R}^m$.

Theorem 2.58 (Surjectivity and solutions to linear systems)

Let $T : \mathbb{R}^n \to \mathbb{R}^m$, $A \in M_{m\times n}(\mathbb{R})$, and $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Then the following are equivalent:

1. T is surjective.
2. $A(\vec{x}) = \vec{b}$ is at least one solution in $\vec{x}$ for each $\vec{b} \in \mathbb{R}^m$
3. rref(A) has a pivot in every row.
4. The columns of A span $\mathbb{R}^m$.

Theorem 2.59 (Injectivity and solutions to linear systems)

Let $T : \mathbb{R}^n \to \mathbb{R}^m$, $A \in M_{m\times n}(\mathbb{R})$, and $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Then the following are equivalent:

1. T is injective.
2. $A(\vec{x}) = \vec{b}$ is at most one solution in $\vec{x}$ for each $\vec{b} \in \mathbb{R}^m$.
3. rref(A) has a pivot in every column.
4. The columns of A are linearly independent.

Again, something interesting happens when $T : \mathbb{R}^n \to \mathbb{R}^m$ is bijective. This forces the matrix to be square, giving even more equivalences for That One Theorem.

Theorem 2.60 (This One/ That One (Linear transformation cont.))

Let $A = [\vec{v}_1 \dots \vec{v}_n] \in M_{n\times n}(\mathbb{R})$, and $T : \mathbb{R}^n \to \mathbb{R}^n$ be the unique linear transformation satisfying $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Then the following are equivalent.

1. $\{\vec{v}_1, \dots, \vec{v}_n\}$ form a basis.

   $\vdots$

12. T is bijective.
13. T is surjective.
14. T is injective.

Exercises

2.5 Inverses and determinants

2.5.1 Inverse

Definition 2.61 (Inverse function)

Let $T : U \to V$ be bijective. The **inverse of** T is the unique function T^{-1} that satisfies

$$T^{-1}(T(x)) = x,$$
$$T(T^{-1}(y)) = y$$

for all $x \in U, y \in V$.

$$\begin{tikzcd} U \arrow[r, bend left, "T"] & V \arrow[l, bend left, "T^{-1}"] \end{tikzcd}$$

How do we know this inverse exists? We know that for each $y \in V$ there is exactly one $x \in U$ that satisifes $T(x) = y$, so we can set $T^{-1}(y)$ to be this particular value of x for each y. This definition of the inverse sounds tautological, so let us apply this to linear transformations.

Proposition 2.62

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a bijective linear transformation. Then T^{-1} is a linear transformation.

Proof. We want to show that for any $\vec{y}_1, \vec{y}_2 \in \mathbb{R}^n$, $c \in \mathbb{R}$, $T^{-1}(\vec{y}_1 + \vec{y}_2) = T^{-1}(\vec{y}_1) + T^{-1}(\vec{y}_2)$ and $T^{-1}(c\vec{y}_1) = cT^{-1}(\vec{y}_1)$.

The trick to solving this is to use the definition for inverses and switch back and forth with T and T^{-1}, and apply linearity of T.

$$\begin{aligned} T^{-1}(\vec{y}_1) + T^{-1}(\vec{y}_2) &= T^{-1}(T(T^{-1}(\vec{y}_1) + T^{-1}(\vec{y}_2))) \\ &= T^{-1}(T(T^{-1}(\vec{y}_1)) + T(T^{-1}(\vec{y}_2))) \text{ (as } T \text{ is linear)} \\ &= T^{-1}(\vec{y}_1 + \vec{y}_2). \end{aligned}$$

For scaling,

$$\begin{aligned} cT^{-1}(\vec{y}_1) &= T^{-1}(T(cT^{-1}(\vec{y}_1))) \\ &= T^{-1}(c(T(T^{-1}(\vec{y}_1)))) \\ &= T^{-1}(c\vec{y}_1). \end{aligned}$$

□

This means that the inverse of T has a matrix representation. If we write $T(\vec{x}) = A\vec{x}$ for some matrix A, then there is some matrix B such that $AB = BA = I_n$.

Definition 2.63 (Matrix Inverses)

Let $A \in M_{m\times n}(\mathbb{R})$. If there exists $X \in M_{n\times m}(\mathbb{R})$ such that

$$XA = I_n,$$

we call X the **left inverse** of A. Similarly, if there exists $Y \in M_{n\times m}(\mathbb{R})$ such that

$$AY = I_m,$$

we call Y the **right inverse** of A. If both the left and right inverses exist and are equal, then we say that A is **invertible**, and denote $X = Y$ the **inverse** of A, written as A^{-1}.

Remark. *If A has a left and right inverse, then the left and right inverses are equal. Indeed, we have for a left inverse X and a right inverse Y of A,*

$$X = X(I_m) = X(AY) = (XA)Y = (I_n)Y = Y.$$

As we have already seen, only bijective linear transformations have inverses, so that means we have more to add to That One Theorem. We also see that being invertible implies having a left and right inverse, so might expect that having a left or right inverse corresponds to the columns being linearly independent or span $\mathbb{R}^m$.

Example 2.64

$A \in M_{m\times n}(\mathbb{R})$ having a left/right inverse indeed corresponds to the conditions that the columns are linearly independent and the columns span $\mathbb{R}^m$. Figure out which conditions correspond to which.

There are two ways to think about this problem: one way is directly translate left and right inverses to the corresponding statements of the column vectors; the other way is to think of them in linear transformations and find out which one is surjective/injective. I will outline both methods.

Method 1: Without knowing where to go, let us suppose we want to prove A has linearly independent columns and see which extra hypothesis we need. Suppose

$$A\vec{x} = \vec{0}.$$

We want to show $\vec{x} = \vec{0}$. This means we want to cancel the A on the left, so we want to use the left inverse X.

$$\vec{x} = XA\vec{x} = X\vec{0} = \vec{0}.$$

So this means having a left inverse corresponds to linear independence. Now to prove the columns of A span $\mathbb{R}^m$, we try to solve

$$A\vec{x} = \vec{b}$$

for any $\vec{b} \in \mathbb{R}^m$. To produce a solution $\vec{x}$ that can cancel A from the right, we will need the right inverse Y, and set $\vec{x} = Y\vec{b}$, then

$$A\vec{x} = AY\vec{b} = \vec{b}.$$

Method 2: We translate the problem into a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ being injective (linear independence) and surjective (span $\mathbb{R}^m$). Why does a left inverse imply injectivity? Let us suppose $T(\vec{x}) = T(\vec{y})$, then the left inverse can immediately cancel out the T from the left and give us $\vec{x} = \vec{y}$.

For surjectivity, let us break down the columns of the right inverse $Y = \begin{bmatrix} \vec{y}_1 & \dots & \vec{y}_m \end{bmatrix}$. Then by the property of matrix multipication, we can consider each column to see

$$A\vec{y}_j = \vec{e}_j.$$

Since the basis vectors can build anything in $\mathbb{R}^m$, we can just use the corresponding linear combination of the $\vec{y}_j$'s to build anything in $\mathbb{R}^m$. This means we have three extra equivalences for a matrix being invertible.

Theorem 2.65 (This/That One)

Let $A = [\vec{v}_1 \dots \vec{v}_n] \in M_{n\times n}(\mathbb{R})$, and $T : \mathbb{R}^n \to \mathbb{R}^n$ be the unique linear transformation satisfying $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Then the following are equivalent.

1. $\{\vec{v}_1, \dots, \vec{v}_n\}$ form a basis.

   $\vdots$

15. A is invertible.
16. A as a left inverse.
17. A has a right inverse.

Finally, we want to compute this inverse A^{-1}. Let us set the columns

$$A = [\vec{a}_1 \ \vec{a}_2 \ \dots \ \vec{a}_n], \ A^{-1} = [\vec{b}_1 \ \vec{b}_2 \ \dots \ \vec{b}_n],$$

then the condition AA^{-1} translates to

$$\begin{aligned} A\vec{b}_1 &= \vec{e}_1, \\ A\vec{b}_2 &= \vec{e}_2, \\ &\vdots \\ A\vec{b}_n &= \vec{e}_n. \end{aligned}$$

Which means we just have to solve n different linear systems. We can do them all at once: by reducing the matrix

$$\left[\begin{array}{cccc|cccc} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n & \vec{b}_1 & \vec{b}_2 & \dots & \vec{b}_n \end{array}\right]$$

to rref, we will get

$$\left[\begin{array}{c|c} I_n & * \end{array}\right]$$

if A is invertible. The $*$ will be the solutions $\vec{b}_k$'s, which means $*$ is the inverse of A.

2.5.2 Isomorphisms

Definition 2.66 (Isomorphism)

Let $T : U \to V$ be a bijective linear transformation. We call T an **isomorphism** and the vector spaces U and V are **isomorphic**. We denote this $U \simeq V$.

Recall many sections ago, we said that many types of vector spaces are very similar to $\mathbb{R}^n$. Now that we have defined isomorphic vector spaces, we have a concrete statement of what it means to be similar.

Theorem 2.67 (Finite dimensional real vector spaces $\simeq \mathbb{R}^n$)

Let V be a finite dimensional real vector spaces. Then $V \simeq \mathbb{R}^n$ for some $n \in \mathbb{N}$.

Proof. Since V is finite dimensional, let $\{v_1, ..., v_k\}$ be a basis for V. We just need to construct an isomorphism from $\mathbb{R}^n \to V$. What is an easy way to define this? We just need to define where each of the standard basis vectors $\vec{e}_j$ goes. This gives us a very natural choice to guess that the isomorphism $T : \mathbb{R}^k \to V$ defined by

$$T(\vec{e}_j) = v_j,$$

or equivalently

$$T(\vec{x}) = x_1 v_1 + x_2 v_2 + \dots + x_k v_k$$

is bijective. This is indeed bijective, which you have confirmed in a previous exercise. □

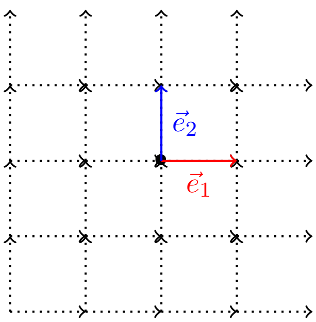
Corollary 2.68: Let V be finite dimensional. Then every basis of V has the same number of vectors.

Proof. Let $\{v_1, \dots, v_k\}$ and $w_1, \dots, w_l$ be two bases for V. Then by the construction in the previous theorem, we have $V \simeq \mathbb{R}^k$ and $V \simeq \mathbb{R}^l$. So $\mathbb{R}^k \simeq \mathbb{R}^l$ and thus $k = l$. □

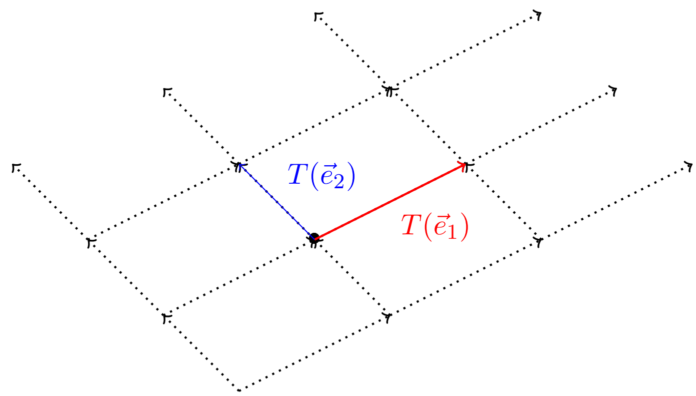
2.5.3 Determinants

We now move towards the last equivalence for That One Theorem. Recall in the intuition behind the Matrix Representation of Linear Transformations, we constructed a grid and showed that a linear transformation is a change in the coordinate system.

The original grid squares become grid 'parallelograms'. Indeed, if you extend to three dimensions,

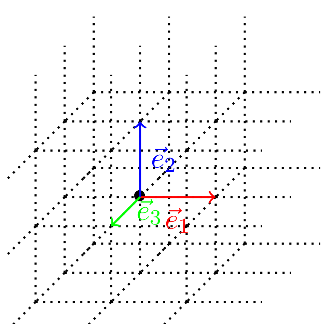


(a) Original grid

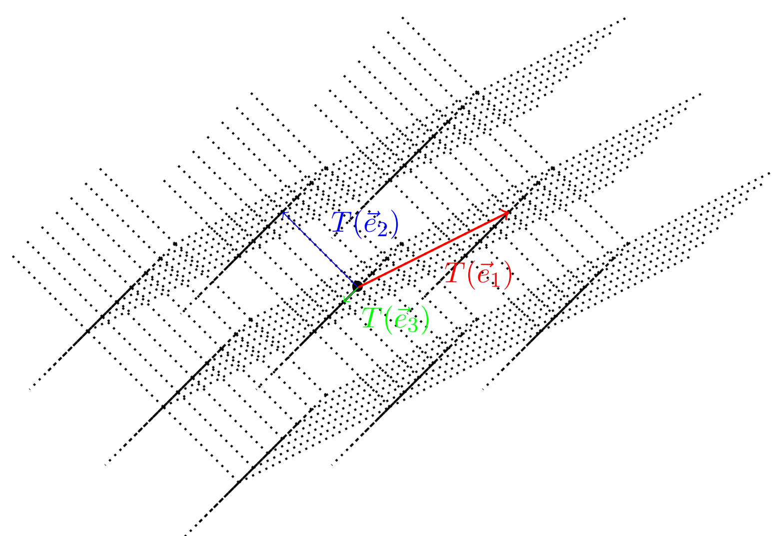


(b) Transformed grid

you can get grid 'parallelepipeds'. To say that the transformation is surjective means that the transformed unit square/cubes have non-zero area/volume and can thus tessellate.



(a) Original grid



(b) Transformed grid

Figure 2.4: I tried my best, please use a bit of imagination

Here we use the god-given formula for the determinant.

Definition 2.69 (Determinant)

We define a determinant function $\det : M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ as follows.

- $\det[a] = a$ for a 1×1 matrix $[a]$
- For higher dimensions $A = [a_{i,j}] \in M_{n\times n}(\mathbb{R})$, we recursively define

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \det A_{i,j}$$

for any $1 \le j \le n$. Any value of j will give the same formula upon expansion.

Remark. *$A_{i,j}$ is the the **minor** of the matrix A where you delete the i-th row and j-th column. If*

$$A = \begin{bmatrix} a & b & * \\ * & * & * \\ c & d & * \end{bmatrix},$$

$$A_{2,3} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Remark. *The recursive formula also works for*

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{j,i} \det A_{j,i}$$

where instead of expanding along the j-th row, we expand along the j-th column.

Proposition 2.70

For small matrices, we have these simplified formulae.

1×1 $\det([a]) = a$.

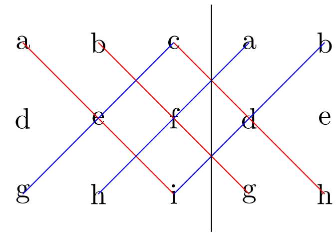
2×2 $\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc.$

3×3 $\det\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\right) = aei + bfg + cdh - ceg - afh - bdi.$

Remark. *The mnenomic for memorizing the 3×3 determinant is as follows: Extend the matrix as*

$$\left[\begin{array}{ccc|cc} a & b & c & a & b \\ d & e & f & d & e \\ g & h & i & g & h \end{array}\right]$$

and take the difference between the red diagonal products and the blue diagonal products.



As you have seen in a previous chapter, the formula for the determinant of a 3×3 matrix coincides with the volume of the parallelepiped spanned by the column vectors.

Example 2.71

Calculate the determinant of the rotational matrix

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

.

We just put this in the formula to get

$$\det R = \cos^2\theta - (-\sin^2\theta) = \cos^2\theta + \sin^2\theta = 1$$

regardles of the value of θ.

Example 2.72

Calculate the determinant of an upper-triangular matrix in the form

$$U = \begin{bmatrix} a_1 & * & * & \dots & * \\ 0 & a_2 & * & \dots & * \\ 0 & 0 & a_3 & \dots & * \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & a_n \end{bmatrix} \in M_{n\times n}(\mathbb{R}).$$

Let's consider the special case $n = 3$. We do not know the values of $*$, but if we write out

$$\left[\begin{array}{ccc|cc} a_1 & * & * & a_1 & * \\ 0 & a_2 & * & 0 & a_2 \\ 0 & 0 & a_3 & 0 & 0 \end{array}\right],$$

the only diagonal that can possible have all non-zero entries gives the product abc. So the determinant of a triangular matrix is always the product of the diagonal terms. The determinant of an upper triangular matrix is always the product of the diagonal terms. To show this, we need to exploit the recursive formula for the determinant. If we expand along the first column, most of the terms in the sum are 0, except the first term that gives

$$(-1)^{1+1}a_1 \det\left(\begin{bmatrix} a_2 & * & \dots & * \\ 0 & a_3 & \dots & * \\ \vdots & & \ddots & \\ 0 & 0 & \dots & a_n \end{bmatrix}\right),$$

which is a_1 times the determinant of a smaller upper triangular matrix. If we know that this $(n-1)\times(n-1)$ matrix has determinant $a_2a_3\dots a_n$, this will prove the formula for a $n \times n$ matrix, then this result will prove the formula for a $(n+1)\times(n+1)$ matrix and so on! This technique is known as Mathematical Induction. If you want to prove a statement is true for all natural numbers n, you can show that it is true when $n = 1$, and show that if the statement is true for $n = k$, it is also true for $n = k+1$.

Proof. We show that the determinant of U is $(a_1 \dots a_n)$ by induction on the value of n.

Base Case: The determinant of the matrix $[a_1]$ is a_1, which is evidently the product of all the diagonal terms.

Induction step: Suppose the determinant of a $k \times k$ upper triangular matrix is the product of the diagonal terms. We compute the determinant of a $(k+1)\times(k+1)$ matrix in the form

$$U = [u_{i,j}] = \begin{bmatrix} a_1 & * & * & \dots & * \\ 0 & a_2 & * & \dots & * \\ 0 & 0 & a_3 & \dots & * \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & a_{k+1} \end{bmatrix}.$$

Applying the recursive formula and expanding along the first column, the only non-zero $u_{j,1}$ term is $u_{1,1} = a_1$, so

$$\begin{aligned} \det U &= (-1)^{1+1}a_1 \det\left(\begin{bmatrix} a_2 & * & \dots & * \\ 0 & a_3 & \dots & * \\ \vdots & & \ddots & \\ 0 & 0 & \dots & a_{k+1} \end{bmatrix}\right) \\ &= a_1(a_2a_3\dots a_{k+1}) \end{aligned}$$

by applying the hypothesis on the $k \times k$ upper triangular matrix. Then the determinant of a $(k+1)\times(k+1)$ matrix is the product of the diagonal terms. By the Principle of Mathematical Induction, the determinant of an upper triangular matrix of any size is the product of its diagonal terms. □

Proposition 2.73

The elementary row operations change the determinant of a matrix in the following way:

- *(Row Swap)* The determinant is multiplied by -1.
- *(Scaling)* The determinant is multiplied by the scaling constant.
- *(Sum)* The determinant does not change.

The proof of this is not very enlightening, so we will skip it. It boils down to working with the general determinant formula and doing intense algebra manipulation.

However, we have the following corollary

Corollary 2.74: The determinant of a matrix can be computed as follows:

1. Reduce the matrix to rref (or row echelon form).
2. Take the product of the diagonal terms.
3. Multiply this product by $1/c$ for every time you scaled a row by c in computing the rref.
4. Multiply this product by -1 for every time you swaped two rows in computing the rref.
5. The final product is the determinant of the matrix.

Theorem 2.75 (This/That One (Determinants))

Let $A = [\vec{v}_1 \dots \vec{v}_n] \in M_{n\times n}(\mathbb{R})$, and $T : \mathbb{R}^n \to \mathbb{R}^n$ be the unique linear transformation satisfying $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. Then the following are equivalent.

1. $\{\vec{v}_1, \dots, \vec{v}_n\}$ form a basis.

   $\vdots$

18. $\det A \neq 0$.

Proof. Recall the motivation for introducing the determinant is the signed n-dimensional volume of the paralellotope formed by the column vectors. If (and only if) the volume is non-zero, they can tesselate the whole $\mathbb{R}^n$ space.

To show the equivalence of $\det A \neq 0$, we can use the condition from That One Theorem that $\text{rref}(A) = I_n$, and the previous corollary for computing determinants. If the determinant is 0, then the product of the diagonal terms of $\text{rref}A$ has a zero, so $\text{rref}(A) \neq I_n$. On the converse, if the determinant is non-zero, all the diagonal terms in $\text{rref}(A)$ are non-zero. The only way this can happen for a matrix in rref is when the matrix is also the identity matrix. □