Linear Algebra

Divakaran D

Contents

Chapter 1

Solving two equations in two unknowns

In school, you would have encountered questions of the form: Solve

$$x + 2y = 0$$
$$2x + y = 3.$$

You might not have realised, but the question is a little ambiguous. It was implicit that you need to find a tuple (x, y) that satisfies both equations. Not a tuple (x, y, z) or (x, y, z, w). If I had instead asked you to solve

$$x + 2y = 0$$
$$2x + y = 3$$
$$x + z = 0$$

you would have assumed you need to find a triple (x, y, z). In other words, the context informs you what you are looking for. However, this can be at times confusing. In this course, we would try to avoid such ambiguities as much as possible. Don't hesitate to point out if/when I turn sloppy. And, I promise, I would hold high expectations from you. Thus, I would rephrase the earlier question as: Find all (x, y) such that $x, y \in \mathbb{R}$ and

$$x + 2y = 0$$
$$2x + y = 3.$$

Definition 1.1. The collection of all tuples (x, y) where x and y are real numbers - $\{(x, y)|x, y \in \mathbb{R}\}$ - is called the Cartesian plane and is denoted as $\mathbb{R}^2$

Thus, the question is asking us to find the set $\{(x, y) \in \mathbb{R}^2 | x + 2y = 0 \text{ and } 2x + y = 3\}$ or equivalently the set $\{(x, y) \in \mathbb{R}^2 | x + 2y = 0\} \cap \{(x, y) \in \mathbb{R}^2 | 2x + y = 3\}$. You may recall from school or you can check using GeoGebra that the set $\{(x, y) \in \mathbb{R}^2 | x + 2y = 0\}$ and $\{(x, y) \in \mathbb{R}^2 | 2x + y = 3\}$ represent lines. In Appendix A we define lines to be sets of this form, but intuition would be sufficient in the course. Thus, **we were looking for the intersection of two lines!**

1.1. Solving equations amount to finding the intersection of lines

More generally, the set of all $(x, y) \in \mathbb{R}^2$ such that

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \end{aligned} \tag{1.1}$$

represents the intersection of two lines. The act of finding all points in the set is called solving the system of equations 1.1. Notice that an equation $px + qy = r$ is really silly when both p and q are 0. The set $\{(x, y) \mid 0 = r\}$ is $\emptyset$ when $r \neq 0$ and is $\mathbb{R}^2$ when $r = 0$. Thus, in the discussion that follows, we would assume that the linear equations in our system are not of this form. More precisely,

Assumption

Given a system of linear equations 1.1, we assume

$$[(a \neq 0) \vee (b \neq 0)] \wedge [(c \neq 0) \vee (d \neq 0)]$$

We will explore how to solve such a system of equations. Note that if $(x, y) \in \mathbb{R}^2$ satisfy the system of equations 1.1, then it should also satisfy the system

$$\begin{aligned} adx + bdy &= de \\ bcx + bdy &= bf \end{aligned} \tag{1.2}$$

In other words, a tuple $(x, y) \in \mathbb{R}^2$ satisfies the system 1.1 only if x should satisfy the equation

$$(ad - bc)x = de - bf \tag{1.3}$$

Case 1: $ad - bc \neq 0$ **-** If $ad - bc \neq 0$, then a tuple (x, y) satisfies the system 1.1 only if x satisfies Equation 1.3. But, x satisfies Equation 1.3 only if $x = \frac{de-bf}{ad-bc}$. Substituting the value back in the system of equations 1.1 we get

$$\begin{aligned} \left[a\left(\frac{de-bf}{ad-bc}\right) + by = e\right] &\equiv \left[y = \frac{ade - bce - ade + abf}{abd - b^2c} = \frac{abf - bce}{abd - b^2c} = \frac{af - ce}{ad - bc}\right] \\ \left[c\left(\frac{de-bf}{ad-bc}\right) + dy = f\right] &\equiv \left[y = \frac{adf - bcf - cde + bcf}{ad^2 - bcd} = \frac{adf - cde}{ad^2 - bcd} = \frac{af - ce}{ad - bc}\right] \end{aligned} \tag{1.4}$$

Thus, the set of solutions of the system of equations 1.1 is

$$\left\{\left(\frac{de-bf}{ad-bc}, \frac{af-ce}{ad-bc}\right)\right\}$$

Example 1.2. Find all $(x, y) \in \mathbb{R}^2$ such that

$$\begin{aligned} x + 2y &= 0 \\ 2x + y &= 3. \end{aligned}$$

As $ad - bc = 1 - 4 = -3 \neq 0$, the set of solutions is $\left\{\left(\frac{de-bf}{ad-bc}, \frac{af-ce}{ad-bc}\right)\right\} = \left\{\left(\frac{0-6}{-3}, \frac{3-0}{-3}\right)\right\} = \{(2, -1)\}$. Figure 1 shows the GeoGebra solution. Note that it agrees with what we have obtained.

Case 2: $ad - bc = 0$ **and** $de - bf \neq 0$ **-** If the system 1.1 has a solution, then there should be an x that satisfies Equation 1.3. But, when $ad - bc = 0$ and $de - bf \neq 0$, Equation 1.3 takes the following form

$$0 = de - bf$$

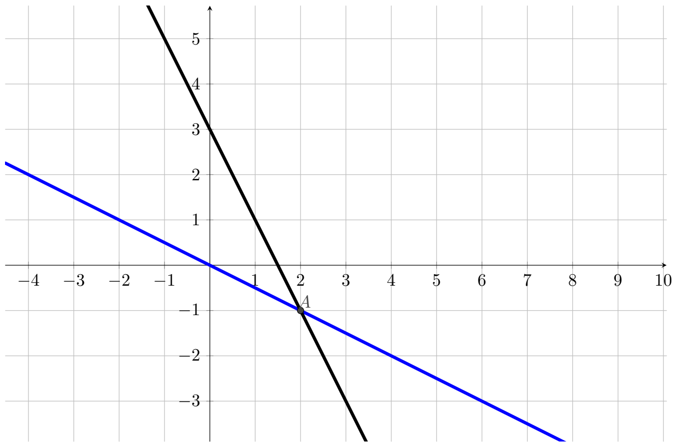


Figure 1. GeoGebra image for the intersection of the two lines $\{(x, y) \in \mathbb{R}^2 \mid x + 2y = 0\}$ and $\{(x, y) \in \mathbb{R}^2 \mid 2x + y = 3\}$. Notice that $ad - bc \neq 0$ and there is a unique solution.

and clearly it has no solution. Thus, there are no solutions for the system 1.1. Geometrically, this is a case when the two lines are parallel.

Example 1.3. Find all $(x, y) \in \mathbb{R}^2$ such that

$$\begin{aligned} x + 2y &= 0 \\ x + 2y &= 3. \end{aligned}$$

As $ad - bc = 2 - 2 = 0$ and $de - bf = 0 - 6 = -6 \neq 0$, there are no solutions. Figure 2 shows the GeoGebra solution. Note that it agrees with what we have obtained.

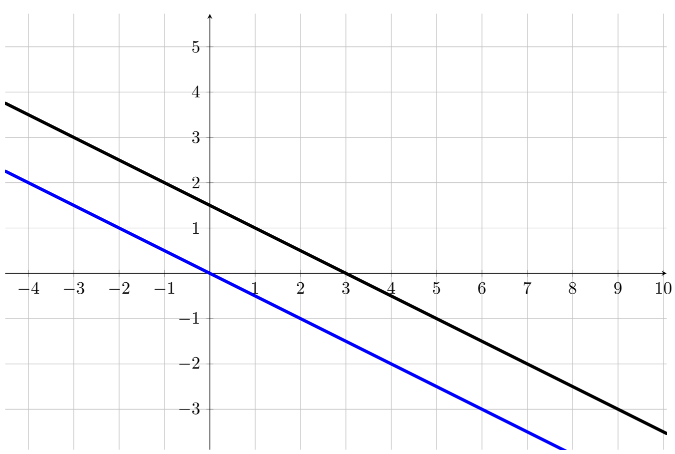


Figure 2. GeoGebra image for the intersection of the two lines $\{(x, y) \in \mathbb{R}^2 \mid x + 2y = 0\}$ and $\{(x, y) \in \mathbb{R}^2 \mid x + 2y = 3\}$. Notice that $ad - bc = 0 \neq de - bf$ and there is no solution.

Case 3: $ad - bc = 0$**,** $de - bf = 0$ **-** If $d \neq 0$ and $b = 0$, then we have $ad = bc = 0$. As, $d \neq 0$, this implies $a = 0$. But, we assumed a and b are not simultaneously 0. Similarly, if $b \neq 0$ and $d = 0$, then we have $bc = ad = 0$. As, $b \neq 0$, this implies $c = 0$. BUt, we assumed c and d are not simultaneously 0. Thus, there are only two subcases:

Case 3a: $ad - bc = 0$**,** $de - bf = 0$**,** $d \neq 0$ **and** $b \neq 0$ **-** Then, $a = \frac{bc}{d}$. We may rewrite Equation 1.1 as

$$\left[\left(\frac{bc}{d}\right)x + by = e\right] \equiv [bcx + bdy = de = bf] \equiv [cx + dy = f]$$
$$cx + dy = f$$

Thus, "in-effect" we have only one equation - $cx + dy = f$ - in this case. We know that the set of all (x, y) satisfying the equation forms a line. This is a case when the two lines coincide.

Example 1.4. Find all $(x, y) \in \mathbb{R}^2$ such that

$$x + 2y = 3$$
$$2x + 4y = 6.$$

Note that $ad - bc = 4 - 4 = 0$, $de - bf = 12 - 12 = 0$, $d = 4 \neq 0$, and $b = 2 \neq 0$. Both equations in the system represent the same line.

Case 3b: $b = 0 = d$ **-** We may rewrite Equation 1.1 as

$$ax = e$$
$$cx = f.$$

Note that

$$\{(x, y) \mid ax = e\} = \left\{\left(\frac{e}{a}, y\right) \mid y \in \mathbb{R}\right\}.$$

Similarly,

$$\{(x, y) \mid cx = f\} = \left\{\left(\frac{f}{c}, y\right) \mid y \in \mathbb{R}\right\}.$$

Thus,

$$\{(x, y) \mid ax = e\} \cap \{(x, y) \mid cx = f\} = \begin{cases} \left\{\left(\frac{f}{c}, y\right) \mid y \in \mathbb{R}\right\} & \text{if } c \neq 0,\ a \neq 0, \text{ and } \frac{e}{a} = \frac{f}{c} \\ \emptyset & \text{otherwise} \end{cases}.$$

Example 1.5. Find all $(x, y) \in \mathbb{R}^2$ such that

$$2x = 4$$
$$x = 3$$

As $c \neq 0$, $a \neq 0$, and $\frac{e}{a} \neq \frac{f}{c}$ this system does not have a solution. Figure 3 shows the GeoGebra solution. Note that it agrees with what we have obtained.

Example 1.6. Find all $(x, y) \in \mathbb{R}^2$ such that

$$2x = 4$$
$$x = 2$$

As $c \neq 0$, $a \neq 0$, and $\frac{e}{a} = \frac{f}{c}$ the set of solutions is $\{(2, y) \mid y \in \mathbb{R}\}$.

Remark 1.7. Notice that the solution exists and is unique only when $ad - bc \neq 0$. Thus, the number $ad - bc$ is important and we will encounter it and its relatives throughout the course.

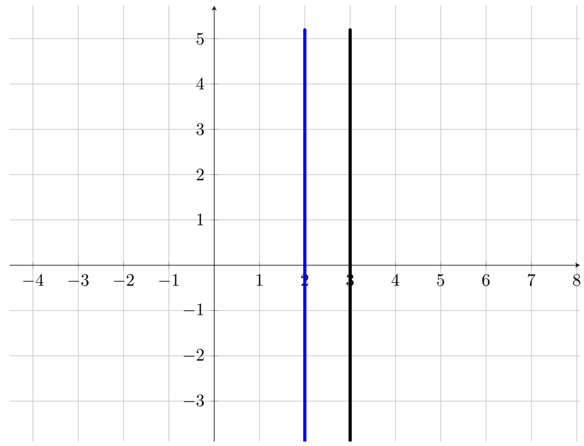


Figure 3. GeoGebra image for the intersection of the two lines $\{(x, y) \in \mathbb{R}^2 \mid 2x = 4\}$ and $\{(x, y) \in \mathbb{R}^2 \mid x = 3\}$. Notice that the two lines are parallel and there is no solution.

Exercise 1.8. Consider the system

$$\begin{aligned} x + ay &= b \\ 2x + 10y &= 4 \end{aligned}$$

Give a value of a for which the system has a unique solution. Does the value of b matter? Explain your answer. Give a value of (a, b) for which the system has no solution. Give a value of (a, b) for which the system has infinitely many values.

1.2. Cartesian plane or $\mathbb{R}^2$

Let us now recall the various structures on $\mathbb{R}^2$ you have encountered in school.

Definition 1.9. For two points (a, b) and (c, d) in $\mathbb{R}^2$, we say $(a, b) = (c, d)$ if $a = c$ and $b = d$.

Addition. Given two elements on $\mathbb{R}^2$, say (a, b) and (c, d), we can add these elements to form an element $(a + c, b + d)$. This defines a function $+ : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ defined as $+((a, b),(c, d)) = (a + c, b + d)$. We will typically write $(a, b) + (c, d)$ instead of $+((a, b),(c, d))$ - as you have been doing for addition of numbers. This is called **infix notation**.

Example 1.10. Let $v = (1, 2)$ and $w = (3, 4)$. Then $v + w = (1, 2) + (3, 4) = (1 + 3, 2 + 4) = (4, 6)$.

Note on notation

Observe, I am using a single letter v to represent the element $(1, 2)$. Later in the book, I will probably write $w = (1, 2, 3)$. You need to keep track of the "type"! As $v, w \in \mathbb{R}^2$, the $+$ sign denotes addition on $\mathbb{R}^2$. But, if $x = 1$ and $y = 2$, then the $+$ sign in $x + y$ would denote addition in $\mathbb{R}$. Most importantly, the expression $x + v$ makes no sense as x and v are objects of different types. Your programming experience would come in handy as there too you had to always keep track of the types.

Elements of $\mathbb{R}^2$ are called vectors - and we will use the language often. This the reason we use v and neighbouring letters of the alphabet for elements in $\mathbb{R}^2$.

Theorem 1.11 (Existence of additive identity)**.** *There exists an element (x, y) in $\mathbb{R}^2$ such that $\forall (a, b) \in \mathbb{R}^2$, $(a, b) + (x, y) = (a, b) = (x, y) + (a, b)$.*

Proof Strategy

In the proof below we will use two strategies that help us prove statements with quantifiers. There are two quantifiers you would encounter, namely the existential quantifies (there exists/$\exists$) and the universal quantifier (for all/ $\forall$).

A statement with an existential quantifier can be proved by giving one example. Note that the statement of the above theorem is an existential statement. We prove the statement by giving one example. That is, we show that $(0,0)$ is the required element. In other words, we show: $\forall (a,b) \in \mathbb{R}^2$, $(a,b)+(0,0) = (a,b) = (0,0)+(a,b)$.

The statement $\forall (a,b) \in \mathbb{R}^2$, $(a,b)+(0,0) = (a,b) = (0,0)+(a,b)$ is a statement with a universal quantifier. To prove something is a true for all elements, we show it is true for an **arbitrary but fixed element**. This is also why mathematicians often use the terms “for all”, “for any”, “for each”, “given any” etc interchangeably. All of these terms are denoted by the symbol $\forall$.

Fix an arbitrary element $(a,b) \in \mathbb{R}^2$, then we need to prove $(a,b)+(0,0) = (a,b) = (0,0)+(a,b)$. The statement $(a,b)+(0,0) = (a,b) = (0,0)+(a,b)$ is actually a compound sentence. It is a short hand for $(a,b)+(0,0) = (a,b)$ and $(a,b) = (0,0)+(a,b)$. Thus, we need to prove both the statements $(a,b)+(0,0) = (a,b)$ and $(a,b) = (0,0)+(a,b)$. But, they are both easy to prove.

$$\begin{aligned} (a,b)+(0,0) &= (a+0,b+0) && \text{(definition of addition on } \mathbb{R}^2\text{)} \\ &= (a,b) && \text{(0 is the additive identity in } \mathbb{R}\text{)} \end{aligned}$$

Similarly,

$$\begin{aligned} (0,0)+(a,b) &= (0+a,0+b) && \boxed{} \\ &= (a,b) && \boxed{} \end{aligned}$$

Theorem 1.12. *Addition on $\mathbb{R}^2$ is commutative. More precisely,* $\forall (a,b),(c,d) \in \mathbb{R}^2$, $(a,b)+(c,d) = (c,d)+(a,b)$.

Proof. Fix an arbitrary pair (a,b) and (c,d) in $\mathbb{R}^2$. We will show that $(a,b)+(c,d) = (c,d)+(a,b)$. As the elements were arbitrary, this would prove the universal statement.

$$\begin{aligned} (a,b)+(c,d) &= (a+c,b+d) && \boxed{} \\ &= (c+a,b+d) && \text{(from commutativity of addition on } \mathbb{R}\text{)} \end{aligned}$$

□

Theorem 1.13. *Addition on $\mathbb{R}^2$ is associative. More precisely,* $\forall (a,b),(c,d),(e,f) \in \mathbb{R}^2$, $(a,b)+((c,d)+(e,f)) = ((a,b)+(c,d))+(e,f)$.

Proof. Fix arbitrary elements $(a,b),(c,d),(e,f) \in \mathbb{R}^2$. We will show that $(a,b)+((c,d)+(e,f)) = ((a,b)+(c,d))+(e,f)$. As the elements were arbitrary, this would prove the universal statement.

$$\begin{aligned} (a,b)+((c,d)+(e,f)) &= (a,b)+(c+e,d+f) && \boxed{} \\ &= (a+(c+e),b+(d+f)) && \boxed{} \\ &= ((a+c)+e,(b+d)+f) && \text{(associativity of addition in } \mathbb{R}\text{)} \\ &= (a+c,b+d)+(e,f) && \boxed{} \\ &= ((a,b)+(c,d))+(e,f) && \boxed{} \end{aligned}$$

$\square$

Exercise 1.14. Show that given any element (a, b), $(a, b) + (-a, -b) = (0, 0) = (-a, -b) + (a, b)$. The element $(-a, -b)$ is called the additive inverse of (a, b).

Exercise 1.15. Plot $(1, 2)$ and $(-1, -2)$ in GeoGebra. Similarly plot $(3, 4)$ and $(-3, -4)$ in GeoGebra.

This allows us to define subtraction on $\mathbb{R}^2$. Given two elements (a, b) and (c, d) we define $(a, b) - (c, d) = (a, b) + (-c, -d)$.

Example 1.16. Let $v = (1, 2)$ and $w = (3, 4)$. Then, $v - w = (1, 2) - (3, 4) = (-2, -2)$.

Exercise 1.17. Let $v = (5, 3)$ and $w = (2, 7)$. Then find $v - w$.

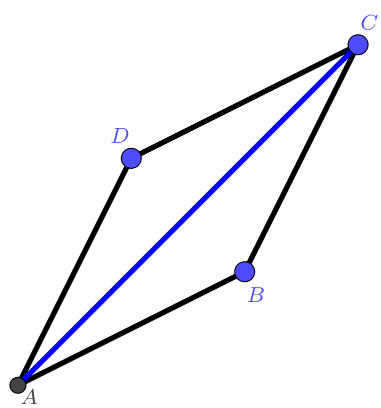
Definition 1.18 (Euclidean Distance)**.** The Euclidean distance is a function $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined as $d((x_1, x_2), (y_1, y_2)) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$.

Example 1.19. Let $v = (1, 2)$ and $w = (3, 4)$. Then $d(v, w) = d((1, 2), (3, 4)) = \sqrt{(1-3)^2 + (2-4)^2} = \sqrt{8} = 2\sqrt{2}$.

Theorem 1.20. *Given any two points $A = (a_1, a_2)$ and $B = (b_1, b_2)$, the points $O = (0, 0)$, $A = (a_1, a_2)$, $B = (b_1, b_2)$, and $C = (a_1 + b_1, a_2 + b_2)$ form vertices of a parallelogram. This allows us to geometrically understand addition on $\mathbb{R}^2$.*

Proof. We would use the following lemma to complete our proof

Lemma 1.21. *Given any quadrilateral, if its opposite sides are equal, then it is a paralellogram.*



Proof. Let $ABCD$ be an arbitrary quadrilateral such that $|AB| = |CD|$ and $|AD| = |BC|$. Then, $\triangle ABC \cong \triangle CDA$ by SSS criterion. Therefore $\angle ACD = \angle CAB$ and $\angle CAD = \angle ACB$. Hence, by the converse of the alternate angle theorem, we get that DC is parallel to AB and AD is parallel to BC. Hence, the quadrilateral $ABCD$ is a parallelogram. $\square$

We will use the above Lemma to show that $OACB$ is a paralellogram. Note that

$$|OA| = d(O, A) = \sqrt{a_1^2 + a_2^2} = \sqrt{((a_1 + b_1) - b_1)^2 - ((a_2 + b_2) - b_2)^2} = d(C, B) = |BC|.$$

Similarly,

$$|OB| = d(O, B) = \sqrt{b_1^2 + b_2^2} = \sqrt{((a_1 + b_1) - a_1)^2 - ((a_2 + b_2) - a_2)^2} = d(C, A) = |AC|.$$

$\square$

1.2.1. Scalar Multiplication. Given a real number α and an element (a,b) in $\mathbb{R}^2$, the point $(\alpha a, \alpha b)$ lies on the line joining $(0,0)$ and (a,b). Thus, we view this "multiplication" as scaling. This defines a function $\cdot : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$.

Example 1.22. Let $k = 3$ and $v = (1,2)$. Then $kv = 3(1,2) = (3,6)$.

GeoGebra Exercise

I have created a GeoGebra file where I use the function Sequence(Expression, Variable, Start Value, End Value, Increment) to plot $a(1,2)$ as a varies from -5 to 5 with an increment of $\frac{1}{N}$. The value of N is determined by a slider.

Exercise 1.23. In this exercise, you will play with the shared GeoGebra code to build intuition about scalar multiplication and learn GeoGebra.

(1) Modify the GeoGebra code to plot all scalar multiples of three other vectors of your choice. What do you observe?
(2) Use the Sequence command to plot the graph of $y = x^2$.

Exercise 1.24. Show that

(1) $\forall (a,b) \in \mathbb{R}^2 \; 1 \cdot (a,b) = (a,b)$
(2) $\forall (a,b) \in \mathbb{R}^2 \; (a,b) + (-1)\cdot(a,b) = (0,0)$
(3) $\forall \; x, y \in \mathbb{R}$ and $(a,b) \in \mathbb{R}^2$, $(xy)\cdot(a,b) = x \cdot (y \cdot (a,b))$
(4) $\forall x \in \mathbb{R}$ and $(a,b),(c,d) \in \mathbb{R}^2$, $x \cdot [(a,b) + (c,d)] = x \cdot (a,b) + x \cdot (c,d)$
(5) $\forall \; x, y \in \mathbb{R}$ and $(a,b) \in \mathbb{R}^2$, $(x+y)\cdot(a,b) = x \cdot (a,b) + y \cdot (a,b)$

1.2.2. Inner Product. You may have studied this as the dot product. This is a function $\langle \;,\; \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined as $\langle (a,b),(c,d)\rangle = ac + bd$.

Example 1.25. Let $v = (1,2)$ and $w = (3,4)$. Then $\langle v, w\rangle = \langle (1,2),(3,4)\rangle = 1.3 + 2.4 = 3 + 8 = 11$.

Exercise 1.26. Show that

(1) $\forall (a,b),(c,d) \in \mathbb{R}^2$, $\langle (a,b),(c,d)\rangle = \langle (c,d),(a,b)\rangle$
(2) $\forall (a,b),(c_1,d_1),(c_2,d_2) \in \mathbb{R}^2$, $\langle (a,b),(c_1,d_1)+(c_2,d_2)\rangle = \langle (a,b),(c_1,d_1)\rangle + \langle (a,b),(c_2,d_2)\rangle$
(3) $\forall x \in \mathbb{R}$ and $\forall (a,b),(c,d) \in \mathbb{R}^2$, $\langle x(a,b),(c,d)\rangle = x\langle (a,b),(c,d)\rangle = \langle (a,b), x(c,d)\rangle$
(4) $\forall (a,b) \in \mathbb{R}^2$, $\langle (a,b),(a,b)\rangle \geq 0$ and the equality holds **if and only if (iff)** $(a,b) = (0,0)$

Exercise 1.27. Find the following inner products:

(1) $\langle (1,2),(3,4)\rangle$
(2) $\langle (4,8),(-5,3)\rangle$
(3) $\langle (1,0),(0,1)\rangle$
(4) $\langle (1,0),(1,0)\rangle$
(5) $\langle (1,1),(1,-1)\rangle$
(6) $\langle (\sqrt{2},\pi),(\frac{1}{2},0)\rangle$

1.2.3. Norm. is a function $\| \; \| : \mathbb{R}^2 \to \mathbb{R}$ defined as $\|(a,b)\| = \langle (a,b),(a,b)\rangle^{\frac{1}{2}} = \sqrt{a^2+b^2}$.

Exercise 1.28. Find the norm of $(2,3)$. Give an example of another element of $\mathbb{R}^2$ with the same norm.

Exercise 1.29. Show that $\|x(a,b)\| = |x|\|(a,b)\|$.

Theorem 1.30. *For all pairs of points (a, b) and (c, d) in $\mathbb{R}^2$*

$$d((a, b), (c, d)) = \sqrt{(a - c)^2 + (b - d)^2} = \|(a - c, b - d)\| = \|(a, b) - (c, d)\|.$$

Exercise 1.31. Find the norm of the following vectors

(1) $(1, 0)$
(2) $(0, 1)$
(3) $(1, 1)$
(4) $(\sqrt{2}, \sqrt{3})$
(5) $(\pi, 0)$
(6) $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$

1.2.4. Polar coordinates. Given any point $(x, y) \in \mathbb{R}^2$, note that $\frac{x}{\sqrt{x^2+y^2}}$ and $\frac{y}{\sqrt{x^2+y^2}}$ are two numbers such that

(1) $\left|\frac{x}{\sqrt{x^2+y^2}}\right| \leq 1 \geq \left|\frac{y}{\sqrt{x^2+y^2}}\right|$
(2) $\left(\frac{x}{\sqrt{x^2+y^2}}\right)^2 + \left(\frac{y}{\sqrt{x^2+y^2}}\right)^2 = 1$

Thus, you can find a θ such that $\cos(\theta) = \frac{x}{\sqrt{x^2+y^2}}$ and $\sin(\theta) = \frac{y}{\sqrt{x^2+y^2}}$. Moreover, it is easy to see that θ is the angle the line joining origin and (x, y) makes with the x-axis. So, if $r = \sqrt{x^2 + y^2}$, then $(x, y) = r(\cos(\theta), \sin(\theta))$. This representation is called the polar representation.

1.3. Linear combinations, span, and solutions to system of linear equations

We already had a through discussion on the solutions of a system of linear equations. In this section, we would take a different perspective which would often come handy. Notice that $(x, y) \in \mathbb{R}^2$ satisfy the system

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \end{aligned}$$

iff

$$x(a, c) + y(b, d) = (e, f).$$

Definition 1.32. Given two elements (v_1, v_2) and (w_1, w_2), in $\mathbb{R}^2$, an expression of the form $\alpha(v_1, v_2) + \beta(w_1, w_2)$ is called a linear combination of the (v_1, v_2) and (w_1, w_2). The collection of all such linear combinations is called the span of (v_1, v_2) and (w_1, w_2).

$$\text{Span}((v_1, v_2), (w_1, w_2)) = \{\alpha(v_1, v_2) + \beta(w_1, w_2) \ |\alpha, \beta \in \mathbb{R}\}.$$

Theorem 1.33. *The system 1.1 has a solution iff $(e, f) \in$ Span$((a, c), (b, d))$.*

We saw in the previous section that if $ad - bc \neq 0$, then $\forall (e, f) \in \mathbb{R}^2$, the system 1.1 has a solution. In other words,

Theorem 1.34. *Let $(a, c), (b, d) \in \mathbb{R}^2$. If $ad - bc \neq 0$, then Span$((a, c), (b, d)) = \mathbb{R}^2$.*

GeoGebra Exercise

I have created a GeoGebra file that plots $aA+bB$. The possible values of a and b are determined by the two sliders c and N. Play with it and build an intuition on the span of the two vectors.

Exercise 1.35. In this exercise, you will play with the shared GeoGebra code to build intuition about scalar multiplication and learn GeoGebra.

(1) For a fixed value of c and N, what values can a take? Describe it in terms of c and N. Let $c = 2$ and $N = 1$, what are all the values that (a, b) can take?
(2) Modify the GeoGebra code try out the span of other pairs of vectors - try it out for atleast two pairs of linearly dependent vectors and two pairs of linearly independent vectors.

Exercise 1.36. Find the span of the following pairs of vectors:

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

Theorem 1.37. *For all* $(a, c), (b, d) \in \mathbb{R}^2$, $ad - bc = 0$ *iff there exists* $(0, 0) \neq (\alpha, \beta) \in \mathbb{R}^2$ *such that* $\alpha(a, c) + \beta(b, d) = (0, 0))$.

Proof Strategy

A statement of the form "p if and only if (iff) q" is a compound statement. It is a short hand for

(1) If p, then q.
(2) If q, then p.

Remember to prove both implications when you encounter an iff. To prove "If p, then q" we assume p and prove q.

Proof. Assume $ad-bc = 0$. Then, $d(a, c)+(-c)(b, d) = (ad-bc, cd-cd) = (0, 0)$. Thus, there exists $(\alpha, \beta) = (d, -c)$ such that $\alpha(a, c) + \beta(b, d) = 0$. $(d, -c) \neq (0, 0)$ as we have assumed $(c, d) \neq (0, 0)$.

Assume there exists $(0, 0) \neq (\alpha, \beta) \in \mathbb{R}^2$ such that $\alpha(a, c) + \beta(b, d) = (0, 0)$. If $\alpha \neq 0$, then $(a, c) = \frac{-\beta}{\alpha}(b, d) = \left(\frac{-\beta}{\alpha}b, \frac{-\beta}{\alpha}d\right)$. If $(a, c) = \left(\frac{-\beta}{\alpha}b, \frac{-\beta}{\alpha}d\right)$, then $ad - bc = \left(\frac{-\beta}{\alpha}b\right)d - b\left(\frac{-\beta}{\alpha}d\right) = 0$. On the other hand, if $\beta \neq 0$, then $(b, d) = \frac{-\alpha}{\beta}(a, c) = \left(\frac{-\alpha}{\beta}a, \frac{-\alpha}{\beta}c\right)$. If $(a, c) = \left(\frac{-\beta}{\alpha}b, \frac{-\beta}{\alpha}d\right)$, then $ad-bc = \left(\frac{-\beta}{\alpha}b\right)d-b\left(\frac{-\beta}{\alpha}d\right) = 0$. If $(b, d) = \left(\frac{-\alpha}{\beta}a, \frac{-\alpha}{\beta}c\right)$, then $ad-bc = a\left(\frac{-\alpha}{\beta}c\right)-\left(\frac{-\alpha}{\beta}a\right)c = 0$. ☐

Definition 1.38. We say $v, w \in \mathbb{R}^2$ are linearly dependent if $\exists(\alpha, \beta) \in \mathbb{R}^2$ such that $(\alpha, \beta) \neq (0, 0)$ and $\alpha v + \beta w = 0$.

Theorem 1.39. *Two vectors* (a, c) *and* (b, d) *are linearly dependent if and only if (iff)* $ad - bc = 0$

We say two vectors v, w are linearly independent if they are not linearly dependent. More precisely,

Definition 1.40. We say $v, w \in \mathbb{R}^2$ are linearly independent if $\forall(\alpha, \beta) \in \mathbb{R}^2$, $\alpha v + \beta w \neq 0$ or $(\alpha, \beta) = (0, 0)$. Equivalently, we say two vectors v and w are linearly independent if $\forall(\alpha, \beta) \in \mathbb{R}^2$, $\alpha v + \beta w = 0 \implies (\alpha, \beta) = (0, 0)$.

Logic refresher

The equivalence of the two different definitions given above is due to the logical equivalence of $p \implies q$ (which, by the way, is another way of saying "if p then q") and $(\neg p) \vee q$. The equivalence can be checked with the help of a truth table.

p	q	$\neg p$	$p \implies q$	$(\neg p) \vee q$
T	T	F	T	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

Exercise 1.41. Check if the following pairs of vectors are linearly independent or linearly dependent. Give rigorous arguments to justify your claims.

(1) $(2, 3), (0, 0)$

(2) $(2, 3), (4, 6)$

(3) $(1, 0), (0, 1)$

(4) $(1, 1), (1, -1)$

Theorem 1.42. *Two vectors (a, c) and (b, d) are linearly independent if and only if (iff) $ad-bc \neq 0$.*

Thus, we may rewrite Theorem 1.34 as

Theorem 1.43. *For all $v, w \in \mathbb{R}^2$, if v, w are linearly independent then Span$(v, w) = \mathbb{R}^2$.*

It is now natural to ask, what happens when v, w are linearly dependent. We will answer the question after giving a geometric characterisation of dependence.

Theorem 1.44. *There exists $(\alpha, \beta) \in \mathbb{R}^2$ such that $(\alpha, \beta) \neq (0, 0)$ and $\alpha v + \beta w = (0, 0)$[1] iff ($\exists$ (there exists) $\gamma \in \mathbb{R}$ such that $v = \gamma w$ or $\exists \gamma \in \mathbb{R}$ such that $w = \gamma v$).*

Proof. Assume, there exists $(\alpha, \beta) \neq (0, 0)$ such that $\alpha v + \beta w = (0, 0)$. If $\alpha \neq 0$, then $v = \frac{\beta}{\alpha} w$. On the other hand, if $\beta \neq 0$, $w = \frac{\alpha}{\beta} v$. Thus, either $v = -\frac{\beta}{\alpha} w$ or $w = -\frac{\alpha}{\beta} v$.

Proof Strategy

The backward implication is a statement of the form $(p \vee q) \implies r$. We often prove statements of this form by using an equivalent form $(p \implies r) \wedge (q \implies r)$. Their equivalence can be seen through the following truth table.

p	q	r	$p \vee q$	$(p \vee q) \implies r$
T	T	T	T	T
T	T	F	T	F
T	F	T	T	T
T	F	F	T	F
F	T	T	T	T
F	T	F	T	F
F	F	T	F	T
F	F	F	F	T

[1]We will often write $(0, 0)$ as 0. This makes sense as $(0, 0)$ is the additive identity

p	q	r	$p \implies r$	$q \implies r$	$(p \implies r) \wedge (q \implies r)$
T	T	T	T	T	T
T	T	F	F	F	F
T	F	T	T	T	T
T	F	F	F	T	F
F	T	T	T	T	T
F	T	F	T	F	F
F	F	T	T	T	T
F	F	F	T	T	T

Similarly, we can prove that $(p \wedge q) \implies r$ is logically equivalent to $(p \implies r) \vee (q \implies r)$.

p	q	r	$p \wedge q$	$(p \wedge q) \implies r$
T	T	T	T	T
T	T	F	T	F
T	F	T	F	T
T	F	F	F	T
F	T	T	F	T
F	T	F	F	T
F	F	T	F	T
F	F	F	F	T

p	q	r	$p \implies r$	$q \implies r$	$(p \implies r) \vee (q \implies r)$
T	T	T	T	T	T
T	T	F	F	F	F
T	F	T	T	T	T
T	F	F	F	T	T
F	T	T	T	T	T
F	T	F	T	F	T
F	F	T	T	T	T
F	F	F	T	T	T

Now, assume $\exists$ (there exists) $\gamma \in \mathbb{R}$ such that $v = \gamma w$ or $\exists \gamma \in \mathbb{R}$ such that $w = \gamma v$. If $\exists$ (there exists) $\gamma \in \mathbb{R}$ such that $v = \gamma w$, then it we may write $v + (-\gamma)w = 0$. Thus, there exists (α, β) (namely $(\alpha, \beta) = (1, -\gamma)$) such that $(\alpha, \beta) \neq (0, 0)$ and $\alpha v + \beta w = 0$. Similarly, if $\exists \gamma \in \mathbb{R}$ such that $w = \gamma v$, then we have $(\alpha, \beta) = (\gamma, -1)$. □

Theorem 1.45. *For all $v, w \in \mathbb{R}^2$, if $\exists$ (there exists) $\gamma \in \mathbb{R}$ such that $v = \gamma w$, then Span$(v, w) \neq \mathbb{R}^2$.*

Russel's generalised conditional

The above statement has the form: $\forall v, w \in \mathbb{R}^2$, if $p(v, w)$ then $q(v, w)$. In the context of mathematics conditional sentences (a sentence like "if $p(v, w)$ then $q(v, w)$") are implicitly assumed to be "universally quantified". Thus, it is common practice in mathematical writings to drop the "for all" at the beginning of the statement of the theorem.

Proof.

$$\begin{aligned}\text{Span}(v,w) &= \{\alpha v+\beta w|\alpha,\beta\in\mathbb{R}\} \\ &= \{\alpha\gamma w+\beta w|\alpha,\beta\in\mathbb{R}\} \\ &= \{(\alpha\gamma+\beta)w|\alpha,\beta\in\mathbb{R}\} \\ &= \{kw|k\in\mathbb{R}\}\end{aligned}$$

The last equality is a bit trickier than the rest, so I would elaborate the argument involved. We need to show that $\{(\alpha\gamma+\beta)w|\alpha,\beta\in\mathbb{R}\}=\{kw|k\in\mathbb{R}\}$

Proof Strategy

Given two sets A and B, we show $A=B$ by showing $A\subseteq B$ and $B\subseteq A$. We show $A\subseteq B$ by showing an arbitrary element of A belongs to B.

Take an arbitrary element in $\{(\alpha\gamma+\beta)w|\alpha,\beta\in\mathbb{R}\}$, say $(\alpha\gamma+\beta)w$. As $\alpha,\beta,\gamma\in\mathbb{R}$, $\alpha\gamma+\beta$ is also a real number, say k. Thus, $(\alpha\gamma+\beta)w=kw\in\{kw|k\in\mathbb{R}\}$. Thus, $\{(\alpha\gamma+\beta)w|\alpha,\beta\in\mathbb{R}\}\subseteq\{kw|k\in\mathbb{R}\}$. Take an arbitrary element in $\{kw|k\in\mathbb{R}\}$ - say kw. Then, $kw=(0\gamma+k)w\in\{(\alpha\gamma+\beta)w|\alpha,\beta\in\mathbb{R}\}$ as $0,k\in\mathbb{R}$. Thus, $\{kw|k\in\mathbb{R}\}\subseteq\{(\alpha\gamma+\beta)w|\alpha,\beta\in\mathbb{R}\}$. Hence,

$$\text{Span}(v,w)=\{(\alpha\gamma+\beta)w|\alpha,\beta\in\mathbb{R}\}=\{kw|k\in\mathbb{R}\}.$$

Let $w=(w_1,w_2)$. If $(w_1,w_2+1)\in\{kw|k\in\mathbb{R}\}$, then $(w_1,w_2+1)=k(w_1,w_2)$. That is, $w_1=kw_1$ and $w_2=kw_2-1$. If $w_1=kw_1$, then $w_1=0$ or $k=1$. Note that $k=1\implies w_2=w_2-1\implies 1=0$. Thus, $k\neq 1$. Thus, $w_1=0$, that is $\{kw|k\in\mathbb{R}\}\subseteq\{(0,y)|y\in\mathbb{R}\}$. Therefore $\{kw|k\in\mathbb{R}\}$ is a strict subset of $\mathbb{R}^2$. If $(w_1,w_2+1)\notin\{kw|k\in\mathbb{R}\}$, then again $\{kw|k\in\mathbb{R}\}$ is a strict subset of $\mathbb{R}^2$. □

Similarly, we can prove

Theorem 1.46. *If $\exists\gamma\in\mathbb{R}$ such that $w=\gamma v$, then Span$(v,w)\neq\mathbb{R}^2$.*

Thus, we can concisely express our findings as follows:

Theorem 1.47. *For all $v,w\in\mathbb{R}^2$, Span$(v,w)=\mathbb{R}^2$ iff v,w are linearly independent.*

We have proved that $(a,c),(b,d)$ are linearly independent iff Span$((a,c),(b,d))=\mathbb{R}^2$ iff $\forall(e,f)$ the system 1.1 has a solution. However, the independence of $(a,c),(b,d)$ (and thus the fact Span$((a,c),(b,d))=\mathbb{R}^2$) not just assure the existence of a solution, it assures us that the solution is unique.

Theorem 1.48. *For all $(a,c),(b,d)\in\mathbb{R}^2$ if $(a,c),(b,d)$ are linearly independent then the solution to system 1.1 is unique.*

Proof. Fix arbitrary elements (a,c) and (b,d). We need to prove "if $(a,c),(b,d)$ are linearly independent then the solution to system 1.1 is unique". Let (x_1,y_1) and (x_2,y_2) be two solutions to system 1.1. We will prove uniqueness by showing $(x_1,y_1)=(x_2,y_2)$ (any two arbitrary solutions are forced to be equal). As (x_1,y_1) is a solution, we have $x_1(a,c)+y_1(b,d)=(e,f)$. As (x_2,y_2) is a solution, we have $x_2(a,c)+y_2(b,d)=(e,f)$. By the following computation (justify every step)

we get,

$$(x_1-x_2)(a,c)+(y_1-y_2)(b,d)=(0,0)$$

As (a,c) and (b,d) are linearly independent, this would mean that $x_1-x_2=0$ and $y_1-y_2=0$. That is, $x_1=x_2$ and $y_1=y_2$ or $(x_1,y_1)=(x_2,y_2)$. $\square$

Recall from the first section that if $ad-bc=0$, then 1.1 has a solution iff $\{(x,y)\mid ax+by=e\}=\{(x,y)\mid cx+dy=f\}$. Moreover, the set of solutions $\{(x,y)\mid ax+by=e \text{ and } cx+dy=f\}=\{(x,y)\mid ax+by=e\}$ or $\{(x,y)\mid ax+by=e \text{ and } cx+dy=f\}=\emptyset$. Let us revisit this idea using what we studied in this section.

Given a system 1.5, we can consider the related system

$$\begin{aligned} ax+by&=0\\ cx+dy&=0 \end{aligned} \tag{1.5}$$

A system of linear equations where all the constant terms are zero is called a homogeneous system of linear equations. Notice that a homogeneous system of linear equations will always have a solution, namely $(0,0)$. If $(a,c),(b,d)$ are linearly independent, then $(0,0)$ is the unique solution.

Theorem 1.49. *If $(a,c),(b,d)$ are linearly dependent, then there exists $(u_1,u_2)\in\mathbb{R}^2$ such that the set of solutions of system 1.5 is equal to $\{k(u_1,u_2)\mid k\in\mathbb{R}\}$.*

Proof. If $(a,c),(b,d)$ are linearly dependent, then $\exists\alpha\in\mathbb{R}$ such that $(a,c)=\alpha(b,d)$ or $\exists\alpha\in\mathbb{R}$ such that $(b,d)=\alpha(a,c)$.

Proof Strategy

The proof of the two cases

(1) $\exists\alpha\in\mathbb{R}$ such that $(a,c)=\alpha(b,d)$
(2) $\exists\alpha\in\mathbb{R}$ such that $(b,d)=\alpha(a,c)$

would be analogous. If one understands the proof in one case, one should be able to prove the other case on their own. In such cases, we may assume that we are in one of the choices - and our choice would not have any impact. Thus, we would say, that "we may choose without loss of generality (WLOG)".

Let us assume without loss of generality that $\exists\alpha\in\mathbb{R}$ such that $(a,c)=\alpha(b,d)$. That is $a=\alpha b$ and $c=\alpha d$. Thus, the system 1.5 takes the form

$$\begin{aligned} \alpha bx+by&=0\\ \alpha dx+dy&=0 \end{aligned} \tag{1.6}$$

Note that, $b = 0$ implies $a = 0$ - but we assumed a and b are not simultaneously 0. Thus, That is, $y = -\alpha x$. In other words, the set of solutions of system 1.5 is the set $\{k(1, -\alpha) \mid k \in \mathbb{R}\}$. In other words, $(u_1, u_2) = (1, -\alpha)$.

Better safe than sorry

To convince yourself that the other case is analogous, start with the assumption $\exists \alpha \in \mathbb{R}$ such that $(b, d) = \alpha(a, c)$. Thus, the system 1.5 takes the form

In other words, the set of solutions of system 1.5 is the set $\{k(-\alpha, 1) \mid k \in \mathbb{R}\}$. That is $(u_1, u_2) = (-\alpha, 1)$.

□

Now, let $(e, f) \in \mathbb{R}^2$ be such that system 1.1 has a solution. As a solution for system 1.1 exists, $\exists (v_1, v_2) \in \mathbb{R}^2$ such that $v_1(a, c) + v_2(b, d) = (e, f)$. Let (w_1, w_2) be an arbitrary solution of system 1.1. Then, from the following computation

we have $(w_1 - v_1, w_2 - v_2)$ is a solution of system 1.5. Thus, $(w_1 - v_1, w_2 - v_2) \in \{k(u_1, u_2) \mid k \in \mathbb{R}\}$. That is, $\exists k \in \mathbb{R}$ such that $(w_1, w_2) = (v_1, v_2) + k(u_1, u_2)$. Therefore, the set of solutions of system 1.1 (in this case) is equal to $\{(v_1, v_2) + k(u_1, u_2) \mid k \in \mathbb{R}\}$.

Remark 1.50. The two different ways of finding the set of solutions correspond the the two different definitions ($L_{p,q,r}$ and $\Lambda_{a,b,c,d}$) respectively discussed in Appendix A.

1.4. Functions and solving equations

Given $a, b, c, d \in \mathbb{R}$, we can define a function $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $F((x, y)) = (ax + by, cx + dy)$. Note that the system 1.1 has a solution iff $(e, f) \in \text{Image}(F)$. Thus,

Theorem 1.51. *The function $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $F((x, y)) = (ax + by, cx + dy)$ is surjective iff $\forall (e, f) \in \mathbb{R}^2$, the system 1.1 has a solution iff Span$(v, w) = \mathbb{R}^2$.*

Abuse of notation

The input for a function $F : \mathbb{R}^2 \to \mathbb{R}^2$ are elements of $\mathbb{R}^2$. Thus, the correct way to evaluate F at a point (x, y) is to write $F((x, y))$. However, you might have seen people write $F(x, y)$. This is convenient and does not usually lead to confusion. Thus, we will also generally represent $F((x, y))$ as $F(x, y)$. Such practices are called abuse of notation - when you encounter more examples you will get a better sense of what they are.

Theorem 1.52. *The function $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $F(x, y) = (ax + by, cx + dy)$ is injective iff (a, c) and (b, d) are linearly independent.*

Proof. Assume F is injective. Note that $F(0,0) = (0,0)$. As F is injective, if $F(x,y) = (0,0) = F(0,0)$, then $(x,y) = (0,0)$. But, $F(x,y) = (ax+by, cx+dy) = x(a,c) + y(b,d)$. That is if $x(a,c) + y(b,d) = 0$, then $(x,y) = (0,0)$. In other words, (a,c) and (b,d) are linearly independent.

Assume (a,c) and (b,d) are linearly independent. That is if $x(a,c)+y(b,d) = 0$, then $(x,y) = (0,0)$. But, $F(x,y) = (ax+by, cx+dy) = x(a,c)+y(b,d)$. Thus, if $F(x,y) = (0,0)$ then $(x,y) = (0,0)$. Now let $v = (v_1, v_2)$ and $w = (w_1, w_2)$ be arbitrary elements such that $F(v_1, v_2) = F(w_1, w_2)$, Then

$$\begin{aligned}
0 &= F(v_1, v_2) - F(w_1, w_2) \\
&= (av_1 + bv_2, cv_1 + dv_2) - (aw_1 + bw_2, cw_1 + dw_2) \\
&= ((av_1 + bv_2) - (aw_1 + bw_2), (cv_1 + dv_2) - (cw_1 + dw_2)) \\
&= (a(v_1 - w_1) + b(v_2 - w_2), c(v_1 - w_1) + d(v_2 - w_2)) \\
&= F(v_1 - w_1, v_2 - w_2)
\end{aligned}$$

Thus, $(v_1 - w_1, v_2 - w_2) = (0,0)$. That is, $v_1 = w_1$ and $v_2 = w_2$. Therefore, $(v_1, v_2) = (w_1, w_2)$. But, as v and w were arbitrary, F is an injection. □

Theorem 1.51 tells us that F is surjective iff $\text{Span}((a,c),(b,d)) = \mathbb{R}^2$ and Theorem 1.52 tells us that F is injective iff (a,c) and (b,d) are linearly independent. However, Theorem 1.47 tells us that $\text{Span}((a,c),(b,d)) = \mathbb{R}^2$ iff (a,c) and (b,d) are linearly independent. Therefore, we have

Theorem 1.53. *For all $a,b,c,d \in \mathbb{R}$, the function $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $F(x,y) = (ax+by, cx+dy)$ is injective iff it is surjective. Moreover, F is injective, surjective, and bijective iff $ad - bc \neq 0$.*

Exercise 1.54. Show that F is injective iff $\{(x,y) \mid F(x,y) = (0,0)\} = \{(0,0)\}$.

Definition 1.55. Given a function $F : \mathbb{R}^2 \to \mathbb{R}^2$ of the form $F(x,y) = (ax+by, cx+dy)$, the set $\{(x,y) \mid F(x,y) = (0,0)\} = \{(0,0)\}$ is called the kernel or null space of F. Note that $(0,0) \in \text{Ker}(F)$ and the function is injective iff $(0,0) = \text{Ker}(F)$

1.5. Matrices and solving equations

An array (generally of numbers) arranged into rows and columns is called a matrix. If the array has n rows and m columns, then we say that the matrix is an $n \times m$ matrix. For example, $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is a 2×2 matrix, $\begin{bmatrix} x \\ y \end{bmatrix}$ is a 2×1 matrix, and $\begin{bmatrix} x & y \end{bmatrix}$ is a 1×2 matrix.

Definition 1.56. Let

$$M_2(\mathbb{R}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid a,b,c,d \in \mathbb{R} \right\}$$

Definition 1.57. Let

$$M_2^1(\mathbb{R}) = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \mid x,y \in \mathbb{R} \right\}$$

The set of solutions 1.1 is the set $\{(x,y) \in \mathbb{R}^2 \mid ax+by = e \text{ and } cx+dy = f\}$. There is a natural bijection from $f : \mathbb{R}^2 \to M_2^1(\mathbb{R})$, namely the function $f(x,y) = \begin{bmatrix} x \\ y \end{bmatrix}$.

Exercise 1.58. Show that f is a bijection.

Theorem 1.59. *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *and* $v_0 = \begin{bmatrix} e \\ f \end{bmatrix}$. *Further, let*

$$V = \{(x, y) \in \mathbb{R}^2 | ax + by = e \text{ and } cx + dy = f\}$$

and

$$W = \{v \in M_2^1(\mathbb{R}) \mid Av = v_0\}.$$

Then, $f(V) = W$.

Proof.

$$\begin{aligned} f(V) &= \{f(v) \mid v \in V\} \\ &= \{f(x, y) \mid (x, y) \in \mathbb{R}^2 \text{ and } (ax + by = e \text{ and } cx + dy = f)\} \\ &= \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \mid (x, y) \in \mathbb{R}^2 \text{ and } (ax + by = e \text{ and } cx + dy = f) \right\} \\ &= \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix} \right\} \\ &= \left\{ v \in M_2^1(\mathbb{R}) \mid Av = v_0 \right\} \\ &= W. \end{aligned}$$

□

This is the reason why the system 1.1 is often expressed as

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix} \tag{1.7}$$

This representation gives a first clue on how to define multiplication of 2 matrices. We would want

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

Thus, you can think of the multiplication $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$ as the action of the function F on (x, y). If there is yet another matrix $\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}$, then it corresponds to the function $F' : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $F'(x, y) = (a'x + b'y, c'x + d'y)$. The action of F' on $F(x, y)$ should thus correspond to

$\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right)$. But, if have a "sensible definition of multiplication, we should have

$$\begin{aligned} \left(\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right) \\ &= \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \begin{bmatrix} ax+by \\ cx+dy \end{bmatrix} \\ &= \begin{bmatrix} a'(ax+by)+b'(cx+dy) \\ c'(ax+by)+d'(cx+dy) \end{bmatrix} \\ &= \begin{bmatrix} x(a'a+b'c)+y(a'b+b'd) \\ x(c'a+d'c)+y(c'b+d'd) \end{bmatrix} \\ &= \begin{bmatrix} (a'a+b'c) & (a'b+b'd) \\ (c'a+d'c) & (c'b+d'd) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \end{aligned}$$

This allows us to arrive at the definition of matrix multiplication you would have encountered in school.

$$\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} (a'a+b'c) & (a'b+b'd) \\ (c'a+d'c) & (c'b+d'd) \end{bmatrix}$$

Definition 1.60. We can define two operations on this set. Addition $+ : M_2(\mathbb{R}) \times M_2(\mathbb{R}) \to M_2(\mathbb{R})$ defined as

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} = \begin{bmatrix} a+a' & b+b' \\ c+c' & d+d' \end{bmatrix}$$

and multiplication $\times : M_2(\mathbb{R}) \to M_2(\mathbb{R})$ defined as

$$\begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} (a'a+b'c) & (a'b+b'd) \\ (c'a+d'c) & (c'b+d'd) \end{bmatrix}$$

Exercise 1.61. Show that

(1) for all $A, B, C \in M_2(\mathbb{R})$, $A + (B + C) = (A + B) + C$ (Addition is associative)

(2) for all $A, B \in M_2(\mathbb{R})$, $A + B = B + A$ (Addition is commutative)

(3) there exists a matrix X $\left(\text{namely } 0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right)$ such that $\forall A \in M_2(\mathbb{R})$, $A+X = A = X+A$ (Existence of additive identity)

(4) for all $A \in M_2(\mathbb{R})$, there exits a $B \in M_2(\mathbb{R})$ such that $A + B = 0 = B + A$. (Existence of additive inverse)

(5) for all $A, B, C \in M_2(\mathbb{R})$, $A(BC) = (AB)C$ (Multiplication is associative)

(6) there exists a matrix X $\left(\text{namely } I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$ such that $\forall A \in M_2(\mathbb{R})$, $AI = A = IA$ (Existence of multiplicative identity)

(7) for all $A, B \in M_2(\mathbb{R})$, $A(B+C) = AB+AC$. Similarly, for all $A, B \in M_2(\mathbb{R})$, $(A+B)C = AC + BC$ (Multiplication distributes over addition)

Example 1.62 (Matrix multiplication is not commutative)**.** Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$. Then,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1.(-1) + 2.0 & 1.0 + 2.1 \\ 3.(-1) + 4.0 & 3.0 + 4.1 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ -3 & 4 \end{bmatrix}$$

but,

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} (-1).1 + 0.3 & (-1).2 + 0.4 \\ 0.1 + 1.3 & 0.2 + 1.4 \end{bmatrix} = \begin{bmatrix} -1 & -2 \\ 3 & 4 \end{bmatrix}$$

Therefore,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \neq \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

Lemma 1.63. *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. *If* $\exists B \in M_2(\mathbb{R})$ *such that* $AB = I = BA$, *then* $\forall \begin{bmatrix} x \\ y \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \neq 0$.

Proof Strategy

A statement "p" is logically equivalent to the statement "$\neg p \implies F$. Intuitively, if the assumption that the statement is false leads to something absurd, then the statement has to be true. This principle is called **reductio ad absurdum** (reduction to absurdity) in Latin and **proof by contradiction** more commonly. Once again, the easiest way to prove the equivalence is using a truth table.

p	$\neg p$	$\neg p \implies F$
T	F	T
T	F	T
F	T	T
F	T	T

Proof. Assume $\exists B \in M_2(\mathbb{R})$ such that $AB = I = BA$. We need to show that $\forall \begin{bmatrix} x \\ y \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \neq 0$. We would intead assume it is false. That is, $\exists \begin{bmatrix} x \\ y \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ such that $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Then,

$$\begin{bmatrix} x \\ y \end{bmatrix} = (BA) \begin{bmatrix} x \\ y \end{bmatrix} = B \left(A \begin{bmatrix} x \\ y \end{bmatrix} \right) = B \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

□

Exercise 1.64. Justify each equality in the last line of the previous lemma.

Example 1.65 (A matrix need not have a multiplicative inverse)**.** The matrix $\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ does not have a multiplicative inverse as $\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

1.6. Elementary matrices

The three 2×2 matrices $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ are called elementary matrices. They obtain a special status due their role in finding the solution of system 1.1. To begin by showing that all these three matrices are invertible. The following exercise descibe their inverses.

Exercise 1.66. Show that

(1)

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

(2)

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

(3)

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

Definition 1.67. The set of all invertible 2×2 matrices is denoted as $GL_2(\mathbb{R})$

$$GL_2(R) = \{A \in M_2(\mathbb{R}) \mid \exists B \in M_2(\mathbb{R}) \text{ such that } AB = I = BA\}.$$

Theorem 1.68. *For all $A \in M_2(\mathbb{R})$, $\forall B \in GL_2(\mathbb{R})$, and $\forall v_0 \in M_2^1(\mathbb{R})$,*

$$\left\{v \in M_2^1(\mathbb{R}) \mid Av = v_0\right\} = \left\{v \in M_2^1(\mathbb{R}) \mid (BA)v = Bv_0\right\}.$$

Proof. Fix an arbitrary $A \in M_2(\mathbb{R})$, $B \in GL_2(\mathbb{R})$, and $v_0 \in M_2^1(\mathbb{R})$. We will show that for this choice, $\{v \in M_2^1(\mathbb{R}) \mid Av = v_0\} = \{v \in M_2^1(\mathbb{R}) \mid (BA)v = Bv_0\}$. As the choice was arbitrary, the more general result is proved. Further, we will prove $\{v \in M_2^1(\mathbb{R}) \mid Av = v_0\} = \{v \in M_2^1(\mathbb{R}) \mid (BA)v = Bv_0\}$ by showing $\{v \in M_2^1(\mathbb{R}) \mid Av = v_0\} \subseteq \{v \in M_2^1(\mathbb{R}) \mid (BA)v = Bv_0\}$ and $\{v \in M_2^1(\mathbb{R}) \mid (BA)v = Bv_0\} \subseteq \{v \in M_2^1(\mathbb{R}) \mid Av = v_0\}$.

Let $v \in \{v \in M_2^1(\mathbb{R}) \mid Av = v_0\}$ be arbitrary. Then $Av = v_0$. Multiplying on the left by B on both sides, we get $B(Av) = Bv_0$. But, $B(Av) = (BA)v$. Therefore $v \in \{v \in M_2^1(\mathbb{R}) \mid (BA)v = Bv_0\}$. As $v \in \{v \in M_2^1(\mathbb{R}) \mid Av = v_0\}$ was arbitrary, we have

$$\left\{v \in M_2^1(\mathbb{R}) \mid Av = v_0\right\} \subseteq \left\{v \in M_2^1(\mathbb{R}) \mid (BA)v = Bv_0\right\}.$$

Let $v \in \{v \in M_2^1(\mathbb{R}) \mid BAv = Bv_0\}$ be arbitrary. Then $(BA)v = Bv_0$. But, B is invertible. So, there exists $C \in M_2(\mathbb{R})$ such that $CB = I = C$. Multiplying both sides of the equation $BAv = Bv_0$

by C, we get $C((BA)v) = C(Bv_0)$. But,

$$\begin{aligned} C((BA)v) &= (C(BA))v \\ &= ((CB)A))v \\ &= (IA)v \\ &= Av \end{aligned}$$

and

$$\begin{aligned} C(Bv_0) &= (CB)v_0 \\ &= Iv_0 \\ &= v_0 \end{aligned}$$

Thus, $Av = v_0$ or in other words $v \in \{v \in M_2^1(\mathbb{R}) \mid Av = v_0\}$. As $v \in \{v \in M_2^1(\mathbb{R}) \mid BAv = Bv_0\}$ was arbitrary, we have

$$\left\{v \in M_2^1(\mathbb{R}) \mid (BA)v = Bv_0\right\} \subseteq \left\{v \in M_2^1(\mathbb{R}) \mid Av = v_0\right\}.$$

□

Notice that

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+c & b+d \\ c & d \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} e+f \\ f \end{bmatrix}$$

As $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is invertible, the previous theorem implies that

$$\left\{\begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix}\right\} = \left\{\begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a+c & b+d \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e+f \\ f \end{bmatrix}\right\}$$

The effect of multiplication by $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is that of "row addition". Multiplication by $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ has a very similar impact. More precisely,

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c+a & b+d \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} e \\ f+e \end{bmatrix}$$

Once again, as $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ is invertible, we have

$$\left\{\begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix}\right\} = \left\{\begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a & b \\ c+a & d+b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f+e \end{bmatrix}\right\}$$

Multiplication by $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ on the other hand has the effect of swapping the rows. More precisely,

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c & d \\ a & b \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} f \\ e \end{bmatrix}$$

Once again, as $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is invertible, we have

$$\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix} \right\} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} c & d \\ a & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ e \end{bmatrix} \right\}$$

Exercise 1.69. Use the fact that $\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ is invertible to show that

$$\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix} \right\} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a-c & b-d \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e-f \\ f \end{bmatrix} \right\} \tag{1.8}$$

Exercise 1.70. Use the fact that $\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ is invertible to show that

$$\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix} \right\} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a & b \\ c-a & d-b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f-e \end{bmatrix} \right\} \tag{1.9}$$

Exercise 1.71. For all $\lambda \in \mathbb{R}$ such that $\lambda \neq 0$, show that $\begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix}$ are invertible. Thus, show that

$$\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix} \right\} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} \lambda a & \lambda b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \lambda e \\ f \end{bmatrix} \right\} \tag{1.10}$$

and

$$\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix} \right\} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} a & b \\ \lambda c & \lambda d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ \lambda f \end{bmatrix} \right\} \tag{1.11}$$

1.7. Gaussian elimination and Matrix multiplication

The ideas in the last section is implicitly used in the method taught in school. Recall that if you are asked to find a solution to

$$x + 2y = 0$$
$$2x + y = 3$$

You use the fact that the set of solutions of this system is same as the set of solutions of the system

$$x + 2y = 0$$
$$x + \frac{1}{2}y = \frac{3}{2}$$

In other words, you are saying (using Equation 1.11)

$$\left\{\begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}\right\} = \left\{\begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} 1 & 2 \\ 1 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{3}{2} \end{bmatrix}\right\}$$

Now we say, that the set of solutions of this new system is same as the set of solutions of (we are subtracting the first equation from the second)

$$x + 2y = 0$$
$$0x + \frac{-3}{2}y = \frac{3}{2}$$

In other words, we are using Equation 1.8 to conclude

$$\left\{\begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} 1 & 2 \\ 1 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{3}{2} \end{bmatrix}\right\} = \left\{\begin{bmatrix} x \\ y \end{bmatrix} \in M_2^1(\mathbb{R}) \mid \begin{bmatrix} 1 & 2 \\ 0 & \frac{-3}{2} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{3}{2} \end{bmatrix}\right\}$$

Now, it is easy to find the value of y and that is then used to find the value of x (this process is called back-substitution). There are two key ideas here

(1) A matrix whose entries below the diagonal are zero is called an upper triangular matrix. If a system is described by an upper triangular matrix, then it is easy to solve the system.

(2) We could find a matrix M, namely $M = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & \frac{1}{2} \end{bmatrix}$ such that $M\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & \frac{-3}{2} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$ an upper triangular matrix. This is a general idea - given any matrix A, you can find a matrix M such that MA is upper triangular.

This example is a special case of a technique called Gaussian elimination. The basic idea is that if the matrix is upper (or lower) triangular, then it is easy to solve the corresponding system of linear equations. Thus, it would be great if given any matrix A we can find an **invertible** (invertibility is essential to ensure the set of solutions remain same) matrix B and an upper triangular matrix U such that $BA = U$. Although that is too much to ask for, we have something close enough.

Definition 1.72. A matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is called upper triangular if $c = 0$ and is called lower triangular if $a = 0$.

Theorem 1.73. *For all $A \in M_2(\mathbb{R})$, there exists* ***an invertible*** *matrix B such that BA is upper triangular or there exists* ***an invertible*** *matrix B such that BA is lower triangular.*

Proof. Fix an arbitrary matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = A \in M_2(\mathbb{R})$. If $a = 0$, then A is a lower triangular matrix to begin with. Thus, we can take B to be the identity matrix. If $c = 0$, then A is an upper triangular matrix to begin with. Thus, we can take B to be the identity matrix again. If $a \neq 0 \neq c$, then $\begin{bmatrix} 1 & 0 \\ -\frac{c}{a} & 1 \end{bmatrix}$ is invertible and $\begin{bmatrix} 1 & 0 \\ -\frac{c}{a} & 1 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & \frac{ad-bc}{a} \end{bmatrix}$. Thus, we have found a matrix $B = \begin{bmatrix} 1 & 0 \\ -\frac{c}{a} & 1 \end{bmatrix}$ such that $BA = \begin{bmatrix} a & b \\ 0 & \frac{ad-bc}{a} \end{bmatrix}$ is upper triangular. □

Exercise 1.74. Explain why the above "proof" is a proof. **Hint: Go through the proof strategy in pages 11-12.**

Exercise 1.75. Prove that if $a \neq 0 \neq c$, then $\begin{bmatrix} 1 & 0 \\ -\frac{c}{a} & 1 \end{bmatrix}$ is invertible and $\begin{bmatrix} 1 & 0 \\ -\frac{c}{a} & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & \frac{ad-bc}{a} \end{bmatrix}$

Exercise 1.76. For the following matrices A, find a matrix B such that BA is either upper triangular or lower triangular.

(1) $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ (2) $\begin{bmatrix} 1 & 0 \\ 3 & 4 \end{bmatrix}$ (3) $\begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}$ (4) $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

Exercises

(1) Find all $(x, y) \in \mathbb{R}^2$ such that
   (a)

(2) Let $(a, b), (c, d)$ belong to $\mathbb{R}^2$. If $\langle (a, b), (c, d) \rangle = 0$ and $\|(c, d)\| = \|(a, b)\|$, then show that $(c, d) = (-b, a)$.

(3) Show that
   (a) $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & a+b \\ c & c+d \end{bmatrix}$
   (b) $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} a+b & b \\ c+d & d \end{bmatrix}$
   (c) $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} b & a \\ d & c \end{bmatrix}$

   Thus, multiplication on the right by elementary matrices correspond to column operations.

Solving m equations in n unknowns

The previous chapter was a warm-up before we tackle the more general question of solving m equations in n unknowns. If there are two unknowns, we may call them x and y. If there are 3 we may call them x, y, and z. But, if we need arbitrarily large number of unknowns, then it is difficult to use the letters of the alphabet to represent them. Therfore, we would use subscripts to denote these unknowns. The n unknowns would typically be written as $x_1, x_2, \ldots, x_n$. In fact, we already used this notation in the previous chapter in some places - mainly because I am so used to it. Thus, the most general system of m equations in n unknowns would have the following form:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
&. \\
&. \\
&. \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{aligned} \tag{2.1}
$$

Thus, the solutions to this system of equations is a tuple $(x_1, x_2, \ldots, x_n)$ where each x_i is a real number. We will start by first understanding the collection of all such tuples.

2.1. Cartesian products

The set $\mathbb{R}^2$ (the Cartesian plane) was the set $\{(x, y) \mid x, y \in \mathbb{R}\}$. This is a an example of a more general construction called the Cartesian product.

Definition 2.1 (Cartesian product)**.** The Cartesian product of two sets A and B is the set $A \times B := \{(a, b) \mid a \in A \text{ and } b \in B\}$

Given a number a, the quantity $a \times a$ is represented as a^2. Following this convention, the set $\mathbb{R} \times \mathbb{R}$ is represented as $\mathbb{R}^2$. Intuitively, we feel $(x_1, \ldots, x_n)$ should similarly be in $\mathbb{R}^n$. But, even when $n = 3$, there is some ambiguity in the meaning for $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$. Do I mean $(\mathbb{R} \times \mathbb{R}) \times \mathbb{R}$ or

$\mathbb{R} \times (\mathbb{R} \times \mathbb{R})$? The associativity of real number multiplication implied that $(a \times a) \times a = a \times (a \times a)$ and thus there was no confusion. However, the Cartesian product is not associative! As sets, $(A \times B) \times C$ is not the same as $A \times (B \times C)$ - an element of $(A \times B) \times C$ has the form $((a, b), c)$ while an element of $A \times (B \times C)$ has the form $(a, (b, c))$. But, we feel they are the same in an intuitive sense - both $((a, b), c)$ and $(a, (b, c))$ are just a triple of numbers (a, b, c). In other words, there is a natural bijection $f : (A \times B) \times C \to A \times (B \times C)$ defined as $f(((a, b), c)) = (a, (b, c))$.

Exercise 2.2. Show that $f : (A \times B) \times C \to A \times (B \times C)$ defined as $f(((a, b), c)) = (a, (b, c))$ is a bijection.

And thus, we would love to treat them as the same. However, if you would like to be really pedantic, the way is to stick to some convention - I decide that whenever I write $A \times B \times C$, I mean $(A \times B) \times C$. Of course, someone else might choose the convention that $A \times B \times C$ stands for $A \times (B \times C)$. The natural bijection we discussed earlier implies that the "mathematics" we both develop would not be so different from each other.

That level of formalism is neither necessary not helpful unless you are working with the foundations of mathematics. For instance, if we want to think $\mathbb{R}^n$ as an n-fold product, it would be better to define it recursively as $\mathbb{R}^{n-1} \times \mathbb{R}$. However, then the elements would be of $\mathbb{R}^3$ would be of the form $((x, y), z)$. Keeping track of these notations can get unwieldy. Thus,

Definition 2.3. We define $\mathbb{R}^n := \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{R}\}$.

As in the case of $\mathbb{R}^2$, we can define addition, scalar multiplication, norm, distance, etc in $\mathbb{R}^n$. We will mimic these constructions now.

Definition 2.4. For two points $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, we say $x = y$ if $\forall i,\ x_i = y_i$.

2.1.1. Addition. Given two elements in $\mathbb{R}^n$, say $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, we can add these elements to form an element $x + y = (x_1 + y_1, \ldots, x_n + y_n)$. This defines a function $+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ defined as $+((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = (x_1 + y_1, \ldots, x_n + y_n)$. We will typically write $(x_1, \ldots, x_n) + (y_1, \ldots, y_n)$ instead of $+((x_1, \ldots, x_n), (y_1, \ldots, y_n))$ - as you have been doing for addition in $\mathbb{R}^2$. This is called **infix notation**.

Example 2.5. Let $v = (1, 2, 3, 4)$ and $w = (5, 6, 7, 8)$. Then $v + w = (1, 2, 3, 4) + (5, 6, 7, 8) = (1 + 5, 2 + 6, 3 + 7, 4 + 8) = (6, 8, 10, 12)$.

Theorem 2.6 (Existence of additive identity)**.** *There exists an element $x = (x_1 \ldots, x_n)$ in $\mathbb{R}^n$ such that $\forall a = (a_1, \ldots, a_n) \in \mathbb{R}^n$, $a + x = a = x + a$. Thus, from now on, we would write $(0, \ldots, 0)$ as 0.*

Proof. To prove an existential statement, it is enough to give an example. We would prove that $x = (0, \ldots, 0)$ has the property $\forall a = (a_1, \ldots, a_n) \in \mathbb{R}^n$, $a + x = a = x + a$. To this end, fix an arbitrary $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. Then, we need to prove $a + x = a$ and $x + a = a$. But, they are both easy to prove.

$$\begin{aligned} (a_1, \ldots a_n) + (0, \ldots, 0) &= (a_1 + 0, \ldots, a_n + 0) && \text{(definition of addition on } \mathbb{R}^n) \\ &= (a_1, \ldots, a_n) && \text{(0 is additive identity in } \mathbb{R}) \end{aligned}$$

Similarly,

$$\begin{aligned} (0, \ldots, 0) + (a_1, \ldots, a_n) &= (0 + a_1, \ldots, 0 + a_n) && \boxed{} \\ &= (a_1, \ldots, a_n) && \boxed{} \end{aligned}$$

∎

Note that the proof is almost identical to the proof we had written to prove the existence of additive identity in $\mathbb{R}^2$. Now mimic the proofs we had in Chapter 1, and

Exercise 2.7. Prove

(1) Addition on $\mathbb{R}^n$ is commutative. More precisely, $\forall a = (a_1, \dots, a_n), b = (b_1, \dots, b_n) \in \mathbb{R}^n$, $a + b = b + a$.

(2) Addition on $\mathbb{R}^n$ is associative. More precisely, $\forall a = (a_1, \dots, a_n), b = (b_1, \dots, b_n), c = (c_1, \dots, c_n) \in \mathbb{R}^n$, $(a + b) + c = a + (b + c)$.

(3) For all $a = (a_1, \dots, a_n) \in \mathbb{R}^n$, there exists an $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ such that $a + x = 0 = x + a$. Use this to define subtraction on $\mathbb{R}^n$

Exercise 2.8. Let $v = (5, 3, 2, 7)$ and $w = (2, 7, 3, 8)$. Then find $v - w$.

2.1.2. Scalar Multiplication. We define a function $\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ defined as $\cdot(\alpha, (a_1, \dots, a_n)) = (\alpha a_1, \dots, \alpha a_n)$. Typically, we just write $\alpha(a_1, \dots, a_n)$ instead of $\cdot(\alpha, (a_1, \dots, a_n))$.

Example 2.9. Let $k = 3$ and $v = (1, 2, 3, 4)$. Then $kv = 3(1, 2, 3, 4) = (3, 6, 9, 12)$.

Exercise 2.10. Show that

(1) $\forall v \in \mathbb{R}^n \; 1 \cdot v = v$

(2) $\forall v \in \mathbb{R}^n \; v + (-1) \cdot v = 0$

(3) $\forall \; x, y \in \mathbb{R}$ and $v \in \mathbb{R}^n$, $(xy) \cdot v = x \cdot (y \cdot v)$

(4) $\forall x \in \mathbb{R}$ and $v, w \in \mathbb{R}^n$, $x \cdot [v + w] = x \cdot v + x \cdot w$

(5) $\forall \; x, y \in \mathbb{R}$ and $v \in \mathbb{R}^n$, $(x + y) \cdot v = x \cdot v + y \cdot v$

2.1.3. Inner Product. You may have studied this as the dot product. This is the function $\langle \; , \; \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined as $\langle (v_1, \dots v_n), (w_1, \dots, w_n) \rangle = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$.

Exercise 2.11. Show that

(1) $\forall v, w \in \mathbb{R}^n$, $\langle v, w \rangle = \langle w, v \rangle$

(2) $\forall u, v, w \in \mathbb{R}^n$, $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$

(3) $\forall x \in \mathbb{R}$ and $\forall v, w \in \mathbb{R}^n$, $\langle xu, w \rangle = x \langle v, w \rangle = \langle v, xw \rangle$

(4) $\forall v \in \mathbb{R}^n$, $\langle v, v \rangle \geq 0$ and the equality holds **if and only if (iff)** $v = 0$

2.1.4. Norm. is the function $\| \; \| : \mathbb{R}^2 \to \mathbb{R}$ defined as $\|(v_1, \dots, v_n)\| = \langle (v_1, \dots, v_n), (v_1, \dots, v_n) \rangle^{\frac{1}{2}} = \sqrt{v_1^2 + \cdots + v_n^2}$.

Exercise 2.12. Find the norm of $(2, 3, 4)$. Give an example of another element of $\mathbb{R}^3$ with the same norm.

Exercise 2.13. Show that $\|x(a, b)\| = |x| \|(a, b)\|$.

2.1.5. Euclidean distance. is the function $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined as $d(v, w) = \|v - w\| = \sqrt{(v_1 - w_1)^2 + \cdots + (v_n - w_n)^2}$.

Exercise 2.14. Find the distance between $(1, 2, 3)$ and $(1, 1, 1)$.

2.2. Span of m vectors in $\mathbb{R}^n$

Definition 2.15. Given m vectors $v_1, \dots, v_m \in \mathbb{R}^n$ and $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ an expression of the form $\alpha_1 v_1 + \dots \alpha_n v_n$ is called a linear combination of the vectors v_i. Span of m vectors is the set of all linear combinations - more precisely,

$$Span(v_1, \dots, v_m) = \{\alpha_1 v_1 + \dots \alpha_n v_m \mid \alpha_1, \dots, \alpha_m \in \mathbb{R}\}$$

GeoGebra Exercises

Exercise 2.16. Consider the element $(1, 2, 3) \in \mathbb{R}^3$. The $\text{Span}((1, 2, 3)) = \{x(1, 2, 3) \mid x \in \mathbb{R}\} = \{(x, 2x, 3x) \mid x \in \mathbb{R}\}$. GeoGebra allows us to plot things in 3D. To demonstrate this, I have created a GeoGebra file that plots $\text{Span}((1, 2, 3))$. Use GeoGebra to plot the span of other vectors in a similar manner. Convince yourself that if $v \neq 0$, then $\text{Span}(v)$ looks like a line.

Exercise 2.17. In Exercise 1.35, we saw how we can use the Sequence function to construct the span of two vectors. Use this idea in the GeoGebra 3D calculator (introduced in the previous exercise) to plot the span of 2-vectors in $\mathbb{R}^3$. Repeat the exercise with two pairs of Linearly dependent vectors and two pairs of linearly independent vectors. What do you observe? How does the span look when the vectors are linearly independent?

Example 2.18. Let $v_1 = (1, 0, 0)$, $v_2 = (0, 1, 0)$. Then,

$$\begin{aligned}
\text{Span}(v_1, v_2) &= \{x(1, 0, 0) + y(0, 1, 0) \mid x, y \in \mathbb{R}\} \\
&= \{(x, 0, 0) + (0, y, 0) \mid x, y \in \mathbb{R}\} \\
&= \{(x, y, 0) \mid x, y \in \mathbb{R}\}
\end{aligned}$$

Thus, the $\text{Span}(v_1, v_2)$ is what is called the xy plane. Similarly, $\text{Span}((1, 0, 0), (0, 0, 1))$ would be the xz plane and $\text{Span}((0, 1, 0), (0, 0, 1))$ would be the yz plane.

Example 2.19. Let $v_1 = (1, 0, 0)$, $v_2 = (0, 1, 0)$, $v_3 = (0, 0, 1)$. Then,

$$\begin{aligned}
\text{Span}(v_1, v_2, v_3) &= \{x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1) \mid x, y, z \in \mathbb{R}\} \\
&= \{(x, 0, 0) + (0, y, 0) + (0, 0, z) \mid x, y, z \in \mathbb{R}\} \\
&= \{(x, y, z) \mid x, y, z \in \mathbb{R}\} \\
&= \mathbb{R}^3
\end{aligned}$$

Thus, three vectors can span $\mathbb{R}^3$. Can two vectors span $\mathbb{R}^3$? Can a single vector span $\mathbb{R}^3$? In Chapter 1, we defined span only for two vectors. As part of one of the proofs in Chapter 1, we argued that $\forall v \in \mathbb{R}^2, \text{Span(v)} \neq \mathbb{R}^2$ - although we did not state it as explicitly. Which theorem am I referring to? Use the arguments given there to

Exercise 2.20. Prove that $\forall v \in \mathbb{R}^2$, $\text{Span(v)} \neq \mathbb{R}^2$.

More generally, prove that

Exercise 2.21. Let $n > 1$. Prove that $\forall v \in \mathbb{R}^n$, $\text{Span(v)} \neq \mathbb{R}^n$.

So, a single vector cannot span $\mathbb{R}^3$. Can two vectors span $\mathbb{R}^3$? The intuition we built in Exercise 2.17 suggests two vectors cannot span $\mathbb{R}^3$.

Exercise 2.22. Prove that if $w = \alpha v$, then $\text{Span}(v, w) = \text{Span}(v)$

Thus, if there exists an $\alpha \in \mathbb{R}$ such that $w = \alpha v$ or there exists an α such that $v = \alpha w$, then $\text{Span}(v, w) \neq \mathbb{R}^2$. Moreover,

Exercise 2.23. Show that there exists an $\alpha \in \mathbb{R}$ such that $w = \alpha v$ or there exists an α such that $v = \alpha w$ iff $(v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1) = (0, 0, 0)$.

Theorem 2.24. *Given any $v, w \in \mathbb{R}^3$, Span$(v, w) \neq \mathbb{R}^3$.*

Proof. Consider the vector $v \times w = (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1)$[1]. If $v \times w = 0$, then by the discussion earlier, $\text{Span}(v, w) \neq \mathbb{R}^3$. Thus, we may assume that $v \times w \neq 0$. (Why? Which proof strategy am I using?) As $v \times w \neq 0$, we may assume without-loss-of-generality that $v_1w_2 - v_2w_1 \neq 0$. We will prove that if $v_1w_2 - v_2w_1 \neq 0$, then $v \times w$ does not belong to $\text{Span}(v, w)$ using a proof by contradiction.

Assume $v \times w \in \text{Span}(v, w)$. Then, there exists $\alpha, \beta \in \mathbb{R}$ such that $(v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1) = \alpha v + \beta w = (\alpha v_1 + \beta w_1, \alpha v_2 + \beta w_2, \alpha v_3 + \beta w_3)$. That is,

$$\begin{aligned}\alpha v_1 + \beta w_1 &= v_2w_3 - v_3w_2\\ \alpha v_2 + \beta w_2 &= v_3w_1 - v_1w_3\\ \alpha v_3 + \beta w_3 &= v_1w_2 - v_2w_1\end{aligned}$$

The first two equations can be rewritten as the following matrix equation

$$\begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} v_2w_3 - v_3w_2 \\ v_3w_1 - v_1w_3 \end{bmatrix}$$

By our assumption $v_1w_2 - v_2w_1 \neq 0$, the matrix $\begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \end{bmatrix}$ is invertible. Thus,

$$\begin{aligned}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} &= \frac{1}{v_1w_2 - v_2w_1} \begin{bmatrix} w_2 & -w_1 \\ -v_2 & v_1 \end{bmatrix} \begin{bmatrix} v_2w_3 - v_3w_2 \\ v_3w_1 - v_1w_3 \end{bmatrix}\\ &= \frac{1}{v_1w_2 - v_2w_1} \begin{bmatrix} w_3(w_2v_2 + w_1v_1) - v_3(w_2^2 + w_1^2) \\ v_3(v_2w_2 + v_1w_1) - w_3(v_1^2 + v_2^2) \end{bmatrix}\end{aligned}$$

Substituting this value of α and β in $\alpha v_3 + \beta w_3 = v_1w_2 - v_2w_1$, we get

$$\begin{aligned}(v_1w_2 - v_2w_1)^2 &= v_3w_3(w_2v_2 + w_1v_1) - v_3^2(w_2^2 + w_1^2) + v_3w_3(v_2w_2 + v_1w_1) - w_3^2(v_1^2 + v_2^2)\\ &= -[(v_2w_3 - v_3w_2)^2 + (v_1w_3 - v_3w_1)^2]\end{aligned}$$

However, $(v_1w_2 - v_2w_1)^2 \geq 0$ and $-[(v_2w_3 - v_3w_2)^2 + (v_1w_3 - v_3w_1)^2] \leq 0$. So, the two can be equal iff $v_1w_2 - v_2w_1 = 0$. But, we assumed $v_1w_2 - v_2w_1 \neq 0$. Thus, we have arrived at a contradiction. Therefore the assumption $v \times w \in \text{Span}(v, w)$ must be wrong and $\text{Span}(v, w) \neq \mathbb{R}^3$. □

We could have proved that if $v \times w \neq 0$, then $v \times w \notin \text{Span}(v, w)$ in a slightly different way. First,

Exercise 2.25. Prove that $\langle v, v \times w \rangle = 0 = \langle w, v \times w \rangle$.

Then, use the theorem,

Theorem 2.26. *Let $v \neq 0, w \neq 0 \in \mathbb{R}^n$. For all $0 \neq u \in \mathbb{R}^n$ if $\langle v, u \rangle = 0 = \langle w, u \rangle$, then $u \notin$ Span(v, w).*

[1] The vector $(v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1)$ is called the cross product of v and w. Some of you might have studied the cross product in school.

But, to prove the above theorem, we need a very famous and important result

Theorem 2.27 (Cauchy-Schwarz inequality)**.** *Let $v, w \in \mathbb{R}^n$. Then, $\langle v, w\rangle \le \|v\|\|w\|$. Moreover, the equality holds iff $\exists\alpha \in \mathbb{R}$ such that $v = \alpha w$ or $\exists\alpha \in \mathbb{R}$ such that $w = \alpha v$.*

Proof. Let us first assume that $\|v\| = 1 = \|w\|$. Then,

$$\begin{aligned} 0 &\le \|v - w\|^2 \\ &= \langle v - w, v - w\rangle \\ &= \langle v, v\rangle + \langle w, w\rangle - \langle v, w\rangle - \langle w, v\rangle \\ &= \langle v, v\rangle + \langle w, w\rangle - 2\langle v, w\rangle \\ &= 2 - 2\langle v, w\rangle \end{aligned}$$

Thus, $\langle v, w\rangle \le 1 = \|v\|\|w\|$. Moreover, the equality holds iff $\|v - w\| = 0$, that is $v - w = 0$ or $v = w$.

Now, let v, w be two arbitrary vectors in $\mathbb{R}^n$. Then consdider the vectors $v' = \frac{1}{\|v\|}v$ and $w' = \frac{1}{\|w\|}w$. Then, by the previous observation, $\langle v', w'\rangle \le \|v'\|^2\|w'\|^2 = 1$. But,

$$\begin{aligned} \langle v', w'\rangle &= \langle \frac{1}{\|v\|}v, \frac{1}{\|w\|}w\rangle \\ &= \frac{1}{\|v\|}\langle v, \frac{1}{\|w\|}w\rangle \\ &= \frac{1}{\|v\|\|w\|}\langle v, w\rangle \end{aligned}$$

Thus,

$$\frac{1}{\|v\|\|w\|}\langle v, w\rangle \le 1$$

and hence we have the result. □

Proof of Theorem 2.26. To prove the result, we first fix a non-zero vector $u \in \mathbb{R}^3$. Then, assume that $\langle v, u\rangle = 0 = \langle w, u\rangle$. We will now prove that $u \notin \text{Span}(v, w)$ using a proof by contradiction.

Assume, $u \in \text{Span}(v, w)$. Then, there exists α, β such that $u = \alpha v + \beta w$. But, $0 = \langle v, u\rangle = \langle v, \alpha v + \beta w\rangle = \alpha\langle v, v\rangle + \beta\langle v, w\rangle = \alpha\|v\|^2 + \beta\langle v, w\rangle$. That is, $\alpha = -\beta\frac{\langle v,w\rangle}{\|v\|^2}$. Similarly, $0 = \langle w, u\rangle = \langle w, \alpha v + \beta w\rangle = \alpha\langle w, v\rangle + \beta\langle w, w\rangle = \alpha\langle w, v\rangle + \beta\|w\|^2$. That is, $\beta = -\alpha\frac{\langle v,w\rangle}{\|w\|^2} = -\left(-\beta\frac{\langle v,w\rangle}{\|v\|^2}\right)\frac{\langle v,w\rangle}{\|w\|^2} = \beta\frac{\langle v,w\rangle^2}{\|v\|^2\|w\|^2}$. In a very similar fashion one can prove that $\alpha = \alpha\frac{\langle v,w\rangle^2}{\|v\|^2\|w\|^2}$. As $u \ne 0$, $(\alpha, \beta) \ne (0, 0)$. Thus, $\langle v, w\rangle^2 = \|v\|^2\|w\|^2$. The famous Cauchy-Schwarz inequality tells us that this is possible iff $\exists\gamma \in \mathbb{R}$ such that $v = \gamma w$ or $\exists\gamma \in \mathbb{R}$ such that $w = \gamma v$. Without loss of generality, let us assume that $w = \gamma v$. Then, $u = \alpha v + \beta(\gamma v) = (\alpha + \beta\gamma)v$. But $0 = \langle u, v\rangle = \langle(\alpha + \beta\gamma)v, v\rangle = (\alpha + \beta\gamma)\|v\|^2$. As $v \ne 0$, $\|v\| \ne 0$. Therefore, $\alpha + \beta\gamma = 0$ or equivalently, $u = (\alpha + \beta\gamma)v = 0v = 0$. But, we assumed $u \ne 0$, so we have a contradiction. Therefore, our assumption that $u \in \text{Span}(u, v)$ should be wrong. □

OK, the span of two vectors $v, w \in \mathbb{R}^3$ is not equal to $\mathbb{R}^3$. Then, what is it?

Example 2.28. If $w = \alpha v$, then

$$\begin{aligned}
\text{Span(v,w)} &= \{xv + yw \mid x, y \in \mathbb{R}\} \\
&= \{xv + y(\alpha v)) \mid x, y \in \mathbb{R}\} \\
&= \{xv + (y\alpha)v \mid x, y \in \mathbb{R}\} \\
&= \{(x + y\alpha)v \mid x, y \in \mathbb{R}\} \\
&= \{zv \mid z \in \mathbb{R}\}
\end{aligned}$$

The last equality is a bit more subtle than the rest. We will prove $\{(x+y\alpha)v \mid x, y \in \mathbb{R}\} = \{zv \mid z \in \mathbb{R}\}$, by showing $\{(x + y\alpha)v \mid x, y \in \mathbb{R}\} \subset \{zv \mid z \in \mathbb{R}\}$ and $\{zv \mid z \in \mathbb{R}\} \subset \{(x + y\alpha)v \mid x, y \in \mathbb{R}\}$. To see the first containment, note that as x, y, and α are real numbers, so is $x + y\alpha$. Thus, $\{(x + y\alpha)v \mid x, y \in \mathbb{R}\} \subset \{zv \mid z \in \mathbb{R}\}$. On the other hand, given any $z \in \mathbb{R}$, choose $(x, y, \alpha) = (z, 0, \alpha)$ and $z = x+y\alpha$. So, $zw \in \{(x+y\alpha)v \mid x, y \in \mathbb{R}\}$. Thus, $\{zv \mid z \in \mathbb{R}\} \subset \{(x+y\alpha)v \mid x, y \in \mathbb{R}\}$.

Example 2.29. Let $v_1 = (1, 0, 0)$, $v_2 = (1, 1, 0)$. Then,

$$\begin{aligned}
\text{Span}(v_1, v_2) &= \{\alpha(1, 0, 0) + \beta(0, 1, 0) \mid \alpha, \beta \in \mathbb{R}\} \\
&= \{(\alpha, 0, 0) + (0, \beta, 0) \mid \alpha, \beta \in \mathbb{R}\} \\
&= \{(\alpha, \beta, 0) \mid \alpha, \beta \in \mathbb{R}\} \\
&= \{(x, y, z) \in \mathbb{R}^3 \mid z = 0\} \\
&= \{(x, y, z) \in \mathbb{R}^3 \mid \langle (x, y, z), (0, 0, 1)\rangle = 0\}
\end{aligned}$$

Example 2.30. Let $v_1 = (2, 1, 0)$, $v_2 = (3, 0, 1)$. Then,

$$\begin{aligned}
\text{Span}(v_1, v_2) &= \{\alpha(2, 1, 0) + \beta(3, 0, 1) \mid \alpha, \beta \in \mathbb{R}\} \\
&= \{(2\alpha + 3\beta, \alpha, \beta) \mid \alpha, \beta \in \mathbb{R}\} \\
&= \{(x, y, z) \in \mathbb{R}^3 \mid x = 2y + 3z\} \\
&= \{(x, y, z) \in \mathbb{R}^3 \mid \langle (x, y, z), (1, -2, -3)\rangle = 0\}
\end{aligned}$$

Example 2.31. Let $v_1 = (7, 2, 1)$, $v_2 = (0, 1, 0)$. Then,

$$\begin{aligned}
\text{Span}(v_1, v_2) &= \{\alpha(7, 2, 1) + \beta(0, 1, 0) \mid \alpha, \beta \in \mathbb{R}\} \\
&= \{(7\alpha, 2\alpha + \beta, \alpha) \mid \alpha, \beta \in \mathbb{R}\} \\
&= \{(x, y, z) \in \mathbb{R}^3 \mid z = 7x\} \\
&= \{(x, y, z) \in \mathbb{R}^3 \mid \langle (x, y, z), (-1, 0, 7)\rangle = 0\}
\end{aligned}$$

In the previous three examples, note that $(1, 0, 0) \times (0, 1, 0) = (0, 0, 1)$, $(2, 1, 0) \times (3, 0, 1) = (1, -2, -3)$, and $(7, 2, 1) \times (0, 1, 0) = (-1, 0, 7)$. Thus, all the three examples illustrate the following more general phenomenon.

GeoGebra Exercises

Exercise 2.32. Use GeoGebra to plot $\text{Span}(v_1, v_2)$ and $v_1 \times v_2$ for the following pairs for vectors.

(1) $v_1 = (1, 0, 0)$, $v_2 = (1, 1, 0)$
(2) $v_1 = (2, 1, 0)$, $v_2 = (3, 0, 1)$
(3) $v_1 = (7, 2, 1)$, $v_2 = (0, 1, 0)$
(4) $v_1 = (1, 2, 3)$, $v_2 = (4, 5, 6)$

Exercise 2.33. Find the Span(v_1, v_2) where $v_1 = (1, 2, 3)$, $v_2 = (4, 5, 6)$. Express Span(v_1, v_2) as $\{(x, y, z) \mid ax + by + cz = 0\}$. Give rigorous arguments to justify your claims.

More precisely, we have the following theorem. However, we will prove this theorem only later in this chapter after introducing more abstraction. Try to prove the theorem using what you know now - this would help you appreciate the need for the abstraction.

Theorem 2.34. *Given two vectors $v, w \in \mathbb{R}^3$ such that $v \times w \neq 0$, Span$(v, w) = \{u \in \mathbb{R}^3 \mid \langle u, v \times w\rangle = 0\}$.*

Exercise 2.35. Let two vectors $v, w \in \mathbb{R}^3$. Show that if $u \in$ Span(v, w), then $\langle u, v \times w\rangle = 0$.

Exercise 2.36. Let $v_1, v_2, v_3 \in \mathbb{R}^n$ be such that $v_3 = \alpha_1 v_1 + \alpha_2 v_2$. Show that Span$(v_1, v_2, v_3) =$ Span(v_1, v_2).

2.3. Subspace

Theorem 2.37. *For all vectors $v_1, \dots, v_m \in \mathbb{R}^n$, Span$(v_1, \dots, v_m)$ satisfies the following conditions.*

(1) $0 \in$ *Span*$(v_1, \dots, v_m)$
(2) For all $u_1, u_2 \in$ *Span*$(v_1, \dots, v_m)$, $u_1 + u_2 \in$ *Span*$(v_1, \dots, v_m)$
(3) For all $u \in$ *Span*$(v_1, \dots, v_m)$, $\alpha u \in$ *Span*$(v_1, \dots, v_m)$

Proof. Note that $0 = 0v_1 + \cdots + 0v_n \in$ Span$(v_1, \dots, v_m)$. Also, if $u_1 \in$ Span$(v_1, \dots, v_m)$, then there exits $\alpha_i \forall i \in \{1, \dots, m\}$ such that $u_1 = \alpha_1 v_1 + \cdots + \alpha_m v_m$. Similarly, if $u_2 \in$ Span$(v_1, \dots, v_m)$, then there exists $\beta_i \forall i \in \{1, \dots, m\}$ such that $u_2 = \beta_1 v_1 + \dots \beta_m v_m$. Thus,

$$\begin{aligned} u_1 + u_2 &= \alpha_1 v_1 + \cdots + \alpha_m v_m + \beta_1 v_1 + \dots \beta_m v_m \\ &= (\alpha_1 + \beta_1) v_1 + \cdots + (\alpha_m + \beta_m) v_m \\ &\in \text{Span}(v_1, \dots, v_m) \end{aligned}$$

□

Similarly, one can prove that

Exercise 2.38. Consider the system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &. \\ &. \\ &. \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned} \tag{2.2}$$

The set $S = \{x = (x_1, \dots, x_n) \mid (x_1, \dots, x_n)$ satisfy the system 2.2$\}$ satisfies the following three conditions

(1) $0 \in S$
(2) For all $u_1, u_2 \in S$, $u_1 + u_2 \in S$
(3) For all $u \in S$, $\alpha u \in S$

As subsets satisfing these three properties keep appearing, it is a good idea to name them.

Definition 2.39. A subset V of $\mathbb{R}^n$ satisfying

(1) $0 \in V$
(2) For all $v_1, v_2 \in V$, $v_1 + v_2 \in V$
(3) For all $v \in V$ and $\alpha \in \mathbb{R}$, $\alpha v \in V$

None of the three conditions above are redundant. More precisely, there exists subsets that satisfy two of the above conditions, but not the third. The following three examples illustrate the various possibilities.

Example 2.40. Let $V = \emptyset \subset \mathbb{R}^2$. Then, for all $v_1, v_2 \in V$, $v_1 + v_2 \in V$. Also, for all $v \in V$ and $\alpha \in \mathbb{R}$, $\alpha v \in V$. However, V is not a subspace as $0 \notin V$.

Example 2.41. Let $V = \mathbb{Z} \times \mathbb{Z} \subset \mathbb{R}^2$. Then, $(0,0) \in \mathbb{Z} \times \mathbb{Z}$. Also, if $v_1, v_2 \in \mathbb{Z} \times \mathbb{Z}$, then $v_1 + v_2 \in \mathbb{Z} \times \mathbb{Z}$, as sum of two integers is an integer. However, $\pi(1,1) \notin \mathbb{Z} \times \mathbb{Z}$ and thus, V is not a subspace.

Example 2.42. Let $V = \{(x,y) \mid x = 0\} \cup \{(x,y) \mid y = 0\}$. Then, $(0,0) \in V$. Let $(x,y) \in V$. Then either $x = 0$ or $y = 0$. If $x = 0$, then $\alpha(x,y) = (\alpha x, \alpha y) = (0, \alpha y) \in V$. If $y = 0$, then $\alpha(x,y) = (\alpha x, \alpha y) = (\alpha x, 0) \in V$. However, although $(1,0) \in V$ and $(0,1) \in V$, $(1,1) = (1,0) + (0,1)$ does not belong to V. Thus, V is not a subspace.

The notion of a subspace allows us to give an alternate characterisation for $\mathrm{Span}(v_1, \ldots, v_n)$. We already saw that $\mathrm{Span}(v_1, \ldots, v_n)$. We will now argue that $\mathrm{Span}(v_1, \ldots, v_n)$ is the "smallest" subspace containing $v_1, \ldots, v_n$. But, to talk about smallest, we should be able to compare two subspaces - there should be an order. When we are talking about subsets, the most natural order is given by the $\subseteq$ relation. We say A is "smaller" (to be precise we mean smaller or equal) than B if $A \subseteq B$. Thus, we may state our claim more precisely, as follows:

Theorem 2.43. *Let $v_1, \ldots, v_m \in \mathbb{R}^n$. Let U be a subspace of $\mathbb{R}^n$ such that $v_i \in U$, $\forall i \in \{1, \ldots, m\}$. Then $V := \mathrm{Span}(v_1, \ldots, v_m) \subseteq U$.*

Proof. We will prove $V \subseteq U$ by taking an arbitrary element in V and showing that it belongs to U. Fix an arbitrary element of V or more precisely fix $(\alpha_1, \ldots, \alpha_m)$ and consider the element $v = \alpha_1 v_1 + \cdots + \alpha_m v_m$. We would like to show that $v \in U$. To begin, as $v_i \in U$ and U is a subspace, $\alpha_i v_i \in U$ for all $i \in \{1, \ldots, m\}$. As $\alpha_1 v_1$ and $\alpha_2 v_2$ belong to U and U is a subspace, $\alpha_1 v_1 + \alpha_2 v_2 \in U$. Now suppose $\alpha_1 v_1 + \cdots + \alpha_k v_k \in U$. As $\alpha_1 v_1 + \cdots + \alpha_k v_k \in U$, $\alpha_{k+1} v_{k+1} \in U$ and U is a vector space, $\alpha_1 v_1 + \cdots + \alpha_{k+1} v_{k+1} \in U$. Thus, $v \in U$. AS $v \in V$ was arbitrary, we have $V \subseteq U$. ☐

Exercise 2.44. For each subset given below, describe the smallest subspace of $\mathbb{R}^3$ that contains it.

(1) $\{(1,0,0)\}$
(2) $\{(\pi,0,0)\}$
(3) $\{(1,0,0),(0,1,0)\}$
(4) $\{(1,1,0)(1,-1,0)\}$
(5) $\{(1,1,0),(-1,-1,0)\}$
(6) $\{(1,1,0),(1,-1,0),(0,1,1)\}$

Optional reading: Span(S) where $S \subset \mathbb{R}^n$

Theorem 2.45. *Let Λ be some set. Let $\mathcal{V} = \{V_\lambda \mid \lambda \in \Lambda\}$ be a set such that V_λ is a subspace of $\mathbb{R}^n$ for all λ. Then, the set $V := \bigcap_{\lambda \in \Lambda} V_\lambda$ is a subspace of $\mathbb{R}^n$.*

Proof. As V_λ is a subspace, $0 \in V\lambda$ for all λ. Thus, $0 \in V$. Let $v_1, v_2 \in V$. Then, $v_1, v_2 \in V_\lambda$ for all λ and V_λ is a subspace. Thus, $v_1 + v + 2 \in V_\lambda$ for all λ. Thus, $v_1 + v_2 \in V$. Let $v \in V$. Then $v \in V_\lambda$ for all λ and V_λ is a subspace. Thus, $\alpha v \in V_\lambda$ for all λ. Therefore $\alpha v \in V$. □

Given $S \subset \mathbb{R}^n$, let $\mathcal{V} = \{V \subset \mathbb{R}^n \mid V \text{ is a subspace and } S \subset V\}$ and $W = \bigcap_{V \in \mathcal{V}} V$. Then, W is a subspace of $\mathbb{R}^n$, by the previous theorem. Moreover, if V is a subspace containing S, then $V \subset W$. Thus, given any set S, we can find the smallest subspace containing it. This fact and Theorem 2.43 motivates the following definition.

Definition 2.46. Let $S \subset \mathbb{R}^n$, then $\text{Span}(S)$ is defined to be the subspace V of $\mathbb{R}^n$ satisfying the following two conditions.

(1) $S \subseteq V$

(2) If U is a subspace of $\mathbb{R}^n$ such that $S \subseteq U$, then $V \subseteq U$.

Thus, it is preferable to write $\text{Span}(\{v_1, \ldots, v_m\})$ instead of $\text{Span}(v_1, \ldots, v_m)$.

Exercise 2.47. Show that if $S \subset T$, then $\text{Span}(S) \subset \text{Span}(T)$.

2.4. Functions and solving equations; associated matrices

Given system 2.1 or 2.2, we can associate to the system a function $F : \mathbb{R}^n \to \mathbb{R}^m$ defined as

$$F(x_1, \ldots, x_n) = \left(\sum_{i=1}^n a_{1i} x_i, \ldots, \sum_{i=1}^n a_{mi} x_i \right)$$

The set of solutions of system 2.1 is the set $F^{-1}(\{(b_1, \ldots, b_m)\})$ and the set of solutions of system 2.2 is the set $F^{-1}(\{0\})$.

Exercise 2.48. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be the function $F(x_1, \ldots, x_n) = \left(\sum_{i=1}^n a_{1i} x_i, \ldots, \sum_{i=1}^n a_{mi} x_i \right)$. Then, show that

(1) $F(v + w) = F(v) + F(w)$.

(2) $F(\alpha v) = \alpha F(v)$

Exercise 2.49. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a function such that $F(v + w) = F(v) + F(w)$ and $F(\alpha v) = \alpha F(v)$. Then, show that there exists a_{ij} for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$ such that $F(x_1, \ldots, x_n) = \left(\sum_{i=1}^n a_{1i} x_i, \ldots, \sum_{i=1}^n a_{mi} x_i \right)$. **Hint:** Recall that if $F : \mathbb{R}^2 \to \mathbb{R}^2$, then $F(x, y) = F(x(1, 0) + y(1, 0)) = xF(1, 0) + yF(0, 1)$.

Definition 2.50. A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is said to be linear, if

(1) $F(v + w) = F(v) + F(w)$.

(2) $F(\alpha v) = \alpha F(v)$

Definition 2.51. Given a linear function $F : \mathbb{R}^n \to \mathbb{R}^m$ we call the set $\{(x_1, \ldots, x_n) \mid F(x_1, \ldots, x_n) = 0\}$ the kernel of F or the null space of F. Usually, we denote this set as $Ker(F)$ or $Null(F)$. Similarly, the set $\{F(x_1, \ldots, x_n) \mid (x_1, \ldots, x_n) \in \mathbb{R}^n\}$ is the image of F (also called the range of F) and is denoted as $Im(F)$.

Exercise 2.52. Given a linear function $F : \mathbb{R}^n \to \mathbb{R}^m$, show that $Ker(F)$ and $Im(F)$ are both subspaces.

Example 2.53. Let $F : \mathbb{R}^2 \to \mathbb{R}^3$ be the function $F(x, y) = (x + 3y, 2x + 2y, 3x + y)$. Then,

$$\begin{aligned} Ker(F) &= \{(x, y) \in R^2 \mid F(x, y) = (0, 0, 0)\} \\ &= \{(x, y) \in R^2 \mid (x + 3y, 2x + 2y, 3x + y) = (0, 0, 0)\} \end{aligned}$$

As $2x + 2y = 0$, we have $y = -x$. Substituting this value of y in $x + 3y = 0$, we get $x - 3x = 0$ or $x = 0$. Substituting this value in $3x + y = 0$, we get $y = 0$. Thus, $(x, y) \in \{(x, y) \in R^2 \mid (x + 3y, 2x + 2y, 3x + y) = (0, 0, 0)\}$ implies that $(x, y) = (0, 0)$. That is,

$$Ker(F) = \{(x, y) \in R^2 \mid (x + 3y, 2x + 2y, 3x + y) = (0, 0, 0)\} = \{(0, 0)\}.$$

Now,

$$\begin{aligned} Im(F) &= \{F(x, y) | (x, y) \in \mathbb{R}^2\} \\ &= \{(x + 3y, 2x + 2y, 3x + y) | (x, y) \in \mathbb{R}^2\} \\ &= \left\{(a, b, c) \in \mathbb{R}^3 \mid b = \frac{a + c}{2}\right\} \end{aligned}$$

Example 2.54. Let $F : \mathbb{R}^3 \to \mathbb{R}^2$ be the function $F(x, y, z) = (x + y, z)$. Then,

$$\begin{aligned} Ker(F) &= \{(x, y, z) \in R^3 \mid F(x, y, z) = (0, 0)\} \\ &= \{(x, y, z) \in R^3 \mid (x + y, z) = (0, 0)\} \\ &= \{(x, -x, 0) \mid x \in \mathbb{R}\} \end{aligned}$$

And,

$$Im(F) = \{F(x, y, z) | (x, y, z) \in \mathbb{R}^3\} = \mathbb{R}^2$$

To see that $Im(F) = \mathbb{R}^2$, we will show that an arbitrary element $(a, b) \in \mathbb{R}^2$ belongs to $Im(F)$. More precisely, note that $F(a, 0, b) = (a + 0, b) = (a, b)$.

Example 2.55. Let $F : \mathbb{R}^2 \to \mathbb{R}^4$ be the function $F(x, y) = \left(\sqrt{2}x + y, 3x + \frac{2}{3}y, \frac{x}{3}, 4y\right)$. Then,

$$\begin{aligned} Ker(F) &= \{(x, y) \in \mathbb{R}^2 \mid F(x, y) = (0, 0, 0, 0)\} \\ &= \left\{(x, y) \in \mathbb{R}^2 \mid \left(\sqrt{2}x + y, 3x + \frac{2}{3}y, \frac{x}{3}, 4y\right) = (0, 0, 0, 0)\right\} \end{aligned}$$

Thus, $(x, y) \in Ker(F)$ iff $\left(\sqrt{2}x + y, 3x + \frac{2}{3}y, \frac{x}{3}, 4y\right) = (0, 0, 0, 0)$. But, if $\left(\sqrt{2}x + y, 3x + \frac{2}{3}y, \frac{x}{3}, 4y\right) = (0, 0, 0, 0)$, then $\frac{x}{3} = 0$ and $4y = 0$. Thus, $x = 0 = y$. Therefore, $(x, y) \in Ker(F)$ iff $(x, y) = (0, 0)$. That is,

$$Ker(F) = \{(0, 0)\}.$$

And,

$$\begin{aligned} Im(F) &= \{F(x, y) | (x, y) \in \mathbb{R}^2\} \\ &= \left\{\left(\sqrt{2}x + y, 3x + \frac{2}{3}y, \frac{x}{3}, 4y\right) | (x, y) \in \mathbb{R}^2\right\} \end{aligned}$$

Thus, if $(a, b, c, d) \in Im(F)$, there exists, $(x, y) \in \mathbb{R}^2$ such that $a = \sqrt{2}x + y$, $b = 3x + \frac{2}{3}y$, $c = \frac{x}{3}$, and $d = 4y$. Therefore, $3\sqrt{2}c + \frac{d}{4} = 3\sqrt{2}\frac{x}{3} + \frac{4y}{4} = \sqrt{2}x + y = a$. Similarly, $9c + \frac{d}{6} = 9\frac{x}{3} + \frac{4y}{6} = 3x + \frac{2}{3}y = b$. Thus, if $(a, b, c, d) \in Im(F)$, then if $(a, b, c, d) \in \left\{(a, b, c, d) \mid a = 3\sqrt{2}c + \frac{d}{4}, b = 9c + \frac{d}{6}\right\}$. That is $Im(F) \subset \left\{(a, b, c, d) \mid a = 3\sqrt{2}c + \frac{d}{4}, b = 9c + \frac{d}{6}\right\}$.

Now, let $(a, b, c, d) \in \left\{(a, b, c, d) \mid a = 3\sqrt{2}c + \frac{d}{4}, b = 9c + \frac{d}{6}\right\}$. Then, choose $x = 3c$ and $y = \frac{d}{4}$. Then, $a = 3\sqrt{2}c + \frac{d}{4} = 3\sqrt{2}\frac{x}{3} + y = 3\sqrt{2}x + y$ and $b = 9c + \frac{d}{6} = 9\frac{x}{3} + \frac{4y}{6} = 3x + \frac{2}{3}y$. Thus, $(a, b, c, d) =$

$\left(\sqrt{2}x+y, 3x+\frac{2}{3}y, \frac{x}{3}, 4y\right) \in Im(F)$. That is $\left\{(a,b,c,d) \mid a = 3\sqrt{2}c + \frac{d}{4}, b = 9c + \frac{d}{6}\right\} \subset Im(F)$. Combining with the earlier observation $Im(F) \subset \left\{(a,b,c,d) \mid a = 3\sqrt{2}c + \frac{d}{4}, b = 9c + \frac{d}{6}\right\}$, we have

$$Im(F) = (a,b,c,d) \in \left\{(a,b,c,d) \mid a = 3\sqrt{2}c + \frac{d}{4}, b = 9c + \frac{d}{6}\right\}$$

Example 2.56. Let $F : \mathbb{R}^4 \to \mathbb{R}^2$ be the function $F(a,b,c,d) = (2a+b+c, b, 3c)$. Then,

$$\begin{aligned} Ker(F) &= \{(a,b,c,d) \in R^4 \mid F(a,b,c,d) = (0,0)\} \\ &= \left\{(a,b,c,d) \in R^4 \mid (2a+b+c, b, 3c) = (0,0,0)\right\} \\ &= \{(0,0,0,d) \mid d \in \mathbb{R}\} \end{aligned}$$

And,

$$\begin{aligned} Im(F) &= \{F(a,b,c,d) | (a,b,c,d) \in \mathbb{R}^4\} \\ &= \{(2a+b+c, b, 3c) | (a,b,c,d) \in \mathbb{R}^4\} \\ &= \mathbb{R}^3 \end{aligned}$$

To see that $Im(F) = \mathbb{R}^3$, we will show that an arbitrary element $(x,y,z) \in \mathbb{R}^3$ belongs to $Im(F)$. More precisely, note that

$$\begin{aligned} F\left(\frac{x}{2} - \frac{y}{2} - \frac{z}{6}, y, \frac{z}{3}, 0\right) &= \left(2\left(\frac{x}{2} - \frac{y}{2} - \frac{z}{6}\right) + y + \frac{z}{3}, y, 3\frac{z}{3}\right) \\ &= \left(x - y - \frac{z}{3} + y + \frac{z}{3}, y, z\right) \\ &= (x,y,z) \end{aligned}$$

Exercise 2.57. Find the $Ker(F)$ and $Im(F)$ for the following functions:

(1) $f : \mathbb{R}^2 \to \mathbb{R}^3$ defined as $f(x,y) = (x, y, x-y)$.
(2) $f : \mathbb{R}^3 \to \mathbb{R}^3$ defined as $f(x,y,z) = (x+4y+13z, 2x+5y+14z, 3x+6y+15z)$.
(3) $f : \mathbb{R}^3 \to \mathbb{R}^2$ defined as $f(x,y,z) = (2x+3y, z)$.
(4) $f : \mathbb{R}^3 \to \mathbb{R}^4$ defined as $f(x,y,z) = (x+y, y+z, x+z, x+y+z)$.
(5) $f : \mathbb{R}^2 \to \mathbb{R}^3$ defined as $f(x,y) = (x+3y, 2x+5y, 7x+9y)$.
(6) $f : \mathbb{R}^4 \to \mathbb{R}^6$ defined as $f(x,y,z,w) = (x+y, x+z, x+w, y+z, y+w, z+w)$.

Definition 2.58. Let $M_m^n(\mathbb{R}) = \left\{ \begin{bmatrix} a_{11} & a_{12} \dots & a_{1n} \\ a_{21} & a_{22} \dots & a_{2n} \\ . & . & . \\ . & . & . \\ . & . & . \\ a_{m1} & \dots & a_{mn} \end{bmatrix} : \ a_{ij} \in \mathbb{R} \forall i \in 1, \dots m \text{ and } j \in \{1, \dots, n\} \right\}$ be the set of $m \times n$ matrices.

Optional reading

Exercise 2.59. Show that $\Phi_n : \mathbb{R}^n \to M_n^1(\mathbb{R})$ defined as $F(x_1, \ldots, x_n) = \begin{bmatrix} x_1 \\ . \\ . \\ . \\ x_n \end{bmatrix}$ is a bijection.

As there is a canonical bijection between $\mathbb{R}^n$ and $M_n^1(\mathbb{R})$, we may often treat them as the same. Thus, for instance, we can talk about inner product on $M_n^1(\mathbb{R})$.

As before, this identification between $\mathbb{R}^n$ and $M_n^1(\mathbb{R})$ (and similarly $\mathbb{R}^m$ and $M_m^1(\mathbb{R})$) allows us to think of F as a function from $M_n^1(\mathbb{R})$ to $M_m^1(\mathbb{R})$. More precisely, given $F : \mathbb{R}^n \to \mathbb{R}^m$, we can define $\overline{F} := \Phi_m \circ F \circ \Phi_n^{-1} : M_n^1(\mathbb{R}) \to M_m^1(\mathbb{R})$. As we saw before, matrix multiplication is defined to ensure that

(1) $$\overline{F}\left(\begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}\right) = \begin{bmatrix} a_{11} & a_{12} \ldots & a_{1n} \\ a_{21} & a_{22} \ldots & a_{2n} \\ . & . & . \\ . & . & . \\ . & . & . \\ a_{m1} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

(2) If $F : \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^m \to \mathbb{R}^k$ are such that

$$\overline{F}\left(\begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}\right) = \begin{bmatrix} a_{11} & a_{12} \ldots & a_{1n} \\ a_{21} & a_{22} \ldots & a_{2n} \\ . & . & . \\ . & . & . \\ . & . & . \\ a_{m1} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

and

$$\overline{G}\left(\begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_m \end{bmatrix}\right) = \begin{bmatrix} b_{11} & b_{12} \ldots & b_{1m} \\ b_{21} & b_{22} \ldots & b_{2m} \\ . & . & . \\ . & . & . \\ . & . & . \\ b_{k1} & \ldots & b_{km} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_m \end{bmatrix}$$

then,

$$\overline{G \circ F}\left(\begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}\right) = \left(\begin{bmatrix} b_{11} & b_{12} \ldots & b_{1m} \\ b_{21} & b_{22} \ldots & b_{2m} \\ . & . & . \\ . & . & . \\ . & . & . \\ b_{k1} & \ldots & b_{km} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \ldots & a_{1n} \\ a_{21} & a_{22} \ldots & a_{2n} \\ . & . & . \\ . & . & . \\ . & . & . \\ a_{m1} & \ldots & a_{mn} \end{bmatrix}\right) \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{bmatrix}$$

To each function $F : \mathbb{R}^n \to \mathbb{R}^m$ of the form $F(x_1, \dots, x_n) = \left(\sum_{i=1}^{n} a_{1i}x_i, \dots, \sum_{i=1}^{n} a_{mi}x_i\right)$, we can associate an $m \times n$ matrix

$$\begin{bmatrix} a_{11} & a_{12}\dots & a_{1n} \\ a_{21} & a_{22}\dots & a_{2n} \\ . & . & . \\ . & . & . \\ . & . & . \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

these two conditions would force the usual definition of matrix multiplication

Definition 2.60. Let $A = \begin{bmatrix} a_{11} & a_{12}\dots & a_{1n} \\ a_{21} & a_{22}\dots & a_{2n} \\ . & . & . \\ . & . & . \\ . & . & . \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12}\dots & b_{1m} \\ b_{21} & b_{22}\dots & b_{2m} \\ . & . & . \\ . & . & . \\ . & . & . \\ b_{k1} & \dots & b_{km} \end{bmatrix}$. Then, we define the product BA as the matrix whose il-th entry is equal to $\sum_{i=1}^{n} b_{ij}a_{jl}$. If $B_1 \dots B_k$ are the rows of B and $A^1, \dots, A^n$ are the columns of A, then $BA = \begin{bmatrix} \langle B_1, A^1\rangle \dots & \langle B_1, A^n\rangle \\ . & . \\ . & . \\ . & . \\ \langle B_k, A^1\rangle \dots & \langle B_k, A^n\rangle \end{bmatrix}$[2].

Once again, let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear function. Then, there exists a_{ij} such that $F(x_1, \dots, x_n) = \left(\sum_{i=1}^{n} a_{1i}x_i, \dots, \sum_{i=1}^{n} a_{mi}x_i\right)$. Then,

$$\begin{aligned} Im(F) &= \{F(x_1, \dots, x_n) \mid (x_1, \dots, x_n) \in \mathbb{R}^n\} \\ &= \left\{\left(\sum_{i=1}^{n} a_{1i}x_i, \dots, \sum_{i=1}^{n} a_{mi}x_i\right) \mid (x_1, \dots, x_n) \in \mathbb{R}^n\right\} \\ &= \left\{\sum_{i=1}^{n} x_i\,(a_{1i}, \dots, a_{mi}) \mid (x_1, \dots, x_n) \in \mathbb{R}^n\right\} \\ &= \left\{\sum_{i=1}^{n} x_i A^i \mid (x_1, \dots, x_n) \in \mathbb{R}^n\right\} \end{aligned}$$

Thus, the image of F is the span of the columns of the corresponding matrix! Therefore, the system 2.1 has a solution for a given $(b_1, \dots, b_n)$ iff $(b_1, \dots, b_n) \in \text{Span}(A^1, \dots, A^n)$. Also, F is surjective iff $\mathbb{R}^m = \text{Span}(A^1, \dots, A^n)$

Theorem 2.61. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear function. Then, F is injective iff $Ker(F) = \{0\}$.*

Proof Strategy

[2] This notation is inspired from Introduction to Linear Algebra, Serge Lang

A statement of the form "if p then q" is logically equivalent to the statement "if $\neg q$, then $\neg p$". The latter is called the **contrapositive** of the former. Often it would be easier to prove the contrapositive. Once again, the easiest way to prove the equivalence is using a truth table.

p	q	$p \implies q$	$\neg q$	$\neg p$	$\neg q \implies \neg q$
T	T	T	F	F	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

Proof. Instead of proving "if F is injective, then $Ker(F) = 0$", we will prove its contrapositive. More precisely, we will prove that if $Ker(F) \neq \{0\}$, then F is not injective. This however is really easy. Assume $Ker(F) \neq \{0\}$, then there exists $0 \neq v \in Ker(F)$. But, then $F(v) = 0 = F(0)$, but $v \neq 0$. Thus, F is not injective.

Now we will prove "If $Ker(F) = \{0\}$, then F is injective". So, we will assume $Ker(F) = \{0\}$ and prove F is injective. In other words, we would assume $F(x_1, \ldots, x_n) = F(y_1, \ldots, y_n)$ and prove that $(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$. Notice that as F is linear, there exists constants a_{ij} such that $F(x_1, \ldots, x_n) = \left(\sum_{i=1}^{n} a_{1i}x_i, \ldots, \sum_{i=1}^{n} a_{mi}x_i\right)$. Thus,

$$\begin{aligned}
0 &= F(x_1, \ldots, x_n) - F(y_1, \ldots, y_n) \\
&= \left(\sum_{i=1}^{n} a_{1i}x_i, \ldots, \sum_{i=1}^{n} a_{mi}x_i\right) - \left(\sum_{i=1}^{n} a_{1i}y_i, \ldots, \sum_{i=1}^{n} a_{mi}y_i\right) \\
&= \left(\sum_{i=1}^{n} a_{1i}(x_i - y_i), \ldots, \sum_{i=1}^{n} a_{mi}(x_i - y_i)\right) \\
&= F(x_1 - y_1, \ldots, x_n - y_n)
\end{aligned}$$

Thus, $(x_1 - y_1, \ldots, x_n - y_n) \in Ker(F)$. Hence $(x_1 - y_1, \ldots, x_n - y_n) = 0$ that is $(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$. □

Theorem 2.62. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear function. Let A be the corresponding matrix and $A^1, \ldots, A^n$ be the columns of the matrix. Then, F is injective iff $x_1A_1 + \cdots + x_nA_n = 0$ implies that $(x_1, \ldots, x_n) = 0$.*

Proof. Assume F is injective. Notice that Then $F(x_1, \ldots, x_n) = x_1A_1 + \cdots + x_nA_m$. Thus, $x_1A_1 + \cdots + x_nA_m$ implies that $F(x_1, \ldots, x_n) = 0$. As, F is injective and $F(0) = 0$, $(x_1, \ldots, x_n) = 0$.

Now, assume $x_1A_1 + \cdots + x_nA_n = 0$ implies that $(x_1, \ldots, x_n) = 0$. Thus, if $(x_1, \ldots, x_n) \in Ker(F)$, then $0 = F(x_1, \ldots, x_n) = x_1A_1 + \cdots + x_nA_m$. Thus, $(x_1, \ldots, x_n) = 0$. So, $Ker(F) = \{0\}$ and hence F is injective. □

The above theorem motivates the next concept central in linear algebra.

2.5. Linear Independence

Definition 2.63. Let $v_1, \ldots, v_k \in \mathbb{R}^n$. We say $(v_1, \ldots, v_k)$ is linearly independent if $\forall(\alpha_1, \alpha_2, \ldots, \alpha_k) \in \mathbb{R}^k$, $\alpha_1v_1 + \cdots + \alpha_kv_k = 0 \implies (\alpha_1, \ldots, \alpha_k) = (0, \ldots, 0)$.

Example 2.64. The vectors $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ are linearly independent. To check linear independence, fix 3 arbitrary real numbers α_1, α_2, and α_3 and equate $(0, 0, 0) = \alpha_1(1, 0, 0) +$

$\alpha_2(0,1,0)+\alpha_3(0,0,1)=(\alpha_1,\alpha_2,\alpha_3)$. Thus, the only solution is $(\alpha_1,\alpha_2,\alpha_3)=(0,0,0)$ and hence the vectors are linearly independent.

Example 2.65. The vectors $(1,-1,0)$, $(1,1,0)$, and $(1,1,1)$ are linearly independent. To check linear independence, fix 3 arbitrary real numbers α_1, α_2, and α_3 and equate $(0,0,0)=\alpha_1(1,-1,0)+\alpha_2(1,1,0)+\alpha_3(1,1,1)=(\alpha_1+\alpha_2+\alpha_3,-\alpha_1+\alpha_2+\alpha_3,\alpha_3)$. Equating the z-coordinates, we get $\alpha_3=0$. Substituting this value of α_3, we get two equations $\alpha_1+\alpha_2=0$ and $-\alpha_1+\alpha_2=0$. That is $\alpha_1+\alpha_2=-\alpha_1+\alpha_2$, that is $\alpha_1=-\alpha_1$. Thus, $\alpha_1=0$. Substituting this value of α_1 in either of the equations, we get $\alpha_3=0$. Thus, the only solution is $(\alpha_1,\alpha_2,\alpha_3)=(0,0,0)$ and hence the vectors are linearly independent.

Example 2.66. The vectors $(1,-1,1)$, $(1,1,1)$, and $(2,0,2)$ are not linearly independent. The easiest way to prove the claim is to just observe that $(1).(1,-1,1)+(1)(1,1,1)+(-1).(2,0,2)=(1+1-2,(-1)+1+0,1+1-2)=(0,0,0)$. Thus, there exists a non-zero $(\alpha_1,\alpha_2,\alpha_3)\in\mathbb{R}^3$ (namely $(1,1,-1)$) such that $\alpha_1v_1+\alpha_2v_2+\alpha_3v_3=0$.

We can also solve it in a systematic way which would be more useful in general. Fix 3 arbitrary real numbers α_1, α_2, and α_3 and equate $(0,0,0)=\alpha_1(1,-1,1)+\alpha_2(1,1,1)+\alpha_3(2,0,2)=(\alpha_1+\alpha_2+2\alpha_3,-\alpha_1+\alpha_2,\alpha_1+\alpha_2+2\alpha_3)$. Thus, we have just two equations $-\alpha_1+\alpha_2=0$ and $\alpha_1+\alpha_2+2\alpha_3=0$. The first equation implies that $\alpha_1=\alpha_2$. Substituting this in the second equation, we get $2\alpha_1+2\alpha_3=0$ or $\alpha_3=-\alpha_1$. Thus, any triple of the form $(x,x,-x)$ is a solution to the system of equations. Of course this can be verified by checking $x(1,-1,1)+x(1,1,1)+(-x)(2,0,2)=(0,0,0)$. The earlier solution was the special case when $x=1$.

Example 2.67. The vectors $(1,2,3)$, $(4,5,6)$, and $(13,14,15)$ are not linearly independent. To check linear independence, fix 3 arbitrary real numbers α_1, α_2, and α_3 and equate $(0,0,0)=\alpha_1(1,2,3)+\alpha_2(4,5,6)+\alpha_3(13,14,15)=(\alpha_1+4\alpha_2+13\alpha_3,2\alpha_1+5\alpha_2+14\alpha_3,3\alpha_1+6\alpha_2+15\alpha_3)$. Thus, we have the system of linear equations.

$$\begin{aligned}\alpha_1+4\alpha_2+13\alpha_3&=0\\2\alpha_1+5\alpha_2+14\alpha_3&=0\\3\alpha_1+6\alpha_2+15\alpha_3&=0\end{aligned}$$

Notice that $(\alpha_1,\alpha_2,\alpha_3)$ satisfies $\alpha_1+4\alpha_2+13\alpha_3=0$ iff it satisfies $2\alpha_1+8\alpha_2+26\alpha_3=0$.

Definition 2.68. Let $v_1,\dots,v_k\in\mathbb{R}^n$. We say $(v_1,\dots,v_k)$ is linearly dependent if they are not linearly independent. In other words, we say $(v_1,\dots,v_k)$ is linearly dependent if $\exists(\alpha_1,\alpha_2,\dots,\alpha_k)\in\mathbb{R}^k$, such that $(\alpha_1,\dots,\alpha_k)\neq(0,\dots,0)$ but $\alpha_1v_1+\cdots+\alpha_kv_k=0$.

Exercise 2.69. Check (using the definition) if the following collection of vectors in $\mathbb{R}^2$ are linearly independent or dependent

(1) $(2,2)$
(2) $(1,1),(1,-1)$
(3) $(1,2),(2,1)$
(4) $(1,2),(3,4),(5,6)$

Theorem 2.70. *Given any three vectors $v_1,v_2,v_3\in\mathbb{R}^2$, (v_1,v_2,v_3) is linearly dependent.*

Proof. If (v_1,v_2) is linearly dependent, then there exists $(\alpha_1,\alpha_2)\neq(0,0)$ such that $\alpha_1v_1+\alpha_2v_2=0$. Thus, $\alpha_1v_1+\alpha_2v_2+0.v_3=0$ and $(\alpha_1,\alpha_2,0)\neq(0,0,0)$. Therefore (v_1,v_2,v_3) is linearly dependent. If (v_1,v_2) is linearly independent, then by Theorem 1.47, $\text{Span}(v_1,v_2)=\mathbb{R}^2$. Thus, there exists $\alpha_1,\alpha_2\in\mathbb{R}$ such that $v_3=\alpha_1v_1+\alpha_2v_2$. Thus, $\alpha_1v_1+\alpha_2v_2+(-1)v_3=0$ and thus (v_1,v_2,v_3) is linearly dependent. □

Exercise 2.71. Check (using the definition) if the following collection of vectors in $\mathbb{R}^3$ are linearly independent or dependent

(1) $(1,0,1),(1,1,0),(1,1,0)$

(2) $(1,1,1),(1,1,-1),(-1,1,1)$

(3) $(1,0,0),(1,2,3),(5,6,9)$

(4) $(1,1,1),(1,2,1),(3,1,1)$

(5) $(1,1,1),(1,2,1),(1,3,1)$

(6) $(1,2,3),(2,1,3)$

Theorem 2.72. *If $(v_1, v_2, \dots, v_k)$ is linearly independent, then any subcollection $(v_{i_1}, v_{i_2}, \dots, v_{i_m})$ (where $1 \le i_1 \le i_2 \le \dots \le i_m \le k$) is also linearly independent.*

Proof (by contradiction). Assume $(v_{i_1}, v_{i_2}, \dots, v_{i_m})$ is linearly dependent, that is, $\exists(a_{i_1}, \dots, a_{i_m}) \neq 0$ such that $a_{i_1}v_{i_1} + \dots + a_{i_m}v_{i_m} = 0$. Define $a_j = 0$ if $j \notin \{i_1, i_2, \dots, i_m\}$. Then $(a_1, \dots, a_n) \neq 0$ but,

$$a_1v_1 + \dots a_kv_k = a_{i_1}v_{i_1} + \dots + a_{i_m}v_{i_m} = 0.$$

Thus, $(v_1, \dots, v_k)$ is linearly dependent. But, this is a contradiction. So, our assumption that $(v_{i_1}, v_{i_2}, \dots, v_{i_m})$ is linearly dependent should be incorrect. ☐

Combining the above result with Theorem 2.70, we get

Corollary 2.73. *Let $v_1, \dots, v_k \in \mathbb{R}^2$ and $k > 2$. Then, $(v_1, v_2, \dots, v_k)$ is linearly dependent.*

Lemma 2.74. *If $v \in Span(v_1, \dots, v_k)$, then $(v, v_1, \dots, v_k)$ is linearly dependent.*

Proof. As $v \in \text{Span}(v_1, \dots, v_k)$, there exists $(\alpha_1, \dots, \alpha_k) \in \mathbb{R}^k$ such that $v = \alpha_1v_1 + \dots + \alpha_kv_k$. Therefore, $(-1)v + \alpha_1v_1 + \dots + \alpha_kv_k = 0$ and therefore $(v, v_1, \dots, v_k)$ is linearly dependent. ☐

Theorem 2.75. *Let V be a subspace of $\mathbb{R}^n$. Let $v_i \in V$ for all $i \in \{1, \dots, k\}$ and let $w_j \in V$ for all $j \in \{1, \dots, l\}$. If $(v_1, \dots, v_k)$ are linearly independent and $Span(w_1, \dots, w_l) = V$, then $k \le l$.*[3]

This proof is very involved. So, we would first give the key ideas of the proof without getting into the details. Hopefully, keeping this idea in mind would help you appreciate the proof better.

Proof Sketch. We start with the list $(w_1, \dots, w_l)$ and adjoin the vector v_k to it to obtain the list $(v_k, w_1, \dots, w_l)$. We show $(v_k, w_1, \dots, w_l)$ is linearly dependent and therefore we may without loss of generality assume $\text{Span}(w_1, \dots, w_l) = \text{Span}(v_k, w_2, \dots, w_l)$. Thus, we replace the list $(w_1, \dots, w_l)$ with $(v_k, w_2, \dots, w_l)$. We repeat this process - at every stage, we add a v and drop a w. If we run out of w's before we run out of v's then a subcollection of v's will span V. This would contradict the fact that $(v_1, \dots, v_k)$ is linearly independent. Thus, there should be at least as many w's as there are v's. ☐

Proof. As $\text{Span}(w_1, \dots, w_l) = V$ and $v_k \in V$, by Lemma 2.74 we know that $(v_k, w_1, \dots, w_l)$ is linearly dependent. Thus, there exists $(a_k, b_1, \dots, b_l) \neq 0$ such that $a_kv_k + b_1w_1 + \dots + b_lw_l = 0$. Notice if $(b_1, \dots, b_l) = 0$, then $a_k \neq 0$ but $0 = a_kv_k + b_1w_1 + \dots + b_lw_l = a_kv_k$. Which implies that v_k is the 0 vector. But then $0v_1 + \dots + 0v_{k-1} + 1v_k = 0$. This contradicts our assumption that $(v_1, \dots, v_k)$ is linearly independent. So, $(b_1, \dots, b_l) \neq 0$. If $l = 1$, then $V = \text{Span}(w_1)$ and $b_1 \neq 0$. Thus, $w_1 = -\frac{a_k}{b_1}v_k$. Hence $\text{Span}(v_k) = \text{Span}(w_1) = V$. As $(v_1, \dots, v_k)$ is linearly independent, this would imply that $k = 1$. Else, $v_1 \in \text{Span}(v_k)$, which means $v_1 = \alpha_kv_k$. Then, $(-1)v_1 + 0v_2 + \dots + 0v_{k-1} + \alpha_kv_k = 0$ and $(v_1, \dots, v_k)$ is linearly dependent - a contradiction. Thus, if $l = 1$, then $k = 1$ and we have $k \le l$. So, we may assume $l > 1$. If $k = 1$, then $k < l$. Thus, we

[3] The statement of this theorem and its proof is from [Axler]. The proof is however significantly elaborated.

may assume that $k > 1$. Without loss of generality (Notice that span does not depend on the order of the vectors. Thus, if needed we can renumber $w_1, \dots, w_l$) we may assume that $b_1 \neq 0$. Thus,

$$w_1 = \left(-\frac{a_k}{b_1}\right) v_k + \left(-\frac{b_2}{b_1}\right) w_2 + \left(-\frac{b_3}{b_1}\right) w_3 + \cdots + \left(-\frac{b_l}{b_1}\right) w_l.$$

Thus, $V = \text{Span}(w_1, \dots, w_l) \subset \text{Span}(v_k, w_2, \dots, w_l)$. But as $\{v_k, w_2, \dots, w_l\} \subset V$, we have $\text{Span}(v_k, w_2, \dots, w_l) \subset V$. That is $\text{Span}(v_k, w_2, \dots, w_l) = V$.

Now, As $\text{Span}(v_k, w_2, \dots, w_l) = V$ and $v_{k-1} \in V$, by Lemma 2.74 $(v_{k-1}, v_k, w_2, \dots, w_l)$ is linearly dependent. Thus, there exists $(a_{k-1}, a_k, b_2, \dots, b_l) \neq 0$ such that $a_{k-1}v_{k-1} + a_k v_k + b_2 w_2 + \cdots + b_l w_l = 0$. Notice if $(b_2, \dots, b_l) = 0$, then $(a_{k-1}, a_k) \neq 0$ but $0 = a_{k-1}v_{k-1} + a_k v_k + b_1 w_1 + \cdots + b_l w_l = a_{k-1}v_{k-1} + a_k v_k$. But then $0v_1 + \cdots + 0v_{k-2} + a_{k-1}v_{k-1} + a_k v_k = 0$. This contradicts our assumption that $(v_1, \dots, v_k)$ is linearly independent. So, $(b_2, \dots, b_l) \neq 0$. If $l = 2$, then $V = \text{Span}(v_k, w_2)$ and $b_2 \neq 0$. Thus, $w_2 = \left(-\frac{a_{k-1}}{b_2}\right) v_{k-1} + \left(-\frac{a_k}{b_2}\right) v_k$. Thus, $V = \text{Span}(v_k, w_2) \subset \text{Span}(v_{k-1}, v_k)$. But, as $\{v_{k-1}, v_k\} \subset V$, $\text{Span}(v_{k-1}, v_k) = V$. As $(v_1, \dots, v_k)$ is linearly independent, this would imply that $k = 2$. Else, $v_1 \in \text{Span}(v_{k-1}, v_k)$, which means $v_1 = \alpha_{k-1}v_{k-1} + \alpha_k v_k$. Then, $(-1)v_1 + 0v_2 + \cdots + 0v_{k-2} + \alpha_{k-1}v_{k-1} + \alpha_k v_k = 0$ and $(v_1, \dots, v_k)$ is linearly dependent - a contradiction. Thus, if $l = 2$ and $k > 1$, then $k = 2$ and we have $k \leq l$. So, we may assume $l > 2$. If $k = 2$, then $k < l$ and we are done. So, we may assume $k > 2$. Without loss of generality (by renumbering $w_2, \dots, w_l$) we may assume that $b_2 \neq 0$. Thus,

$$w_2 = \left(-\frac{a_{k-1}}{b_2}\right) v_{k-1} + \left(-\frac{a_k}{b_2}\right) v_k + \left(-\frac{b_3}{b_2}\right) w_3 + \left(-\frac{b_4}{b_2}\right) w_4 + \cdots + \left(-\frac{b_l}{b_2}\right) w_l.$$

Thus, $V = \text{Span}(v_k, w_2, \dots, w_l) \subset \text{Span}(v_{k-1}, v_k, w_3, \dots, w_l)$. But as $\{v_{k-1}, v_k, w_3, \dots, w_l\} \subset V$, we have $\text{Span}(v_{k-1}, v_k, w_3, \dots, w_l) \subset V$. That is $\text{Span}(v_{k-1}, v_k, w_3, \dots, w_l) = V$.

We can continue this process and assume $l > i$ and $k > i$. As $\text{Span}(v_{k-i+1}, \dots, v_k, w_{i+1}, \dots, w_l) = V$ and $v_{k-i} \in V$, by Lemma 2.74 $(v_{k-i}, \dots, v_k, w_{i+1}, \dots, w_l)$ is linearly dependent. Thus, there exists $(a_{k-i}, \dots, a_k, b_{i+1}, \dots, b_l) \neq 0$ such that $a_{k-i}v_{k-i} + \cdots + a_k v_k + b_{i+1}w_{i+1} + \cdots + b_l w_l = 0$. Notice if $(b_{i+1}, \dots, b_l) = 0$, then $(a_{k-i}, \dots, a_k) \neq 0$ but $0 = a_{k-i}v_{k-i} + \cdots + a_k v_k + b_{i+1}w_{i+1} + \cdots + b_l w_l = a_{k-i}v_{k-i} + \cdots + a_k v_k$. But then $0v_1 + \cdots + 0v_{k-i-1} + a_{k-i}v_{k-i} + \cdots + a_k v_k = 0$. This contradicts our assumption that $(v_1, \dots, v_k)$ is linearly independent. So, $(b_{i+1}, \dots, b_l) \neq 0$. If $l = i + 1$, then $\text{Span}(v_{k-i+1}, \dots, v_k, w_l) = V$ and $b_l \neq 0$. Thus,

$$w_l = \left(-\frac{a_{k-i}}{b_l}\right) v_{k-i} + \cdots + \left(-\frac{a_k}{b_l}\right) v_k.$$

That is $V = \text{Span}(v_{k-i+1}, \dots, v_k, w_l) \subset \text{Span}(v_{k-i}, \dots, v_k)$. As $\{v_{k-i}, \dots, v_k\} \subset V$, we have $\text{Span}(v_{k-i}, \dots, v_k) = V$. If $k - i > 1$, then $v_1 \in \text{Span}(v_{k-i}, \dots, v_k) = V$ and hence $(v_1, \dots, v_k)$ is linearly dependent. Therefore $k - i \leq 1$, that is $k \leq i + 1 = l$. **Thus, we will not run out of w's before v's**.

Thus, the above process works for each i. When $i = k-1$, the process terminates and the proof becomes complete. Thus, we have the result. □

Corollary 2.76. *If $(v_1, \dots, v_k)$ is a linearly independent list of vectors in $\mathbb{R}^n$, then $k \leq n$.*

Proof. Note that $\text{Span}(e_1, \dots, e_n) = \mathbb{R}^n$. Thus, $k \leq n$. □

Corollary 2.77. *If Span$(v_1, \dots, v_k) = \mathbb{R}^n$, then $k \geq n$.*

Proof. Note that $(e_1, \dots, e_n)$ is linearly independent. Thus, $k \geq n$. □

2.6. Basis and Dimension

Definition 2.78. Let V be a subspace of $\mathbb{R}^n$, then a list of vectors $(v_1, \ldots, v_k)$ is called a basis if

(1) $\text{Span}(v_1, \ldots, v_k) = V$
(2) $(v_1, \ldots, v_k)$ is linearly independent.

Example 2.79. The list $((1,0),(0,1))$ is a basis of $\mathbb{R}^2$. The list $((1,1),(1,-1))$ is also a basis of $\mathbb{R}^2$. The list $((1,0),(0,1),(1,1))$ is not a basis as it is not linearly independent.

Example 2.80. Let e_i be the vector in $\mathbb{R}^n$ whose i-th entry is 1 and all other entries are 0. Then $(e_1, \ldots, e_n)$ is a basis for $\mathbb{R}^n$. Notice that we will often denote $(1,0) \in \mathbb{R}^2$ and $(1,0,0) \in \mathbb{R}^3$ by e_1. In most situations this will not lead to any confusion instead it will aid in clarity.

Example 2.81. Let $V = \text{Span}((1,2,3),(4,5,6),(13,14,15))$. We will start by checking if the list $((1,2,3),(4,5,6),(13,14,15))$ is linearly independent. Assume $\alpha(1,2,3)+\beta(4,5,6)+\gamma(13,14,15) = (0,0,0)$. That is,

$$\begin{aligned} \alpha + 4\beta + 13\gamma &= 0 \\ 2\alpha + 5\beta + 14\gamma &= 0 \\ 3\alpha + 6\beta + 15\gamma &= 0 \end{aligned}$$

As $\alpha + 4\beta + 13\gamma = 0$, $2\alpha + 8\beta + 26\gamma = 0$. Thus, $0 = (2\alpha+8\beta+26\gamma)-(2\alpha+5\beta+14\gamma) = 3\beta+12\gamma$. That is, $\beta = -4\gamma$. Substituting this value of β in the equation $3\alpha + 6\beta + 15\gamma = 0$, we get $0 = 3\alpha + 6((-4)\gamma) + 15\gamma = 3\alpha - 9\gamma$. That is, $\alpha = 3\gamma$. Taking $\gamma = 1$, we can check that $3(1,2,3) + (-4)(4,5,6) + (13,14,15) = (0,0,0)$. In other words, $(13,14,15) = (-3)(1,2,3) + 4(4,5,6)$. Thus, $(13,14,15) \in \text{Span}((1,2,3),(4,5,6))$. Therefore,

$$\text{Span}((1,2,3),(4,5,6)) = \text{Span}((1,2,3),(4,5,6),(13,14,15)) = V.$$

Now, let us check if $((1,2,3),(4,5,6))$ is linearly independent. Assume $\alpha(1,2,3) + \beta(4,5,6) = (0,0,0)$. That is

$$\begin{aligned} \alpha + 4\beta &= 0 \\ 2\alpha + 5\beta &= 0 \\ 3\alpha + 6\beta &= 0 \end{aligned}$$

As $\alpha + 4\beta = 0$, $2\alpha + 8\beta = 0$. Thus, $0 = (2\alpha + 8\beta) - 2\alpha + 5\beta = 3\beta$. Therefore $\beta = 0$ and hence $\alpha = 0$. That is, $((1,2,3),(4,5,6))$ is linearly independent. Thus, $((1,2,3),(4,5,6))$ is a basis for V.

Example 2.82. Let $V = \text{Span}((1,2,3),(4,5,6),(13,14,13))$. We will start by checking if the list $((1,2,3),(4,5,6),(13,14,13))$ is linearly independent. Assume $\alpha(1,2,3)+\beta(4,5,6)+\gamma(13,14,13) = (0,0,0)$. That is,

$$\begin{aligned} \alpha + 4\beta + 13\gamma &= 0 \\ 2\alpha + 5\beta + 14\gamma &= 0 \\ 3\alpha + 6\beta + 13\gamma &= 0 \end{aligned}$$

Thus, $0 = (3\alpha+6\beta+13\gamma)-(\alpha+4\beta+13\gamma) = 2\alpha+2\beta$. That is, $\beta = -\alpha$. Substituting this value of β in $2\alpha + 5\beta + 14\gamma = 0$, we get $0 = 2\alpha - 5\alpha + 14\gamma = (-3)\alpha + 14\gamma$. That is $\gamma = \frac{3}{14}\alpha$. Substituting these values of β and γ in $\alpha + 4\beta + 13\gamma = 0$, we get $0 = \alpha - 4\alpha + \frac{39}{14}\alpha = -\frac{3}{14}\alpha$. Thus, $\alpha = 0$ and hence $(\alpha, \beta, \gamma) = \left(\alpha, -\alpha, \frac{3}{14}\alpha\right) = (0,0,0)$. Therefore, $((1,2,3),(4,5,6),(13,14,13))$ is a basis of V.

Example 2.83. Let $V = \{(x,y,z) \in \mathbb{R}^3 \mid 2x+3y=0\}$. Then,

$$\begin{aligned} V &= \left\{ \left(x, -\frac{2}{3}x, z\right) \mid (x,z) \in \mathbb{R}^2 \right\} \\ &= \left\{ x\left(1, -\frac{2}{3}, 0\right) + z(0,0,1) \mid (x,z) \in \mathbb{R}^2 \right\} \\ &= \operatorname{Span}\left(\left(1, -\frac{2}{3}, 0\right), (0,0,1) \right) \end{aligned}$$

It is easy to see that $\left(\left(1,-\frac{2}{3},0\right),(0,0,1)\right)$ is linearly independent as $\alpha\left(1,-\frac{2}{3},0\right)+\beta(0,0,1)=0$ clearly implies $(\alpha,\beta)=(0,0)$. Thus, $\left(\left(1,-\frac{2}{3},0\right),(0,0,1)\right)$ is a basis for V.

Theorem 2.84. *If $(v_1,\ldots,v_k)$ and $(w_1,\ldots,w_l)$ are both basis of a subspace V of $\mathbb{R}^n$, then $k=l$.*

Proof. As $(v_1,\ldots,v_k)$ is a basis, $(v_1,\ldots,v_k)$ is linearly independent. And, as $(w_1,\ldots,w_l)$ is a basis, $\operatorname{Span}(w_1,\ldots,w_l)=V$. Thus, by Theorem 2.75, $k \le l$. However, as $(v_1,\ldots,v_k)$ is a basis, $\operatorname{Span}(v_1,\ldots,v_k)=V$. And, as $(w_1,\ldots,w_l)$ is a basis, $(w_1,\ldots,w_l)$ is linearly independent. Thus, by Theorem 2.75, $l \le k$. Hence $k=l$. □

Thus, any two basis of a subspace $V \subset \mathbb{R}^n$ has the same cardinality. But, can we find a basis for any subspace of $\mathbb{R}^n$?

Theorem 2.85. *Given any subspace $V \neq \{0\}$ of $\mathbb{R}^n$, we can find a basis for V.*

Proof. As $V \neq \{0\}$, there exists at least one non-zero vector $v_1 \in V$. If $\operatorname{Span}(v_1)=V$, then v_1 is a basis for V. If not, let $v_2 \in V \setminus \operatorname{Span}(v_1)$ be an arbitrary element. Then (v_1,v_2) is linearly independent. If $\operatorname{Span}(v_1,v_2)=V$, then (v_1,v_2) is a basis for V if not let $v_3 \in V \setminus \operatorname{Span}(v_1,v_2)$ be an arbitrary element. Then (v_1,v_2,v_3) is linearly independent. If $\operatorname{Span}(v_1,v_2,v_3)=V$, then (v_1,v_2,v_3) is a basis for V if not let $v_4 \in V \setminus \operatorname{Span}(v_1,v_2,v_3)$ be an arbitrary element. Then (v_1,v_2,v_3,v_4) is linearly independent. We can keep repeating this process. If $(v_1,\ldots,v_k)$ is linearly independent and $\operatorname{Span}(v_1,\ldots,v_k) \neq V$, then we can take an arbitrary vector $v_{k+1} \in V \setminus \operatorname{Span}(v_1,\ldots,v_k)$. But, thus process has to stop as otherwise we will obtain a list $(v_1,\ldots,v_{n+1})$ that is linearly independent contradicting Corollary 2.76. □

Hidden in the above proof is the following useful fact

Theorem 2.86. *Let V be a subspace of $\mathbb{R}^n$ and let $(v_1,\ldots,v_l)$ be a linearly independent list of vectors in V. Then, there exists additional vectors $v_{l+1},\ldots,v_k$ such that $(v_1,\ldots,v_k)$ is a basis of V.*

Exercise 2.87. Prove the above theorem.

Similarly, we have

Theorem 2.88. *Let V be a subspace of $\mathbb{R}^n$ and let $Span(v_1,\ldots,v_l)=V$. Then, there exists a subcollection $(v_{i_1},\ldots,v_{i_k})$ where $1 \le v_{i_1} < \cdots < v_{i_k} \le v_l$ such that $(v_{i_1},\ldots,v_{i_k})$ is a basis of V.*

Proof. Let m be the smallest natural number such that $v_m \neq 0$. Then, $\operatorname{Span}(v_m,\ldots,v_l) = \operatorname{Span}(v_1,\ldots,v_l)=V$. Thus, we can throw away $v_1,\ldots,v_{m-1}$ and renumber the vectors. In other words, we may assume without loss of generality that $v_1 \neq 0$. Thus, start with the list $\mathcal{B}=(v_1)$. If $v_2 \notin \mathcal{B}$, then add v_2 to $\mathcal{B}$ (by convention we add to the right, that is, $\mathcal{B}=(v_1,v_2)$). Else, leave $\mathcal{B}$ as it is. If $v_3 \notin \mathcal{B}$, then add v_3 to $\mathcal{B}$. Else, leave $\mathcal{B}$ as it is. Repeat these steps n times and we have the required list. □

Definition 2.89. Let V be a subspace $\mathbb{R}^n$. If $V = \{0\}$, we say the dimension of V is 0. If $V \neq \{0\}$, then the dimension of V is the cardinality of any of its basis.

Example 2.90. As $(e_1, \ldots, e_n)$ is a basis for $\mathbb{R}^n$, the dimension of $\mathbb{R}^n$ is n.

Example 2.91. Let $V = \text{Span}((1,2,3),(4,5,6),(13,14,15))$. As $((1,2,3),(4,5,6))$ is a basis for V, dimension of V is 2.

Example 2.92. Let $V = \text{Span}((1,2,3),(4,5,6),(13,14,13))$. As, $((1,2,3),(4,5,6),(13,14,13))$ is a basis of V, dimension of V is 3.

Example 2.93. Let $V = \{(x,y,z) \in \mathbb{R}^3 \,|2x+3y=0\}$. As. $\left(\left(1,-\frac{2}{3},0\right),(0,0,1)\right)$ is a basis for V, dimension of V is 2

Exercise 2.94. In the following questions, find the dimension of the nullspace/kernel and range/image (by explicitly finding a $(v_1, \ldots, v_k)$ and showing that $(v_1, \ldots, v_k)$ are linearly independent and $\text{Span}(v_1, \ldots, v_k) = V$.). Check that the sum of the dimensions of the nullspace and the range add up to the dimension of the domain.

i) $f : \mathbb{R}^2 \to \mathbb{R}^3$ defined as $f(x,y) = (x,y,x-y)$.
ii) $f : \mathbb{R}^3 \to \mathbb{R}^3$ defined as $f(x,y,z) = (x+4y+13z, 2x+5y+14z, 3x+6y+15z)$.
iii) $f : \mathbb{R}^3 \to \mathbb{R}^2$ defined as $f(x,y,z) = (2x+3y, z)$.
iv) $f : \mathbb{R}^3 \to \mathbb{R}^4$ defined as $f(x,y,z) = (x+y, y+z, x+z, x+y+z)$.

We will now show that the observation in the above exercise was no coincidence.

Theorem 2.95. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear function. Then $dim(Ker(F)) + dim(Im(F)) = n$.*

Proof. We know that $Ker(F)$ is a subspace of $\mathbb{R}^n$. Thus, by Theorem 2.85. $Ker(F)$ has some basis $(v_1, \ldots, v_k)$. Thus, $(v_1, \ldots, v_k)$ is linearly independent. Therefore, by Theorem 2.86, we can extend $(v_1, \ldots, v_k)$ to a list $(v_1, \ldots, v_n)$ which form a basis of $\mathbb{R}^n$. Let $F(v_{k+1}) = w_{k+1}$, $F(v_{k+2}) = w_{k+2}, \ldots$ $F(v_n) = w_n$. As $(v_1, \ldots, v_n)$ is a basis of $\mathbb{R}^n$, given any vector $v \in \mathbb{R}^n$ there exists $(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. As F is linear,

$$F(v) = \alpha_1 F(v_1) + \cdots + \alpha_n F(v_n) = \alpha_1 0 + \ldots \alpha_k 0 + \alpha_{k+1} w_{k+1} + \alpha_n w_n = \alpha_{k+1} w_{k+1} + \alpha_n w_n.$$

As v was arbitrary, this implies that $Im(F) = \text{Span}(w_{k+1}, \ldots, w_n)$. Moreover, if $\alpha_{k+1} w_{k+1} + \cdots + \alpha_n w_n = 0$, then, $F(\alpha_{k+1} v_{k+1} + \cdots + \alpha_n v_n) = \alpha_{k+1} w_{k+1} + \alpha_n w_n = 0$. That is $\alpha_{k+1} v_{k+1} + \cdots + \alpha_n v_n \in Ker(F)$. As $(v_1, \ldots, v_k)$ is a basis of $Ker(F)$, there exists $(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ such that $\alpha_1 v_1 + \cdots + \alpha_k v_k = \alpha_{k+1} v_{k+1} + \cdots + \alpha_n v_n$. In other words, $\alpha_1 v_1 + \cdots + \alpha_k v_k + (-\alpha_{k+1}) v_{k+1} + \cdots + (-\alpha_n) v_n = 0$. As $(v_1, \ldots, v_n)$ is linearly independent, this would imply that $\alpha_i = 0$ for all $i \in \{1, \ldots, n\}$. Thus, in particular $(\alpha_{k+1}, \ldots, \alpha_n) = 0$. Hence, $(w_{k+1}, \ldots, w_n)$ is linearly independent. Therefore, $(w_{k+1}, \ldots, w_n)$ is a basis for $Im(F)$. Thus, $dim(Ker(F)) + dim(Im(F)) = k + (n-k) = n$.

□

Abstract vector spaces

In the previous chapter, we saw the importance of linear functions. To define linearity, we need addition and scalar multiplication on both domain and co-domain. Of course, we do not want these operations to be defined extremely weirdly, so we expect them to satisfy certain properties. This motivates the definition of vector spaces which are the natural domains and co-domains of linear functions.

Definition 3.1. Given a set V and functions $+ : V \times V \to V$ and $. : \mathbb{R} \times V \to V$, the triple $(v, +, .)$ where is called a vector space over $\mathbb{R}$ if it satisfies the following properties.

(1) $\forall v_1, v_2, v_3 \in V$, $v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$ (Associativity of vector addition)

(2) $\forall v_1, v_2 \in V$, $v_1 + v_2 = v_2 + v_1$ (Commutativity of vector addition)

(3) $\exists 0 \in V$ such that $v + 0 = v = 0 + v$ (Existence of additive identity)

(4) $\forall v \in V$, $\exists v' \in$V such that $v + v' = 0 = v' + v$ (Existence of additive inverse)

(5) $\forall \alpha, \beta, \gamma \in \mathbb{R}$ and $\forall v \in V$, $(\alpha\beta).v = \alpha(\beta.v)$ (Associativity of scalar multiplication)

(6) $\forall \alpha \in \mathbb{R}$ and $v_1, v_2 \in V$, $\alpha.(v_1 + v_2) = \alpha.v_1 + \alpha.v_2$

(7) $\forall \alpha, \beta \in \mathbb{R}$ and $v \in V$, $(\alpha + \beta).v = \alpha.v + \beta.v$.

(8) $\forall v \in V$, $1.v = v$.

Example 3.2. Let $V = \{0\}$, $0 + 0 = 0$ and $\forall \alpha \in \mathbb{R}, \alpha.0 = 0$. Then, $(V, +, .)$ is a vector space over $\mathbb{R}$.

Example 3.3. For all vectors $(v_1, \dots, v_n), (w_1, \dots, w_n) \in \mathbb{R}^n$ define $(v_1, \dots, v_n) + (w_1, \dots, w_n) = (v_1 + w_1, \dots, v_n + w_n)$. Similarly, for all $\alpha \in \mathbb{R}$ and $(v_1, \dots, v_n) \in \mathbb{R}^n$ define $\alpha.(v_1, \dots, v_n) = (\alpha v_1, \dots, \alpha v_n)$. Then, $(\mathbb{R}^n, +, .)$ is a vector space over $\mathbb{R}$.

Example 3.4. Let X be any set and let $\mathcal{F}(X, \mathbb{R})$ be the set of all functions from X to $\mathbb{R}$. We can define $+$ as follows. Given two functions $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$, we define $f + g : X \to \mathbb{R}$ as the function $(f + g)(x) = f(x) + g(x)$. Similarly, if $f : X \to \mathbb{R}$ is a function and $\alpha \in \mathbb{R}$, then we define αf to be the function defined as $(\alpha f)(x) = \alpha.(f(x))$. Then, $(\mathcal{F}(X, \mathbb{R}), +, .)$ is a vector space over $\mathbb{R}$.

Notice, that if you take $X = \mathbb{N}$, then you get the set of all real sequences.

Example 3.5. Let $\mathcal{P}$ be the set of all polynomials with real coefficients. Given two polynomials $a_0 + \cdots + a_n x^n$ and $b_0 + \cdots + b_m x^m$ we may assume without loss of generality that $m \geq n$. Recall from school that the sum of these two polynomials is the polynomial $(a_0 + b_0) + \dots (a_n + b_n)x^n + b_{n+1}x^{n+1} + \dots b^n x^n$. Given a polynomial $a_0 + \dots + a_n x^n$ and $\alpha \in \mathbb{R}$, the scalar multiplication is defined as $\alpha.(a_0 + \dots + a_n x^n) = (\alpha a_0) + \dots (\alpha a_n)x^n$. Then $(\mathcal{P}, +, .)$ is a vector space over $\mathbb{R}$.

Example 3.6. Let $\mathcal{P}(n)$ be the set of polynomials (with real coefficients) whose degree is less than or equal to n. Notice that any polynomial in $\mathcal{P}(n)$ is of the form $a_0 + \dots + a_n x^n$. As in the previous example, we can define $(a_0 + \dots + a_n x^n) + (b_0 + \dots + b_n x^n) = (a_0 + b_0) + \dots (a_n + b_n)x^n$ and $\alpha.(a_0 + \dots + a_n x^n) = (\alpha a_0) + \dots (\alpha a_n)x^n$. Then $(\mathcal{P}(n), +, .)$ is a vector space over $\mathbb{R}$.

Example 3.7. If V is a subspace of $\mathbb{R}^n$, then it inherits addition and scalar multiplication from $\mathbb{R}^n$ making $(V, +, .)$ a vector space over $\mathbb{R}$.

Exercise 3.8. Show that all the above examples are indeed examples of vector spaces over $\mathbb{R}$.

More generally,

Definition 3.9. Let $(V, +, .)$ is a vector space. A subset W of V is called a subspace if

(1) $0 \in W$
(2) $\forall v, w \in W$, $v + w \in W$
(3) $\forall \alpha \in \mathbb{R}$ and $\forall w \in W$, $\alpha w \in W$

Example 3.10. Let $(V, +, .)$ is a vector space and $W \subset V$. Then $(W, +, .)$ is a vector space iff W is a subspace of V.

Example 3.11. We saw earlier that $(\mathcal{F}(X, \mathbb{R}), +, .)$ is a vector space for all sets X. In particular, $(\mathcal{F}(\mathbb{R}, \mathbb{R}), +, .)$ is a vector space. If $\mathcal{C}^0(\mathbb{R}, \mathbb{R})$ denotes the set of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$ then $\mathcal{C}^0(\mathbb{R}, \mathbb{R})$ is a subspace of $(\mathcal{F}(X, \mathbb{R}), +, .)$. Similarly, if $\mathcal{C}^1(\mathbb{R}, \mathbb{R})$ denotes the set of all continuously differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ then $\mathcal{C}^1(\mathbb{R}, \mathbb{R})$ is a subspace of $(\mathcal{F}(X, \mathbb{R}), +, .)$.

Example 3.12. The set $\mathcal{P}(n)$ is a subspace of $(\mathcal{P}, +, .)$.

Exercise 3.13. Show that the intersection of two vector subspaces is a vector subspace.

Exercise 3.14. Prove that the union of two subspaces of V is a subspace of V if and only if one of the subspaces is contained in the other.

Exercise 3.15. Let p be a polynomial. What extra condition should p satisfy so that

$$\{(x_1, x_2, \dots, x_n) |\ p(x_1, x_2, \dots, x_n) = 0\} \subset \mathbb{R}^n$$

is a subspace? Prove your claim.

Redundancy of axioms

Mathematicians often like to have a minimal set of axioms. Famously, many mathematicians thought that Euclid's fifth postulate was redundant, that is, it was a consequence of the other four postulates. It was proved much later that the fifth postulate is not a consequence of the other four - we can come up with "geometries" where the first four axioms are satisfied, but the fifth axiom is not satisfied. This was the birth of hyperbolic and spherical geometry.

Similarly, one may ask if axioms of vector spaces are essential. Here I reproduce the proof in [**Bryant**] of the fact that Axiom 2 is redundant. If $v, w \in V$, then notice that $(v + w) + (v + w) = (1 + 1)(v + w) = (1 + 1)v + (1 + 1)w = (v + v) + (w + w)$. Let v' be the inverse of v and w' be the inverse of w. Then,

$$\begin{aligned}(v' + ((v+w) + (v+w))) + w' &= ((v' + (v+w)) + (v+w)) + w' \\ &= (((v'+v) + w) + (v+w)) + w' \\ &= ((0+w) + (v+w)) + w' = (w + (v+w)) + w' \\ &= w + ((v+w) + w') = w + (v + (w+w')) = w + (v+0) \\ &= w + v.\end{aligned}$$

And,

$$\begin{aligned}(v' + ((v+v) + (w+w))) + w' &= ((v' + (v+v)) + (w+w)) + w' \\ &= (((v'+v) + v) + (w+w)) + w' \\ &= ((0+v) + (w+w)) + w' = (v + (w+w)) + w' \\ &= v + ((w+w) + w') = v + (w + (w+w')) = v + (w+0) \\ &= v + w.\end{aligned}$$

Exercise 3.16. Justify each equality in the above two computations.

But, $(v'+((v+w)+(v+w)))+w' = (v'+((v+v)+(w+w)))+w'$ and therefore $w+v = v+w$. Notice that we used only associativity, existence of additive identity and existence of additive inverse. Thus, commutativity is redundant.

In [**Rigby-Wiegold**] the authors show that actually you need only 6 axioms. I believe, according to their definition, the empty set would become a vector field. But, with a small modification, we can solve this issue. Basically, we can replace Axiom 3 and Axiom 4 with a single axiom: $\exists$ **a vector** $0 \in V$ **such that** $\forall v \in V$**,** $0.v = 0$. This vector is the additive identity as $v + 0 = 1.v + 0.v = (1+0).v = 1.v = v$. Also, $(-1)v$ is the additive inverse as $v + (-1)v = (1 + (-1)).v = 0.v = 0$. Also, we cannot just keep our current "Existence of additive identity" as it is and delete "Existence of additive inverse". Because, then although v + (-1)v = 0.v, we cannot conclude that is the zero vector.

It is still unclear to me if the other axioms are redundant or not. It is a very interesting exercise to come up with examples that satisfy all but one axiom. Some interesting examples for vector spaces over a field (not necessarily $\mathbb{R}$) is provided in [**user7530**].

Exercise 3.17. Let $V = \{(x,0) \mid x \in \mathbb{R}\} \cup \{(0,y) \mid y \in \mathbb{R}\}$. Define $+$ as

(1) $\forall x, x' \in \mathbb{R}$, $(x,0) + (x',0) = (x+x',0)$
(2) $\forall y, y' \in \mathbb{R}$, $(0,y) + (0,y') = (0,y+y')$
(3) $\forall x, y \in \mathbb{R} \setminus \{0\}$, $(x,0) + (0,y) = (0,0)$

Similarly, define . as $\forall \alpha \in \mathbb{R}$ and $\forall x \in \mathbb{R}$, $\alpha.(x,0) = (\alpha x, 0)$ and $\alpha.(0,x) = (0,\alpha x)$. Which of the axioms of a vector space is violated by $(V,+,.)$?

Exercise 3.18. Let $V = \emptyset$. Let $+$ be the empty function - the function corresponding to the set $\emptyset = \emptyset \times \emptyset$. Let . be the empty function - the function corresponding to $\emptyset = \mathbb{R} \times \emptyset$. Which of the axioms of a vector space is violated by $(V,+,.)$?

Exercise 3.19. Take $V = [0,\infty)$, $+$ be **multiplication** and let . be defined as $\alpha.x = x^{\alpha}$ if $x \neq 0$ and $\alpha.0 = 0$. Which of the axioms of a vector space is violated by $(V,+,.)$? This example is from [**user7530**].

Exercise 3.20. Let $V = \mathbb{R}$ and let $+$ be the usual addition on $\mathbb{R}$. Let $.$ be defined as $\alpha.v = \alpha v$ for all $\alpha \in \mathbb{Q}$ and $\alpha.v = 0$ for all $\alpha \notin \mathbb{Q}$. Which of the axioms of a vector space is violated by $(V, +, .)$?

Exercise 3.21. Let $V = \{0, 1\}$. Define $+$ as $0 + 0 = 0 = 1 + 1$ and $0 + 1 = 1 = 1 + 0$. Define $.$ as $\alpha.0 = 0$ for all α, $\alpha.1 = 1$ for $\forall \alpha \neq 0$ and $0.1 = 0$. Which of the axioms of a vector space is violated by $(V, +, .)$?

Exercise 3.22. Let $V = \{0, 1\}$. Define $+$ as $0 + 0 = 0 = 1 + 1$ and $0 + 1 = 1 = 1 + 0$. Define $.$ as $\alpha.v = 0$ for all $\alpha \in \mathbb{R}$ and $v \in V$. Which of the axioms of a vector space is violated by $(V, +, .)$?

Exercise 3.23. Let $V = \mathbb{R}^2$, let $+$ be the usual addition on $\mathbb{R}^2$ and let $.$ be defined as $\alpha.(x, y) = (\alpha x, 0)$. Which of the axioms of a vector space is violated by $(V, +, .)$?[1]

3.1. Important properties of vector spaces

Throughout this section, $(V, +, .)$ is an arbitrary vector space.

Theorem 3.24. *Additive identity is unique. That is, there exists only one vector $w \in V$ such that $\forall v \in V$, $v + w = v = w + v$.*

Proof. Let 0 and $0'$ be such that $\forall v \in V$. $v + 0 = v = 0 + v$ and $v + 0' = v = 0' + v$. Taking $v = 0$, the equality $v + 0' = v = 0' + v$ gives us $0 = 0 + 0'$. But, taking $v = 0'$, the equality $v + 0 = v = 0 + v$ gives us $0 + 0' = 0'$. Thus, $0 = 0 + 0' = 0'$. In other words, any two additive identities have to be equal. So, the additive identity is unique. □

Theorem 3.25. *Additive inverse is unique. That is, given any vector v, there exists a unique w such that $v + w = 0 = w + v$.*

Proof. Assume the contrary. Assume there exists vector u, v, w such that $u + v = 0 = v + u$ and $u + w = 0 = w + u$ (v and w are both inverses of u). Then,

$$v = v + 0 = v + (u + w) = (v + u) + w = 0 + w = w$$

In other words, any two additive inverses have to be equal. So, additive inverse is unique. □

Theorem 3.26. *Let $\alpha \in \mathbb{R}$ and $v \in V$. Then, $\alpha.v = 0$ iff $\alpha = 0$ or $v = 0$.*

Proof. First note that $0.v = (0 + 0).v = 0.v + 0.v$. Let w be the additive inverse of $0.v$. Then, $0 = w + 0.v = w + (0.v + 0.v) = (w + 0.v) + 0.v = 0 + 0.v = 0.v$. Thus, $0.v = 0$. Similarly, $\alpha.0 = \alpha.(0 + 0) = \alpha.0 + \alpha.0$. Let w be the additive inverse of $\alpha.0$. Then, $0 = w + (\alpha.0) = w + (\alpha.0 + \alpha.0) = (w + \alpha.0) + \alpha.0 = 0 + \alpha.0 = \alpha.0$. Thus, $\alpha.0 = 0$. Thus, if $\alpha = 0$ or $v = 0$, then $\alpha.v = 0$.

Now, let us prove the converse - if $\alpha.v = 0$, then $\alpha = 0$ or $v = 0$. In other words, we will assume $\alpha.v = 0$ and prove $\alpha = 0$ or $v = 0$. We will do this by assuming $\alpha \neq 0$ and proving $v = 0$. In other words, we have $\alpha.v = 0$ and $\alpha \neq 0$. As $\alpha \neq 0$, $\frac{1}{\alpha} \in \mathbb{R}$. Thus,

$$v = 1.v = \left(\frac{1}{\alpha}\alpha\right).v = \frac{1}{\alpha}.(\alpha.v) = \frac{1}{\alpha}.0 = 0.$$

□

Theorem 3.27. *Let $(V, +, .)$ be a vector space. Then, V is singleton or uncountable.*

[1]This example is taken from a comment of Jyrki Lahtonen for a question on mathstackexchange.

Proof. We will assume $V \neq \{0\}$ and prove that V is uncountable. As $V \neq \{0\}$, there exists a non-zero vector $v \in V$. Given any two real numbers α and β, $\alpha.v = \beta.v$ implies that $(\alpha - \beta).v = 0$. As $v \neq 0$, the previous theorem tells us that $(\alpha - \beta) = 0$ or $\alpha = \beta$. Thus, for each real number α we get a distinct vector $\alpha.v$. Thus, there are atleast as many vectors as there are real numbers. Hence V is uncountable. □

Theorem 3.28. *Given an vector v, the vector $(-1).v$ is the additive inverse of v.*

Proof.

$$v + (-1).v = 1.v + (-1).v = (1 + (-1)).v = 0.v = 0$$

□

Remark 3.29. We use the article **the** to remind us that the additive inverse is unique.

Remark 3.30. Thus, we will denote the additive inverse of v by $-v$!

3.2. Maximally linearly independent list of vectors

Definition 3.31. Let $v_1, \ldots, v_k \in V$. We say $(v_1, \ldots, v_k)$ is linearly independent if $\forall(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$, $\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$ implies that $(\alpha_1, \ldots, \alpha_k) = 0$.

Definition 3.32. Let $v_1, \ldots, v_k \in V$. We say $(v_1, \ldots, v_k)$ is linearly dependent if it is not linearly independent. More precisely, $(v_1, \ldots, v_k)$ is linearly dependent if $\exists(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ such that $\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$ but $(\alpha_1, \ldots, \alpha_k) \neq 0$.

Theorem 3.33. *If $(v_1, v_2, \ldots, v_k)$ is linearly independent, then any subcollection $(v_{i_1}, v_{i_2}, \ldots, v_{i_m})$ (where $1 \leq i_1 \leq i_2 \leq \cdots \leq i_m \leq k$) is also linearly independent.*

Proof (by contradiction). Assume $v_{i_1}, v_{i_2}, \ldots, v_{i_m}$ are linearly dependent, that is, $\exists(a_{i_1}, \ldots, a_{i_m}) \neq 0$ such that $a_{i_1} v_{i_1} + \cdots + a_{i_m} v_{i_m} = 0$. Define $a_j = 0$ if $j \notin \{i_1, i_2, \ldots, i_m\}$. Then $(a_1, \ldots, a_n) \neq 0$ but,

$$a_1 v_1 + \ldots a_k v_k = a_{i_1} v_{i_1} + \cdots + a_{i_m} v_{i_m} = 0.$$

Thus, $v_1, \ldots, v_k$ are linearly dependent. But, this is a contradiction. So, our assumption that $v_{i_1}, v_{i_2}, \ldots, v_{i_m}$ are linearly dependent should be incorrect. □

Remark 3.34. Notice that the theorem and proof is verbatim the same as the statement and proof of Theorem 2.72. I would like you to observe and appreciate how similar the proofs are. It would be best if you tried to generalise results you have studied in the previous chapter to abstract vector spaces.

Definition 3.35. Let $v_1, \ldots, v_k \in V$. We say $(v_1, \ldots, v_k)$ is maximally linearly independent if $(v_1, \ldots, v_k)$ is linearly independent, but $(v_1, \ldots, v_k, v)$ is not linearly independent for any $v \in V$.

Example 3.36. The vectors $((1, 0), (0, 1))$ are maximally linearly independent in $\mathbb{R}^2$. Given any $(x, y) \in \mathbb{R}^2$, $x(1, 0) + y(0, 1) + (-1)(x, y) = (0, 0)$. Thus, $((1, 0), (0, 1), (x, y))$ is not linearly independent for any $(x, y) \in \mathbb{R}^2$.

Example 3.37. The vectors $((1, 0, 0), (0, 1, 0))$ are linearly independent in $\mathbb{R}^3$, but not maximally linearly independent as $((1, 0, 0), (0, 1, 0), (0, 0, 1))$ are linearly independent.

Definition 3.38. Given vectors $v_1, \ldots, v_k \in V$, $\text{Span}(v_1, \ldots, v_k)$ is the smallest subspace containing $\{v_1, \ldots, v_k\}$.

Exercise 3.39. Prove that $\text{Span}(v_1, \ldots, v_k) = \{\alpha_1 v_1 + \cdots + \alpha_k v_k \mid (\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k\}$.

Exercise 3.40. Let V be a vector space. Let $(v_1, \dots, v_n)$ and $(w_1, \dots, w_m)$ be two lists of vectors in V. Show that if $v_i \in span(w_1, \dots, w_m)$ for all $i \in \{1, \dots n\}$, then $span(v_1, \dots, v_n) \subset span(w_1, \dots, w_m)$.

Theorem 3.41. *Let $(V, +, .)$ be a vector space and let $(v_1, \dots, v_k)$ be maximally linearly independent list of vectors in V. Then, $Span(v_1, \dots, v_k) = V$.*

Proof. Let $v \in V$ be arbitrary. As $(v_1, \dots, v_k)$ is maximally linearly independent, $(v_1, \dots, v_k, v)$ is linearly dependent. That is there exits $0 \neq (\alpha_1, \dots, \alpha_k, \alpha) \in \mathbb{R}^{k+1}$ such that $\alpha_1 v_1 + \dots \alpha_k v_k + \alpha v = 0$. Further, $\alpha \neq 0$, as $\alpha = 0$ implies $(v_1, \dots, v_k)$ is linearly dependent which is a contradiction. Thus,

$$v = \left(-\frac{\alpha_1}{\alpha}\right).v_1 + \cdots + \left(-\frac{\alpha_k}{\alpha}\right).v_k.$$

That is $v \in \text{Span}(v_1, \dots, v_k)$. As $v \in V$ was arbitrary, we have the result. □

Theorem 3.42. *The collection $(v_1, \dots, v_k)$ is maximally linearly independent in $\mathbb{R}^n$ iff $k = n$ and $(v_1, v_2, \dots, v_n)$ is linearly independent.*

Proof. If $(v_1, \dots, v_k)$ is maximally linearly independent, then $(v_1, \dots, v_k)$ is linearly independent. Thus, by Corollary 2.76, $k \leq n$. As $(v_1, \dots, v_n)$ is maximally linearly independent, by Theorem 3.41, $\text{Span}(v_1, \dots, v_k) = \mathbb{R}^n$. Thus, by Corollary 2.77, $k \geq n$. Thus, $k = n$. Of course, if $(v_1, \dots, v_n)$ is maximally linearly independent, then it is linearly independent. Thus, we have the result. □

Example 3.43. The list $((1, 1, 1), (0, 1, 0), (0, 0, 1))$ is a maximally linearly independent list in $\mathbb{R}^3$. Notice that by the previous theorem, it is enough to show that $((1, 1, 1), (0, 1, 0), (0, 0, 1))$ is linearly independent. To show this consider $(\alpha, \beta, \gamma) \in \mathbb{R}^3$ such that $(0, 0, 0) = \alpha(1, 1, 1) + \beta(0, 1, 0) + \gamma(0, 0, 1) = (\alpha, \alpha + \beta, \alpha + \gamma)$. Thus, $(\alpha, \beta, \gamma) = (0, 0, 0)$.

Exercise 3.44. Show that the following are maximally linearly independent list of vectors in $\mathbb{R}^3$.

(1) $((1, 1, 0), (1, -1, 0), (0, 0, 1))$
(2) $((1, 2, 3), (4, 5, 6), (13, 14, 13))$

Example 3.45. Consider the list $(1, x, x^2)$ in $\mathcal{P}(2)$. Let $\alpha 1 + \beta.x + \gamma.x^2 0 = 0 + 0.x + 0.x^2$. Two polynomials are equal iff the coefficients of x^i is equal for all i. Thus, $\alpha = 0$, $\beta = 0$, and $\gamma = 0$. Thus, $(1, x, x^2)$ is linearly independent. Moreover, given any polynomial p of degree less than or equal to 2, there exists $a, b, c \in \mathbb{R}$ such that $p = a + bx + cx^2$. Thus, $(-a).1 + (-b).x + (-c).x^2 + 1.(a + bc + cx^2) = 0$. Thus, $(1, x, x^2, p)$ is linearly dependent for every $p \in \mathcal{P}(2)$. In other words, $(1, x, x^2)$ is a maximally linearly independent list.

Example 3.46. Consider the list $(1 + x + x^2, x, x^2)$ in $\mathcal{P}(2)$. Let $\alpha(1 + x + x^2) + \beta.x + \gamma.x^2 = 0 = 0 + 0.x + 0.x^2$. Two polynomials are equal iff the coefficients of x^i is equal for all i. But, $\alpha(1 + x + x^2) + \beta.x + \gamma.x^2 = (\alpha) + (\alpha + \beta)x + (\alpha + \gamma)x^2$. Thus, $\alpha = 0$, $\alpha + \beta = 0$, and $\alpha + \gamma = 0$. Thus, $(\alpha, \beta, \gamma) = (0, 0, 0)$ and hence $(1 + x + x^2, x, x^2)$ is linearly independent. Moreover, given any polynomial p of degree less than or equal to 2, there exists $a, b, c \in \mathbb{R}$ such that $p = a + bx + cx^2$. Thus, $a.(1 + x + x^2) + (b - a).x + (c - a).x^2 + (-1).(a + bc + cx^2) = 0$. Thus, $(1 + x + x^2, x, x^2, p)$ is linearly dependent for every $p \in \mathcal{P}(2)$. In other words, $(1 + x + x^2, x, x^2)$ is a maximally linearly independent list.

Exercise 3.47. Show that the following are maximally linearly independent list of vectors in $\mathcal{P}(2)$.

(1) $(1 + x, 1 - x, x^2)$
(2) $(1 + 2x + 3x^2, 4 + 5x + 6x^2, 13 + 14x + 13x^2)$

3.3. Minimal spanning list of vectors

Theorem 3.48. *Let $(V, +, .)$ is a vector space and let $v_1, \ldots, v_k$ be elements of V such that Span$(v_1, \ldots, v_k) = V$. Given any $l - k$ vectors $v_{k+1}, \ldots, v_l$, Span$(v_1, \ldots, v_l) =$ Span$(v_1, \ldots, v_k)$.*

Proof. As $\{v_1, \ldots, v_k\} \subset \{v_1, \ldots, v_l\}$, the smallest subset containing $\{v_1, \ldots, v_k\}$ is contained in the smallest subspace containing $\{v_1, \ldots, v_l\}$. Thus, $\text{Span}(v_1, \ldots, v_k) \subseteq \text{Span}(v_1, \ldots, v_l)$. But as $\text{Span}(v_1, \ldots, v_k) = V$, in particular $\{v_1, \ldots, v_l\} \subset \text{Span}(v_1, \ldots, v_k)$ and $\text{Span}(v_1, \ldots, v_k)$ is a subspace. Thus, the smallest subspace containing $\{v_1, \ldots, v_l\}$ is contained in $\text{Span}(v_1, \ldots, v_k)$. Thus, $\text{Span}(v_1, \ldots, v_l) \subseteq \text{Span}(v_1, \ldots, v_k)$ ☐

Thus, we should be interested in the "smallest" spanning list.

Definition 3.49. Let $(V, +, .)$ be a vector space over $\mathbb{R}$ and let $v_1, \ldots, v_k$ be vectors in V. We say the list $(v_1, \ldots, v_k)$ is a minimal spanning set if

(1) $\text{Span}(v_1, \ldots, v_k) = V$
(2) $\text{Span}(v_2, \ldots, v_k) \neq V$, $\text{Span}(v_1, \ldots, v_{k-1}) \neq V$ and for any $i \in \{2, \ldots, n-1\}$, we have $\text{Span}(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k) \neq V$.

Example 3.50. The list $((1,1,1),(0,1,0),(0,0,1))$ is a minimal spanning list in $\mathbb{R}^3$. First notice that given any $(a,b,c) \in \mathbb{R}^3$, $(a,b,c) = a(1,1,1) + (b-a)(0,1,0) + (c-1)(0,0,1)$. Thus, $\mathbb{R}^3 = \text{Span}((1,1,1),(0,1,0),(0,0,1))$. Further note that, $\text{Span}((0,1,0),(0,0,1)) = \{\alpha(0,1,0) + \beta(0,0,1) \mid (\alpha,\beta) \in \mathbb{R}^2\} = \{(0,\alpha,\beta) \mid (\alpha,\beta) \in \mathbb{R}^2\}$. Thus, the x-coordinate of every element in $\text{Span}((0,1,0),(0,0,1))$ is 0. Thus, $(1,1,1) \notin \text{Span}((0,1,0),(0,0,1))$. Similarly, we see that $\text{Span}((1,1,1),(0,1,0)) = \{\alpha(1,1,1) + \beta(0,1,0) \mid (\alpha,\beta) \in \mathbb{R}^2\} = \{(\alpha,\alpha+\beta,\alpha) \mid (\alpha,\beta) \in \mathbb{R}^2\}$. If $(0,0,1) \in \text{Span}((1,1,1),(0,1,0))$, then there exists $(\alpha,\beta) \in \mathbb{R}^2$ such that $(0,0,1) = (\alpha,\alpha+\beta,\alpha)$. That is $0 = \alpha = 1$. But, this is not possible. So, $(0,0,1) \notin \text{Span}((1,1,1),(0,1,0))$. Finally, we see that $\text{Span}((1,1,1),(0,0,1)) = \{\alpha(1,1,1) + \beta(0,0,1) \mid (\alpha,\beta) \in \mathbb{R}^2\} = \{(\alpha,\alpha,\alpha+\beta) \mid (\alpha,\beta) \in \mathbb{R}^2\}$. If $(0,1,0) \in \text{Span}((1,1,1),(0,0,1))$, then there exists $(\alpha,\beta) \in \mathbb{R}^2$ such that $(0,1,0) = (\alpha,\alpha,\alpha+\beta)$. That is $0 = \alpha = 1$. But, this is not possible. So, $(0,1,0) \notin \text{Span}((1,1,1),(0,0,1))$. Thus, $((1,1,1),(0,1,0),(0,0,1))$ is a minimal spanning list.

Exercise 3.51. Show that the following are minimal spanning list of vectors in $\mathbb{R}^3$.

(1) $((1,1,0),(1,-1,0),(0,0,1))$
(2) $((1,2,3),(4,5,6),(13,14,13))$

Example 3.52. Consider the list $(1,x,x^2)$ in $\mathcal{P}(2)$. First notice that given any polynomial $a + bx + cx^2 \in \mathcal{P}(2)$, $a + bx + cx^2 = a.1 + b.x + c.x^2$. Thus, $\mathcal{P}(2) = \text{Span}(1,x,x^2)$. Further note that, $\text{Span}(x,x^2) = \{\alpha.x + \beta.x^2 \mid (\alpha,\beta) \in \mathbb{R}^2\}$. Thus, the constant term of every polynomial in $\text{Span}(x,x^2)$ is 0. Thus, $1 \notin \text{Span}(x,x^2)$. Similarly, we see that $\text{Span}(1,x^2) = \{\alpha.1 + \beta.x^2 \mid (\alpha,\beta) \in \mathbb{R}^2\} = \{\alpha.1 + 0.x + \beta.x^2 \mid (\alpha,\beta) \in \mathbb{R}^2\}$. Thus, the co-efficient of x of every polynomial in $\text{Span}(1,x^2)$ is 0. Thus, $x \notin \text{Span}(1,x^2)$. Finally, we see that $\text{Span}(1,x) = \{\alpha.1 + \beta.x \mid (\alpha,\beta) \in \mathbb{R}^2\} = \{\alpha.1 + \beta.x + 0.x^2 \mid (\alpha,\beta) \in \mathbb{R}^2\}$. Thus, the co-efficient of x^2 of every polynomial in $\text{Span}(1,x)$ is 0. Thus, $x^2 \notin \text{Span}(1,x)$. Thus, $(1,x,x^2)$ is a minimal spanning list.

Example 3.53. Consider the list $(1+x+x^2,x,x^2)$ in $\mathcal{P}(2)$. First notice that given any polynomial $a + bx + cx^2 \in \mathcal{P}(2)$, $a + bx + cx^2 = a.(1+x+x^2) + (b-a).x + (c-a).x^2$. Thus, $\mathcal{P}(2) = \text{Span}(1 + x + x^2, x, x^2)$. Further note that, $\text{Span}(x,x^2) = \{\alpha.x + \beta.x^2 \mid (\alpha,\beta) \in \mathbb{R}^2\}$. Thus, the constant term of every polynomial in $\text{Span}(x,x^2)$ is 0. Thus, $1 + x + x^2 \notin \text{Span}(x,x^2)$. Similarly, we see that $\text{Span}(1+x+x^2,x^2) = \{\alpha.(1+x+x^2) + \beta.x^2 \mid (\alpha,\beta) \in \mathbb{R}^2\} = \{\alpha.1 + \alpha.x + \alpha + \beta.x^2 \mid (\alpha,\beta) \in \mathbb{R}^2\}$. If $x \in \text{Span}(1+x+x^2,x^2)$, then there exists $(\alpha,\beta) \in \mathbb{R}^2$ such that $0.1 + 1.x + 0.x^2 = \alpha.1 + \alpha.x + \alpha + \beta.x^2$.

That is $0 = \alpha = 1$. As this is not possible, $x \notin \text{Span}(1 + x + x^2, x^2)$. Finally, we see that $\text{Span}(1+x+x^2, x) = \{\alpha.(1+x+x^2)+\beta.x \mid (\alpha,\beta) \in \mathbb{R}^2\} = \{\alpha.1+(\alpha+\beta).x+\alpha.x^2 \mid (\alpha,\beta) \in \mathbb{R}^2\}$. If $x^2 \in \text{Span}(1+x+x^2, x)$, then there exists $(\alpha,\beta) \in \mathbb{R}^2$ such that $0.1 + 0.x + 1.x^2 = \alpha.1 + (\alpha + \beta).x + \alpha.x^2$. That is $0 = \alpha = 1$. As this is not possible, $x \notin \text{Span}(1 + x + x^2, x^2)$. Thus, $(1, x, x^2)$ is a minimal spanning list.

Exercise 3.54. Show that the following are minimal spanning list of vectors in $\mathcal{P}(2)$.

(1) $(1 + x, 1 - x, x^2)$

(2) $(1 + 2x + 3x^2, 4 + 5x + 6x^2, 13 + 14x + 13x^2)$

Theorem 3.55. *Let $(V, +, .)$ be a vector space over $\mathbb{R}$ and let $(v_1, \ldots, v_k)$ be a minimal spanning list. Then, $(v_1, \ldots, v_k)$ is linearly independent.*

Proof. Assume $(v_1, \ldots, v_k)$ is linearly dependent. Then, there exists $(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ such that $(\alpha_1, \ldots, \alpha_k) \neq 0$ but $\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$. As, $(\alpha_1, \ldots, \alpha_k) \neq 0$, there exists an $i \in \{1, \ldots, k\}$ such $\alpha_i \neq 0$. Without loss of generality we may assume that $\alpha_1 \neq 0$. Thus, $v_1 = \left(-\frac{\alpha_2}{\alpha_1}\right) v_2 + \cdots + \left(-\frac{\alpha_k}{\alpha_1}\right) v_k$. Thus, $v_1 \in \text{Span}(v_2, \ldots, v_k)$. Thus, $\text{Span}(v_2, \ldots, v_k) = \text{Span}(v_1, \ldots, v_k) = V$. That is, $(v_1, \ldots, v_k)$ is not a minimal spanning list. Hence our assumption that $(v_1, \ldots, v_k)$ is linearly dependent has to be wrong. □

Theorem 3.56. *Let $(V, +, .)$ be a vector space over $\mathbb{R}$. Then $(v_1, \ldots, v_k)$ is a minimal spanning list iff $(v_1, \ldots, v_k)$ is a maximally linearly independent list.*

Proof. Assume $(v_1, \ldots, v_k)$ is a minimal spanning list. Then, by Theorem 3.55, $(v_1, \ldots, v_k)$ is linearly independent. Further, as $\text{Span}(v_1, \ldots, v_k) = V$, given any vector $v \in V$, there exists $(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ such that $v = \alpha_1 v_1 + \ldots \alpha_k v_k$. Thus, $\alpha_1 v_1 + \ldots \alpha_k v_k + (-1)v = 0$. As, $-1 \neq 0$, we have $(v_1, \ldots, v_k, v)$ is linearly dependent. As $v \in V$ was arbitrary, we know that $(v_1, \ldots, v_k, v)$ is linearly dependent for every $v \in V$. Thus, $(v_1, \ldots, v_k)$ is a maximally linearly independent list.

Assume $(v_1, \ldots, v_k)$ is a maximally linearly independent list. Then, by Theorem 3.41, we have $\text{Span}(v_1, \ldots, v_k) = V$. Further, if $v_i \in \text{Span}(v_1, \ldots, v_{i-1}, v_{i+1}, v_k)$, then there exists a tuple $(\alpha_1, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_k) \in \mathbb{R}^{k-1}$ such that $v_i = \alpha_1 v_1 + \cdots + \alpha_{i-1} v_{i-1} + \alpha_{i+1} v_{i+1} + \ldots \alpha_k v_k$. That is, $\alpha_1 v_1 + \cdots + \alpha_{i-1} v_{i-1} (-1) v_i + \alpha_{i+1} v_{i+1} + \ldots \alpha_k v_k$. That is $(v_1, \ldots, v_k)$ is linearly dependent - a contradiction. Thus, $v_i \notin \text{Span}(v_1, \ldots, v_{i-1}, v_{i+1}, v_k)$ for all i. Thus, $(v_1, \ldots, v_k)$ is a minimal spanning set. □

Theorem 3.57. *Let $(V, +, .)$ be a vector space over $\mathbb{R}$. Let $(v_1, \ldots, v_k)$ be linearly independent and Span$(v_1, \ldots, v_k) = V$. Then,*

(1) $(v_1, \ldots, v_k)$ is a minimal spanning list

(2) $(v_1, \ldots, v_k)$ is a maximally linearly independent list.

Proof. As $\text{Span}(v_1, \ldots, v_k) = V$, given any vector $v \in V$ there exists $(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ such that $v = \alpha_1 v_1 + \cdots + \alpha_k v_k$. Thus, $\alpha_1 v_1 + \cdots + \alpha_k v_k + (-1)v = 0$. As $-1 \neq 0$, $(v_1, \ldots, v_k, v)$ is linearly dependent. As v was arbitrary, for all $v \in V$, $(v_1, \ldots, v_k, v)$ is linearly dependent. But, $(v_1, \ldots, v_k)$ is linearly independent. Thus, $(v_1, \ldots, v_k)$ is maximally linearly independent. And by Theorem 3.56, $(v_1, \ldots, v_k)$ is also a minimal spanning list. □

3.4. Finite dimensional vector spaces

Inspired by what we had seen for $\mathbb{R}^n$, given any vector space $(V, +, .)$ we would ideally like to construct a linearly independent spanning list. However, as we can see from the following example, it is not always possible.

Example 3.58. Let $(\mathcal{P}, +, .)$ be the vector space of polynomials. Given any k polynomials $(p_1, \ldots, p_k)$, let $n = \max\{deg(p_1), \ldots, deg(p_k)\}$. Then, note that $\{p_1, \ldots, p_k\} \subseteq \mathcal{P}(n)$. Thus, $\text{Span}(p_1, \ldots, p_k) \subseteq \mathcal{P}(n) \subsetneq \mathcal{P}$. Thus, we cannot find any finite list of vectors that span $\mathcal{P}$. If we cannot find a spanning set, we certainly cannot find a minimal spanning set.

Definition 3.59. A vector space $(V, +, .)$ is called a finite dimensional vector space (f.d.v.s) if there exits a list $(v_1, \ldots, v_k)$ of vectors in V that span V, that is $\text{Span}(v_1, \ldots, v_k) = V$.

Definition 3.60. A list of vectors $(v_1, \ldots, v_n)$ is called a basis of a vector space $(V, +, .)$ if

(1) $\text{Span}(v_1, \ldots, v_n) = V$.
(2) $(v_1, \ldots, v_n)$ is linearly independent.

Remark 3.61. Thus, by Theorem 3.57, $(v_1, \ldots, v_n)$ is a basis iff $(v_1, \ldots, v_n)$ is a minimal spanning list and maximally linearly independent list.

Example 3.62. The vector space $(\mathbb{R}^n, +, .)$ is a finite dimensional vector space and $(e_1, \ldots, e_n)$ is a basis.

Example 3.63. The list $((1, 1, 0), (1, -1, 0), (0, 0, 1))$ is a basis for $\mathbb{R}^3$. Another basis for $\mathbb{R}^3$ is the list $((1, 2, 3), (4, 5, 6), (13, 14, 13))$.

Example 3.64. The list $(1, x, x^2, x^3, x^4, x^5)$ is a basis for $\mathcal{P}(\in)$. First observe that 0 in $\mathcal{P}(2)$ is the polynomial $0.1 + 0.x + 0.x^2$. Thus, $\alpha_1.1 + \alpha_2.x + \alpha_3.x^2 = 0.1 + 0x + 0.x^2$. As two polynomials are equal iff their corresponding coefficients are equal, have $(\alpha_1, \alpha_2, \alpha_3) = 0$. Thus, $(1, x, x^2)$ is linearly independent. Further, given any polynomial p of degree less than or equal to 2, we can find $a, b, c \in \mathbb{R}$ such that $p = a + bx + cx^2$. Therefore, $p = a.1 + b.x + c.x^2 \in \text{Span}(1, x, x^2)$.

More generally,

Example 3.65. The vector space of polynomials of degree less than or equal to n, $\mathcal{P}(n)$ is a finite dimensional vector space, with a basis $(1, x, \ldots, x^n)$.

Example 3.66. The list $(1 + x, 1 - x, x^2)$ is a basis for $\mathcal{P}(2)$ and the list $(1 + 2x + 3x^2, 4 + 5x + 6x^2, 13 + 14x + 13x^2)$ is another basis for $\mathcal{P}(2)$.

Example 3.67. Let $I_n = \{1, \ldots, n\}$. Then, the vector space $(\mathcal{F}(I_n, \mathbb{R}), +, .)$ is a finite dimensional vector space. Moreover, if we define functions $f_i : I_n \to \mathbb{R}$ as

$$f_i(j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

then $(f_1, \ldots, f_n)$ form a basis for $(\mathcal{F}(I_n, \mathbb{R}), +, .)$. To see that given any function $f : I_n \to \mathbb{R}$, define the linear combination $f(1)f_1 + \ldots f(n)f_n$. Then, notice that $(f(1)f_1 + \cdots + f(n)f_n)(i)$

$$\begin{aligned} &= f(1)f_1(i) + \cdots + f(i-1)f_{i-1}(i) + f(i)f_i(i) + f(i+1)f_{i+1} + \cdots + f(n)f_n(i) \\ &= f(1).0 + \cdots + f(i-1).0 + f(i).1 + f(i+1).0 + \cdots + f(n).0 \\ &= f(i) \end{aligned}$$

As i was arbitrary, $f(1)f_1 + \dots f(n)f_n = f$. As f was arbitrary, $\text{Span}(f_1, \dots, f_n) = \mathcal{F}(I_n, \mathbb{R})$. Moreover, if $\alpha_1 f_1 + \dots + \alpha_n f_n = 0$, then $\alpha_1 f_1 + \dots + \alpha_n f_n(i) = 0$ for all i (as here the zero vector is the constant function 0). But, $\alpha_1 f_1 + \dots + \alpha_n f_n(i) = \alpha_i$. Therefore, $(\alpha_1, \dots, \alpha_n) = 0$, thus, $(f_1, \dots, f_n)$ is linearly independent.

Note that the trivial vector space $(0, +, .)$ is finite dimensional as $\text{Span}(0) = \{0\}$. However, as the only vector in the vector space is the 0 vector, we cannot find a linearly independent list of vectors - unless we accept the empty list! There is some merit in accepting the empty list as that would give us a better reason to say that the trivial vector space has dimension 0. Else, you can just assume that vector space is non-trivial.

Theorem 3.68. *Every non-trivial finite dimensional vector space has a basis.*

Proof. Let $(V, +, .)$ be an arbitrary non-trivial finite dimensional vector space. As V is finite dimensional, there exists vectors $(v_1, \dots, v_k)$ such that $\text{Span}(v_1, \dots, v_k) = V$. If $(v_1, \dots, v_k)$ is not linearly independent, then there exists, $(\alpha_1, \dots, \alpha_k) \in \mathbb{R}^k$ such that $(\alpha_1, \dots, \alpha_k) \neq 0$ and $\alpha_1 v_1 + \dots + \alpha_k v_k = 0$. As the span does not depend on the ordering of the vectors, we might assume without loss of generality that $\alpha_k \neq 0$. Thus, $v_k = -\left(\frac{\alpha_1}{\alpha_k}\right) v_1 + \dots + -\left(\frac{\alpha_{k-1}}{\alpha_k}\right) v_{k-1}$. Thus, $v_k \in \text{Span}(v_1, \dots, v_{k-1})$ and hence $\text{Span}(v_1, \dots, v_{k-1}) = \text{Span}(v_1, \dots, v_k)$. Keep repeating the process until you get a linearly independent list of vectors. It is a valid concern that we might end up in the scenario where $\text{Span}(v) = V$, but (v) is not linearly independent. But, a single vector is linearly dependent iff it is the 0 vector. Thus, this would imply that $V = \{0\}$. But, we had assumed that V is a non-trivial finite dimensional vector space, so this cannot happen. □

Notice that in the above proof, we took a spanning list and threw away some vectors to form a basis. We can do the opposite process, if we start with a linearly independent list. More precisely,

Theorem 3.69. *Let $(V, +, .)$ be a finite dimensional vector space. Then, given any linearly independent list $(v_1, \dots, v_k)$, $\exists\ v_{k+1}, \dots, v_n \in V$ such that $(v_1, \dots, v_n)$ is a basis of V.*

To make the proof easier to understand, we introduce the operation of appending an element to a list. More precisely, given a list $(v_1, \dots, v_m)$ and a vector v, we say $(v_1, \dots, v_m) \oplus v = (v_1 \dots, v_m, v)$.

Proof. As V is a finite dimensional vector space, there exists vectors $(w_1, \dots, w_l)$ such that $\text{Span}(w_1, \dots, w_l) = V$. Let us say $B = (v_1, \dots, v_k)$ to begin with. The idea is to successively modify the list B, till we get a basis. First, if $w_1 \in \text{Span}(B)$ then keep B as it is, else assign the name B to $B \oplus w_1$. Keep repeating this process. At the ith state, if $w_i \in \text{Span}(B)$ then keep B as it is, else assign the name B to $B \oplus w_i$. Notice that at each step, we are adding a vector that does not belong to $\text{Span}(B)$, thus the list continues to remain linearly independent. Moreover, after l steps, we know that $\{w_1, \dots, w_l\} \subset \text{Span}(B)$. Thus, B forms a basis. □

Corollary 3.70. *Let $(V, +, .)$ be a finite dimensional vector space, let $n = dim(V)$, and let $(v_1, \dots, v_n)$ be a linearly independent list of vectors. Then, $(v_1, \dots, v_n)$ is a basis of V.*

Theorem 3.71. *Let $(V, +, .)$ be a vector space. A list $(v_1, \dots, v_n)$ is a basis iff $\forall v \in V$, there exists a unique $(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ such that $v = \alpha_1 v_1 + \dots + \alpha_n v_n$.*

Proof. Let us assume that $(v_1, \dots, v_n)$ is a basis. Then, $\text{Span}(v_1, \dots, v_n) = V$. Thus, given any $v \in V$, there exists $(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ such that $v = \alpha_1 v_1 + \dots + \alpha_n v_n$. So, we just need to show that this representation is unique. But, if $(\beta_1, \dots, \beta_n) \in \mathbb{R}^n$ and $v = \beta_1 v_1 + \dots + \beta_n v_n$. Thus, $\alpha_1 v_1 + \dots + \alpha_n v_n = \beta_1 v_1 + \dots + \beta_n v_n$. That is $(\alpha_1 - \beta_1) v_1 + \dots + (\alpha_n - \beta_n) v_n = 0$. As, $(v_1, \dots, v_n)$

is linearly independent, this would imply that $\alpha_i = \beta_i$ for all i. Thus, the representation is indeed unique.

Now assume that $\forall v \in V$, there exists a unique $(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Then, clearly, $\text{Span}(v_1, \ldots, v_n) = V$. Moreover, note that, $0 = 0.v_1 + \cdots + 0.v_n$. Thus, by uniqueness, if $\alpha_1 v_1 + \ldots \alpha_n v_n = 0$, then $\alpha_i = 0$ for all i. Therefore, $(v_1, \ldots, v_n)$ is linearly independent. Hence $(v_1, \ldots, v_n)$ is a basis. □

Lemma 3.72. *If $v \in Span(v_1, \ldots, v_k)$, then $(v, v_1, \ldots, v_k)$ is linearly dependent.*

Proof. As $v \in \text{Span}(v_1, \ldots, v_k)$, there exists $(\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ such that $v = \alpha_1 v_1 + \cdots + \alpha_k v_k$. Therefore, $(-1)v + \alpha_1 v_1 + \cdots + \alpha_k v_k = 0$ and therefore $(v, v_1, \ldots, v_k)$ is linearly dependent. □

Theorem 3.73. *Let $(V, +, .)$ be a finite dimensional vector space. Let $v_i \in V$ for all $i \in \{1, \ldots, k\}$ and let $w_j \in V$ for all $j \in \{1, \ldots, l\}$. If $(v_1, \ldots, v_k)$ are linearly independent and $Span(w_1, \ldots, w_l) = V$, then $k \leq l$.*[2]

The proof verbatim the same as the proof of Theorem 2.75, the analogous result for subspaces in $\mathbb{R}^n$. This proof is very involved. So, we would first give the key ideas of the proof without getting into the details. Hopefully, keeping this idea in mind would help you appreciate the proof better.

Proof Sketch. We start with the list $(w_1, \ldots, w_l)$ and adjoin the vector v_k to it to obtain the list $(v_k, w_1, \ldots, w_l)$. We show $(v_k, w_1, \ldots, w_l)$ is linearly dependent and therefore we may without loss of generality assume $\text{Span}(w_1, \ldots, w_l) = \text{Span}(v_k, w_2, \ldots, w_l)$. Thus, we replace the list $(w_1, \ldots, w_l)$ with $(v_k, w_2, \ldots, w_l)$. We repeat this process - at every stage, we add a v and drop a w. If we run out of w's before we run out of v's then a subcollection of v's will span V. This would contradict the fact that $(v_1, \ldots, v_k)$ is linearly independent. Thus, there should be at least as many w's as there are v's. □

Proof. As $\text{Span}(w_1, \ldots, w_l) = V$ and $v_k \in V$, by Lemma 3.72 we know that $(v_k, w_1, \ldots, w_l)$ is linearly dependent. Thus, there exists $(a_k, b_1, \ldots, b_l) \neq 0$ such that $a_k v_k + b_1 w_1 + \cdots + b_l w_l = 0$. Notice if $(b_1, \ldots, b_l) = 0$, then $a_k \neq 0$ but $0 = a_k v_k + b_1 w_1 + \cdots + b_l w_l = a_k v_k$. Which implies that v_k is the 0 vector. But then $0v_1 + \cdots + 0v_{k-1} + 1v_k = 0$. This contradicts our assumption that $(v_1, \ldots, v_k)$ is linearly independent. So, $(b_1, \ldots, b_l) \neq 0$. If $l = 1$, then $V = \text{Span}(w_1)$ and $b_1 \neq 0$. Thus, $w_1 = -\frac{a_k}{b_1} v_k$. Hence $\text{Span}(v_k) = \text{Span}(w_1) = V$. As $(v_1, \ldots, v_k)$ is linearly independent, this would imply that $k = 1$. Else, $v_1 \in \text{Span}(v_k)$, which means $v_1 = \alpha_k v_k$. Then, $(-1)v_1 + 0v_2 + \cdots + 0v_{k-1} + \alpha_k v_k = 0$ and $(v_1, \ldots, v_k)$ is linearly dependent - a contradiction. Thus, if $l = 1$, then $k = 1$ and we have $k \leq l$. So, we may assume $l > 1$. If $k = 1$, then $k < l$. Thus, we may assume that $k > 1$. Without loss of generality (Notice that span does not depend on the order of the vectors. Thus, if needed we can renumber $w_1, \ldots, w_l$) we may assume that $b_1 \neq 0$. Thus,

$$w_1 = \left(-\frac{a_k}{b_1}\right) v_k + \left(-\frac{b_2}{b_1}\right) w_2 + \left(-\frac{b_3}{b_1}\right) w_3 + \cdots + \left(-\frac{b_l}{b_1}\right) w_l.$$

Thus, $V = \text{Span}(w_1, \ldots, w_l) \subset \text{Span}(v_k, w_2, \ldots, w_l)$. But as $\{v_k, w_2, \ldots, w_l\} \subset V$, we have $\text{Span}(v_k, w_2, \ldots, w_l) \subset V$. That is $\text{Span}(v_k, w_2, \ldots, w_l) = V$.

Now, As $\text{Span}(v_k, w_2, \ldots, w_l) = V$ and $v_{k-1} \in V$, by Lemma 2.74 $(v_{k-1}, v_k, w_2, \ldots, w_l)$ is linearly dependent. Thus, there exists $(a_{k-1}, a_k, b_2, \ldots, b_l) \neq 0$ such that $a_{k-1}v_{k-1} + a_k v_k + b_2 w_2 + \cdots + b_l w_l = 0$. Notice if $(b_2, \ldots, b_l) = 0$, then $(a_{k-1}, a_k) \neq 0$ but $0 = a_{k-1}v_{k-1} + a_k v_k + b_1 w_1 + \cdots + b_l w_l = a_{k-1}v_{k-1} + a_k v_k$. But then $0v_1 + \cdots + 0v_{k-2} + a_{k-1}v_{k-1} + a_k v_k = 0$. This contradicts our assumption that $(v_1, \ldots, v_k)$ is linearly independent. So, $(b_2, \ldots, b_l) \neq 0$. If $l = 2$, then

[2] The statement of this theorem and its proof is from [Axler]. The proof is however significantly elaborated.

$V = \text{Span}(v_k, w_2)$ and $b_2 \neq 0$. Thus, $w_2 = \left(-\frac{a_{k-1}}{b_2}\right) v_{k-1} + \left(-\frac{a_k}{b_2}\right) v_k$. Thus, $V = \text{Span}(v_k, w_2) \subset \text{Span}(v_{k-1}, v_k)$. But, as $\{v_{k-1}, v_k\} \subset V$, $\text{Span}(v_{k-1}, v_k) = V$. As $(v_1, \ldots, v_k)$ is linearly independent, this would imply that $k = 2$. Else, $v_1 \in \text{Span}(v_{k-1}, v_k)$, which means $v_1 = \alpha_{k-1} v_{k-1} + \alpha_k v_k$. Then, $(-1)v_1 + 0v_2 + \cdots + 0v_{k-2} + \alpha_{k-1}v_{k-1} + \alpha_k v_k = 0$ and $(v_1, \ldots, v_k)$ is linearly dependent - a contradiction. Thus, if $l = 2$ and $k > 1$, then $k = 2$ and we have $k \leq l$. So, we may assume $l > 2$. If $k = 2$, then $k < l$ and we are done. So, we may assume $k > 2$. Without loss of generality (by renumbering $w_2, \ldots, w_l$) we may assume that $b_2 \neq 0$. Thus,

$$w_2 = \left(-\frac{a_{k-1}}{b_2}\right) v_{k-1} + \left(-\frac{a_k}{b_2}\right) v_k + \left(-\frac{b_3}{b_2}\right) w_3 + \left(-\frac{b_4}{b_2}\right) w_4 + \cdots + \left(-\frac{b_l}{b_2}\right) w_l.$$

Thus, $V = \text{Span}(v_k, w_2, \ldots, w_l) \subset \text{Span}(v_{k-1}, v_k, w_3, \ldots, w_l)$. But as $\{v_{k-1}, v_k, w_3, \ldots, w_l\} \subset V$, we have $\text{Span}(v_{k-1}, v_k, w_3, \ldots, w_l) \subset V$. That is $\text{Span}(v_{k-1}, v_k, w_3, \ldots, w_l) = V$.

We can continue this process and assume $l > i$ and $k > i$. As $\text{Span}(v_{k-i+1}, \ldots, v_k, w_{i+1}, \ldots, w_l) = V$ and $v_{k-i} \in V$, by Lemma 2.74 $(v_{k-i}, \ldots, v_k, w_{i+1}, \ldots, w_l)$ is linearly dependent. Thus, there exists $(a_{k-i}, \ldots, a_k, b_{i+1}, \ldots, b_l) \neq 0$ such that $a_{k-i}v_{k-i} + \cdots + a_k v_k + b_{i+1}w_{i+1} + \cdots + b_l w_l = 0$. Notice if $(b_{i+1}, \ldots, b_l) = 0$, then $(a_{k-i}, \ldots, a_k) \neq 0$ but $0 = a_{k-i}v_{k-i} + \cdots + a_k v_k + b_{i+1}w_{i+1} + \cdots + b_l w_l = a_{k-i}v_{k-i} + \cdots + a_k v_k$. But then $0v_1 + \cdots + 0v_{k-i-1} + a_{k-i}v_{k-i} + \cdots + a_k v_k = 0$. This contradicts our assumption that $(v_1, \ldots, v_k)$ is linearly independent. So, $(b_{i+1}, \ldots, b_l) \neq 0$. If $l = i + 1$, then $\text{Span}(v_{k-i+1}, \ldots, v_k, w_l) = V$ and $b_l \neq 0$. Thus,

$$w_l = \left(-\frac{a_{k-i}}{b_l}\right) v_{k-i} + \cdots + \left(-\frac{a_k}{b_l}\right) v_k.$$

That is $V = \text{Span}(v_{k-i+1}, \ldots, v_k, w_l) \subset \text{Span}(v_{k-i}, \ldots, v_k)$. As $\{v_{k-i}, \ldots, v_k\} \subset V$, we have $\text{Span}(v_{k-i}, \ldots, v_k) = V$. If $k - i > 1$, then $v_1 \in \text{Span}(v_{k-i}, \ldots, v_k) = V$ and hence $(v_1, \ldots, v_k)$ is linearly dependent. Therefore $k - i \leq 1$, that is $k \leq i + 1 = l$. **Thus, we will not run out of w's before v's**.

Thus, the above process works for each i. When $i = k-1$, the process terminates and the proof becomes complete. Thus, we have the result. □

Theorem 3.74. *Let $(V, +, .)$ be a finite dimensional vector space. Let $(v_1, \ldots, v_m)$ and $(w_1, \ldots, w_n)$ both be a basis for V. Then, $m = n$.*

Proof. As $(v_1, \ldots, v_n)$ is a basis, $(v_1, \ldots, v_m)$ is linearly independent and $\text{Span}(v_1, \ldots, v_m) = V$. And as $(w_1, \ldots, w_n)$ is a basis, $(w_1, \ldots, w_n)$ is linearly independent and $\text{Span}(w_1, \ldots, w_n) = V$. But, $(v_1, \ldots, v_m)$ is linearly independent and $\text{Span}(w_1, \ldots, w_n) = V$ would imply (using previous theorem) that $m \leq n$. Also, $(w_1, \ldots, w_n)$ is linearly independent and $\text{Span}(v_1, \ldots, v_m) = V$ would imply (by previous theorem) that $n \leq m$. Thus, $m = n$. □

Definition 3.75. Let $(V, +, .)$ be a finite dimensional vector space. We say the dimension of V (denoted as $dim(V)$) is n if there exists a basis $(v_1 \ldots, v_n)$ of V.

Example 3.76. Let $M_2(\mathbb{R}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mid a, b, c, d \in \mathbb{R} \right\}$ and let + and . be defined as follows:

$$\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} + \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix} = \begin{bmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{bmatrix}$$

$$\alpha \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \alpha a & \alpha b \\ \alpha c & \alpha d \end{bmatrix}$$

Then, $M_2(\mathbb{R})$ is a vector space. Moreover, $\left(\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right)$ is a basis for $M_2(\mathbb{R})$. To show that Span $\left(\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = M_2(\mathbb{R})$, note that given any element $\begin{bmatrix}a & b\\c & d\end{bmatrix} \in M_2(\mathbb{R})$,

$$\begin{bmatrix}a & b\\c & d\end{bmatrix} = a\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + b\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + c\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + d\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}.$$

As the zero in this vector space is the matrix $\begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$, it is also clear that

$$\begin{bmatrix}0 & 0\\0 & 0\end{bmatrix} = \alpha\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + \beta\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + \gamma\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + \delta\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix} = \begin{bmatrix}\alpha & \beta\\\gamma & \delta\end{bmatrix}$$

iff $(\alpha, \beta, \gamma, \delta) = 0$. Thus, $\left(\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 1\\0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\1 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right)$ is linearly independent. Thus, we can conclude that $dim(M_2(\mathbb{R})) = 4$.

Example 3.77. Conside the vector space $\mathcal{F}(\{1,2,3\}, \mathbb{R})$. Note that any function $f : \{1,2,3\} \to \mathbb{R}$ is defined uniquely once we specify $f(1), f(2)$, and $f(3)$. Let $f_1 : \{1,2,3\} \to \mathbb{R}$ be the function with $f_1(1) = 1$, $f_1(2) = 0$, and $f_1(3) = 0$. Let $f_2 : \{1,2,3\} \to \mathbb{R}$ be the function with $f_2(1) = 0$, $f_2(2) = 1$, and $f_2(3) = 0$. Let $f_3 : \{1,2,3\} \to \mathbb{R}$ be the function with $f_3(1) = 0$, $f_3(2) = 0$, and $f_3(3) = 1$. Then, (f_1, f_2, f_3) is a basis for $\mathcal{F}(\{1,2,3\}, \mathbb{R})$. To show linear independence, note that the zero vector is the function $0 : \{1,2,3\} \to \mathbb{R}$ such that $0(1) = 0$, $0(2) = 0$, and $0(3) = 0$. Thus, $\alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3 = 0$ iff

$$\begin{aligned}\alpha_1 &= (\alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3)(1) = 0(1) = 0\\ \alpha_2 &= (\alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3)(2) = 0(2) = 0\\ \alpha_3 &= (\alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3)(3) = 0(3) = 0\end{aligned}$$

Thus, (f_1, f_2, f_3) is linearly independent. To show that (f_1, f_2, f_3) span $\mathcal{F}(\{1,2,3\}, \mathbb{R})$, note that given any $f : \{1,2,3\} \to \mathbb{R}$, $f = f(1).f_1 + f(2).f_2 + f(3)f_3$ as

$$\begin{aligned}f(1) &= f(1).1 + f(2).0 + f(3).0 = (f(1).f_1 + f(2).f_2 + f(3)f_3)(1)\\ f(2) &= f(1).0 + f(2).1 + f(3).0 = (f(1).f_1 + f(2).f_2 + f(3)f_3)(2)\\ f(3) &= f(1).0 + f(2).0 + f(3).1 = (f(1).f_1 + f(2).f_2 + f(3)f_3)(3)\end{aligned}$$

Theorem 3.78. *Let $(V, +, .)$ be a finite dimensional vector space and let W be a subspace of V. Then, W is also a finite dimensional vector space. Moreover, $dim(W) \leq dim(V)$.*

Proof. As V is finite dimensional, there exists a list $(v_1, \ldots, v_n)$ such that $Span(v_1, \ldots, v_n) = V$. If $W = \{0\}$, then W is a f.d.v.s by convention and $dim(W) = 0$. If $W \neq \{0\}$, let $w_1 \in W$ be a a non-zero vector. If $Span(w_1) = W$, then W is not finite dimensional. If not, let $w_2 \in W \setminus Span(w_1)$ - that is (w_1, w_2) is linearly independent. If $Span(w_1, w_2) = W$, then W is finite dimensional. If not, let $w_3 \in W \setminus Span(w_1, w_2)$ - that is (w_1, w_2, w_3) is linearly independent. Keep continuing this process. After i steps, you would have a linearly independent list $(w_1, \ldots, w_i)$. But, note that $i \leq n$ by Theorem 3.73. Thus, the process has to terminate.

Further, let $(v_1, \dots, v_n)$ is a basis of V and $(w_1, \dots, w_k)$ be a basis of W. Then, $(w_1, \dots, w_k)$ is a linearly independent list of vectors in V and $(v_1, \dots, v_n)$ is a spanning list. Thus, by Theorem 3.73, $k \leq n$. That is, $dim(W) \leq dim(V)$. □

Exercise 3.79. Let $\mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ be the set of linear functions from $\mathbb{R}^3$ to $\mathbb{R}^2$. Given $f, g \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$, define $f + g : \mathbb{R}^3 \to \mathbb{R}^2$ as $(f + g)(v) = f(v) + g(v)$. Given $f \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ and $\alpha \in \mathbb{R}$, define $\alpha.f : \mathbb{R}^3 \to \mathbb{R}^2$ as $(\alpha.f)(v) = \alpha f(v)$. Then, show that $(\mathcal{L}(\mathbb{R}^3, \mathbb{R}^2), +, .)$ forms a vector space. Is $\mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ finite dimensional space? Prove your claim

Exercise 3.80. Let P_3 be the set of all polynomials of degree less than or equal to 3. Prove or disprove: there exists a basis (p_0, p_1, p_2, p_3) of P_3 such that none of the polynomials p_0, p_1, p_2, p_3 has degree 2.

Exercise 3.81. Let $V = \mathbb{R}^3$. Consider the two subspaces $U = span((1, 0, 0), (3, 1, 1), (-1, 1, 1))$ and $W = span((0, 1, 2), (0, 0, 1))$. Find the dimension of $U \cap W$. Explain your answer!

Exercise 3.82. Let $\mathcal{P}(n)$ be set of all polynomials of degree less than or equal to n. Show that $(1, 1 + x, 1 + x + x^2, 1 + x + x^2 + x^3, \dots, 1 + x + \cdots + x^n)$ is a basis of $\mathcal{P}(n)$.

Exercise 3.83. Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. Show that the set $\{v \in \mathbb{R}^n : L(v) = \lambda v\}$ is a subspace for all $\lambda \in \mathbb{R}$.

Exercise 3.84. Let $U = \{f \in \mathcal{F}(\mathbb{N}, \mathbb{R}) \mid f(n) = 0 \ \forall n \geq 10\}$. Show that U is a subspace of $(\mathcal{F}(\mathbb{N}, \mathbb{R}), +, .)$ and find the dimension of U.

Linear maps

Vector spaces were introduced to serve as domains and codomains of "linear" functions. Thus, it would be a waste to not study linear maps.

Definition 4.1. Given two vector spaces $(V, +, .)$ and $(W, \oplus, \otimes)$, a function $L : V \to W$ is called linear if

(1) $L(u + v) = L(u) \oplus L(v)$
(2) $L(\alpha.v) = \alpha \otimes L(v)$

We have seen several examples of linear functions from $\mathbb{R}^m$ to $\mathbb{R}^n$. Let us now look at some examples of linear maps between more abstract vector spaces.

Example 4.2. Let $\mathcal{P}$ be the vector space of polynomials. Then, the function $D : \mathcal{P} \to \mathcal{P}$ defined as

$$D(a_0 + \cdots + a_n x^n) = a_1 + 2a_2 x + \ldots n a_n x^n$$

is a linear map. If you want to restrict your attention to finite dimensional vector space, you can restrict the domain to $\mathcal{P}(n)$. Notice that, if a polynomial p has degree n, then degree of $D(p)$ is less than or equal to $n - 1$. Thus, $D : \mathcal{P}(n) \to \mathcal{P}(n - 1)$ is also a linear map.

Example 4.3. Let $\mathcal{P}$ be the vector space of polynomials. Then, the function $I : \mathcal{P} \to \mathcal{P}$ defined as

$$I(a_0 + \cdots + a_n x^n) = a_0 x + \frac{a_1}{2} x^2 + \frac{a_2}{3} x^3 + \ldots \frac{a_n}{n+1} x^{n+1}$$

is a linear map. If you want to restrict your attention to finite dimensional vector space, you can restrict the domain to $\mathcal{P}(n)$. Notice that, if a polynomial p has degree n, then degree of $I(p)$ is less than or equal to $n + 1$. Thus, $I : \mathcal{P}(n) \to \mathcal{P}(n + 1)$ is also a linear map.

Theorem 4.4. *Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two vector spaces. Then, $L(0) = 0$.*

Proof. By Theorem 3.26 $0.v = 0$ for all vector v. As L is linear, $L(0) = L(0.v) = 0 \otimes L(v)$. Applying Theorem 3.26 again, we get $L(0) = 0$. □

4.1. Null set and Range

Definition 4.5. Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two vector spaces and $L : V \to W$ be a linear map. Then, the range or image of L denoted as $Im(L)$ is the subset $\{L(v) \mid v \in V\}$. Note that $Im(L)$ is a subspace because

(1) $0 = L(0) \in Im(L)$
(2) If $w_1, w_2 \in Im(L)$, then there exits $v_1, v_2 \in V$ such that $w_1 = L(v_1)$ and $w_2 = L(v_2)$. Thus, $w_1 \oplus w_2 = L(v_1) \oplus L(v_2) = L(v_1 + v_2) \in Im(L)$.
(3) If $w \in Im(L)$, then there exists $v \in V$ such that $L(v) = w$. Thus, $\alpha \otimes w = \alpha \otimes L(v) = L(\alpha v) \in Im(L)$.

The following theorem is then obvious

Theorem 4.6. *Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two vector spaces and $L : V \to W$ be a linear map. Then, L is surjective iff $Im(L) = W$.*

Definition 4.7. Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two vector spaces and $L : V \to W$ be a linear map. Then, the null space or kernel of L denoted as $Ker(L)$ or $Null(L)$ is the subset $\{v \in V \mid L(v) = 0\}$. Note that $Ker(L)$ is a subspace because

(1) $L(0) = 0$
(2) If $v_1, v_2 \in Ker(L)$, then $L(v_1) = 0 = L(v_2)$. Thus, $L(v_1 + v_2) = L(v_1) \oplus L(v_2) = 0 \oplus 0 = 0$. Hence $v_1 + v_2 \in Ker(L)$.
(3) If $v \in Ker(L)$, then $L(\alpha.v) = \alpha \otimes L(v) = \alpha \otimes 0 = 0$. Thus, $\alpha.v \in Ker(L)$.

Theorem 4.8. *Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two vector spaces and $L : V \to W$ be a linear map. Then, L is injective iff $Ker(L) = \{0\}$.*

Proof. Assume L is injective. We already proved that $L(0) = 0$. Thus, if $v \in V$ and $L(v) = 0 = L(0)$, then $v = 0$ by injectivity of L. Thus, $Ker(L) = \{0\}$.

Now assume $Ker(L) = \{0\}$. To show L is injective, we would assume $L(v) = L(w)$ and prove $v = w$. If $L(v) = L(w)$, then $L(v) \oplus (-1) \otimes L(w) = 0$. Thus, then $L(v + (-1).w) = L(v) \oplus (-1) \otimes L(w) = 0$. Thus, $v + (-1).w \in Ker(L) = \{0\}$. Thus, $v = w$. □

We saw that given a linear map $L : V \to W$, $Ker(L)$ is a subspace of V. It is therefore natural to ask if given a subspace U of a (finite dimensional vector space)[1] $(V+, .)$, does there exists a vector space $(W, \oplus, \otimes)$ and a linear map $L : V \to W$ such that $Ker(L) = U$? Before answering the general question, it might be useful to work with some examples.

Example 4.9. Let $V = \mathbb{R}^2$ and let $U = \{(x, mx) \mid x \in \mathbb{R}\}$. Then note that $L : \mathbb{R}^2 \to \mathbb{R}$ defined as $L(x, y) = y - mx$ does the trick - in other words, $Ker(L) = U$.

Example 4.10. Let $V = \mathbb{R}^3$ and let $U = \{(x, y, z) \mid 3x + 2y + 4z = 0\}$. Then, $L : \mathbb{R}^3 \to \mathbb{R}$ defined as $L(x, y, z) = 3x + 2y + 4z$ does the trick.

Example 4.11. Let $V = \mathbb{R}^3$ and let $U = \{(x, y, z) \mid 3x + 2y + 4z = 0 \text{ and } y + z = 0\}$. Then, $L : \mathbb{R}^3 \to \mathbb{R}^2$ defined as $L(x, y, z) = (3x + 2y + 4z, y + z)$ does the trick.

In the above two examples, the subspace U was given in a comfortable form which enabled us to find the function quickly.

Example 4.12. Let $V = \mathbb{R}^3$ and let $U = \text{Span}((1, 2, 3))$. Of course, we could express this subspace as the solution of two linear equations (or intersection of two planes), but we will take a different approach. We note that the basis of U can be extended to form a basis of V, namely, $((1, 2, 3), (0, 1, 0), (0, 0, 1))$ forms a basis of V. Now, given any vector (x, y, z), note that $(x, y, z) = x(1, 2, 3) + (y - 2x)(0, 1, 0) + (z - 3x)(0, 0, 1)$. As $(1, 2, 3) \in Ker(L)$, we would certainly

[1] You can ask the question for more general vector spaces, but the answer is a lot harder

want $L(1,2,3) = 0$. Moreover, we want $(L(0,1,0), L(0,0,1))$ to be linearly independent if we want L to be injective. Thus, the codomain should be atleast two dimensional. Hence we may assume WLOG that the codomain is $\mathbb{R}^2$ and $L(0,1,0) = (1,0)$ and $L(0,0,1) = (0,1)$. Thus,

$$\begin{aligned} L(x,y,z) &= L(x(1,2,3) + (y-2x)(0,1,0) + (z-3x)(0,0,1)) \\ &= xL(1,2,3) + (y-2x)L(0,1,0) + (z-3x)L(0,0,1) \\ &= 0 + (y-2x)(1,0) + (z-3x)(0,1) \\ &= (y-2x, z-3x) \end{aligned}$$

The approach in the previous example, guides the proof of the more general result

Theorem 4.13. *Let $(V+,.)$ be a finite dimensional vector space and let U be a subspace. Then, there exists a vector space $(W,\oplus,\otimes)$ and a linear map $L : V \to W$ such that $Ker(L) = U$.*

Proof. As U is a subspace of V, by Theorem 3.78, U is finite dimensional and $dim(U) \leq dim(V)$. Let $(v_1,\ldots,v_k)$ be a basis for U. Using Theorem 3.69, we can extend this list to form a basis $(v_1,\ldots,v_n)$. Therefore given any $v \in V$, there exists $\alpha_1,\ldots,\alpha_n \in \mathbb{R}$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Define $L : V \to \mathbb{R}^{n-k}$ as $L(\alpha_1 v_1 + \cdots + \alpha_n v_n) = (\alpha_{k+1}, \alpha_{k+2}, \ldots, \alpha_n)$. Clearly, $L(\alpha_1 v_1 + \cdots + \alpha_n v_n) = 0$ iff $(\alpha_{k+1}, \alpha_{k+2}, \ldots, \alpha_n) = 0$. That is,

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n = \alpha_1 v_1 + \cdots + \alpha_k v_k \in \text{Span}(v_1,\ldots,v_k) = U.$$

□

Theorem 4.14 (Rank nullity theorem)**.** *Let $(V,+,.)$ and $(W,\oplus,\otimes)$ be two finite dimensional vector spaces and $L : V \to W$ be a linear map. Then, $dim(Ker(L)) + dim(Im(L)) = dim(V)$.*

Proof. We know that $Ker(L)$ is a subspace of V. Thus, by Theorem 3.78, $Ker(L)$ has some basis $(v_1,\ldots,v_k)$. Thus, $(v_1,\ldots,v_k)$ is linearly independent. Therefore, by Theorem 3.69, we can extend $(v_1,\ldots,v_k)$ to a list $(v_1,\ldots,v_n)$ which form a basis of V. Let $L(v_{k+1}) = w_{k+1}$, $L(v_{k+2}) = w_{k+2}, \ldots$ $L(v_n) = w_n$. As $(v_1,\ldots,v_n)$ is a basis of V, given any vector $v \in V$ there exists $(\alpha_1,\ldots,\alpha_n) \in \mathbb{R}^n$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. As L is linear,

$$\begin{aligned} L(v) &= (\alpha_1 \otimes L(v_1)) \oplus \cdots \oplus (\alpha_n \otimes L(v_n)) \\ &= (\alpha_1 \otimes 0) \oplus \cdots \oplus (\alpha_k \otimes 0) \oplus (\alpha_{k+1} \otimes w_{k+1}) \oplus \cdots \oplus (\alpha_n \otimes w_n) \\ &= (\alpha_{k+1} \otimes w_{k+1}) \oplus \cdots \oplus (\alpha_n \otimes w_n) . \end{aligned}$$

As v was arbitrary, this implies that $Im(L) = \text{Span}(w_{k+1},\ldots,w_n)$. Moreover, if $(\alpha_{k+1} \otimes w_{k+1}) \oplus \cdots \oplus (\alpha_n \otimes w_n) = 0$, then, $L(\alpha_{k+1} v_{k+1} + \cdots + \alpha_n v_n) = (\alpha_{k+1} \otimes w_{k+1}) \oplus \cdots \oplus (\alpha_n \otimes w_n) = 0$. That is $\alpha_{k+1} v_{k+1} + \cdots + \alpha_n v_n \in Ker(L)$. As $(v_1,\ldots,v_k)$ is a basis of $Ker(L)$, there exists $(\alpha_1,\ldots,\alpha_k) \in \mathbb{R}^k$ such that $\alpha_1 v_1 + \cdots + \alpha_k v_k = \alpha_{k+1} v_{k+1} + \cdots + \alpha_n v_n$. In other words, $\alpha_1 v_1 + \cdots + \alpha_k v_k + (-\alpha_{k+1}) v_{k+1} + \cdots + (-\alpha_n) v_n = 0$. As $(v_1,\ldots,v_n)$ is linearly independent, this would imply that $\alpha_i = 0$ for all $i \in \{1,\ldots,n\}$. Thus, in particular $(\alpha_{k+1},\ldots,\alpha_n) = 0$. Hence, $(w_{k+1},\ldots,w_n)$ is linearly independent. Therefore, $(w_{k+1},\ldots,w_n)$ is a basis for $Im(L)$. Thus, $dim(Ker(L)) + dim(Im(L)) = k + (n-k) = n = dim(V)$. □

Theorem 4.15. *Let $(V,+,.)$ and $(W,\oplus,\otimes)$ be two finite dimensional vector spaces and $L : V \to W$ be a linear map. If $(v_1,\ldots,v_k)$ is linearly independent in V and L is injective, then $(L(v_1),\ldots,L(v_k))$ is linearly independent in W.*

Proof. Assume $\alpha_1 L(v_1) + \cdots + \alpha_k L(v_k) = 0$. Then,

$$L(\alpha_1 v_1 + \cdots + \alpha_k v_k) = (\alpha_1 \otimes L(v_1)) \oplus \cdots \oplus (\alpha_k \otimes L(v_k)) = 0$$

Thus, $\alpha_1 v_1 + \cdots + \alpha_k v_k \in Ker(L) = \{0\}$ as L is injective. As $(v_1, \ldots, v_k)$ is linearly independent, this would imply that $(\alpha_1, \ldots, \alpha_k) = 0$. Thus, $(L(v_1), \ldots, L(v_k))$ is linearly independent. □

Theorem 4.16. *Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two finite dimensional vector spaces and $L : V \to W$ be a linear map. If $Span(v_1, \ldots, v_k) = V$ and and L is surjective, then $Span(L(v_1), \ldots, L(v_k)) = W$.*

Proof. As L is surjective, given any $w \in W$, there exists a $v \in V$ such that $L(v) = w$. As $\text{Span}(v_1, \ldots, v_k) = V$, there exists $(\alpha_1, dots, \alpha_k) \in \mathbb{R}^k$ such that $v = \alpha_1 v_1 + \cdots + \alpha_k v_k$. Thus,

$$w = L(v) = L(\alpha_1 v_1 + \cdots + \alpha_k v_k) = (\alpha_1 \otimes L(v_1)) \oplus \cdots \oplus (\alpha_k \otimes L(v_k)) \in \text{Span}(L(v_1), \ldots, L(v_k)).$$

As $w \in W$ was arbtitrary, $\text{Span}(L(v_1), \ldots, L(v_k)) = W$. □

Theorem 4.17. *Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two finite dimensional vector spaces and $L : V \to W$ be a linear map. If $dim(V) > dim(W)$, then L is not injective.*

Proof. The rank nullity theorem tells us that $dim(Ker(L)) = dim(V) - dim(Im(L))$. But, by Theorem 3.78, $dim(Im(L)) \leq dim(W)$. Thus,

$$dim(Ker(L)) = dim(V) - dim(Im(L)) \geq dim(V) - dim(W) > 0.$$

Therefore, $Ker(L) \neq \{0\}$ and thus by Theorem 4.8, L is not injective. □

Theorem 4.18. *A system of n equations in m unknowns*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_m &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_m &= 0 \\ &. \\ &. \\ &. \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= 0 \end{aligned}$$

has a non-trivial solution if $m > n$.

Proof. Define $L : \mathbb{R}^m \to \mathbb{R}^n$ as

$$L(x_1, \ldots, x_n) = (a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_m, \ldots, a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m).$$

Notice that the set of solutions of the system is equal to $Ker(L)$. But, the previous theorem tells us that L cannot be injective if $m > n$. Thus, $Ker(L) \neq \{0\}$ and hence we have non-trivial solutions. □

Theorem 4.19. *Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two finite dimensional vector spaces and $L : V \to W$ be a linear map. If $dim(V) < dim(W)$, then L is not surjective.*

Proof. The rank nullity theorem tells us that $dim(Im(L)) = dim(V) - dim(Ker(L)) \leq dim(V) < dim(W)$. Thus, $Im(L) \neq W$ and hence L is not surjective. □

4.2. Isomorphisms

Theorem 4.20. *Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two vector spaces and $L : V \to W$ be a bijective linear map. Then, L^{-1} is a linear map.*

Proof. Let $w_1, w_2 \in W$. Let $L^{-1}(w_1) = v_1$ and $L^{-1}(w_2) = v_2$. That is, $L(v_1) = w_1$ and $L(v_2) = w_2$. Thus,

$$\begin{aligned} L^{-1}(w_1 + w_2) &= L^{-1}(L(v_1) + L(v_2)) = L^{-1}(L(v_1 + v_2)) \\ &= v_1 + v_2 = L^{-1}(w_1) + L^{-1}(w_2) \end{aligned}$$

Let $w \in W$ and let $v = L^{-1}(w)$. That is $L(v) = w$. Thus,

$$L^{-1}(\alpha \otimes w) = L^{-1}(\alpha \otimes L(v)) = L^{-1}(L(\alpha.v)) = \alpha.v = \alpha.L^{-1}(w)$$

□

Definition 4.21. Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two vector spaces. A bijective linear map $L : V \to W$ is called an isomorphism. We say V and W are isomorphic if there exists an isomorphism $L : V \to W$.

Example 4.22. Let $\mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ be the vector space of linear functions from $\mathbb{R}^3$ to $\mathbb{R}^2$. Then, $\mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$ is isomorphic to $M_2^3(\mathbb{R})$.

Given a linear map $L : \mathbb{R}^3 \to \mathbb{R}^2$, let $L(e_i) = (a_{1i}, a_{2i})$ for i equal to $1, 2, 3$. Define $\Phi : \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2) \to M_2^3(\mathbb{R})$ as

$$\Phi(L) = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}.$$

We will first show that Φ is a linear map. Let $L, L' \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^2)$. Let $L(e_i) = (a_{1i}, a_{2i})$ for i equal to $1, 2, 3$ and $L'(e_i) = (a'_{1i}, a'_{2i})$ for i equal to $1, 2, 3$. Then, $(L + L')(e_i) = L(e_i) + L'(e_i) = (a_{1i} + a'_{1i}, a_{2i} + a'_{2i})$

$$\begin{aligned} \Phi(L + L') &= \begin{bmatrix} a_{11} + a'_{11} & a_{12} + a'_{12} & a_{13} + a'_{13} \\ a_{21} + a'_{21} & a_{22} + a'_{22} & a_{23} + a'_{23} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} + \begin{bmatrix} a'_{11} & a'_{12} & a'_{13} \\ a'_{21} & a'_{22} & a'_{23} \end{bmatrix} = \Phi(L) + \Phi(L') \end{aligned}$$

Similarly, $(\alpha L)(e_i) = \alpha L(e_i) = \alpha(a_{1i}, a_{2i}) = (\alpha a_{1i}, \alpha a_{2i})$. Thus,

$$\Phi(\alpha L) = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \alpha a_{13} \\ \alpha a_{21} & \alpha a_{22} & \alpha a_{23} \end{bmatrix} = \alpha \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = \alpha \Phi(L).$$

Moreover, $\Phi(L)$ is the zero matrix iff $L(e_i) = (0, 0, 0)$ for $i = 1, 2, 3$. That is, $L(x, y, z) = xL(1, 0, 0) + yL(0, 1, 0) + zL(0, 0, 1) = x(0, 0, 0) + y(0, 0, 0) + z(0, 0, 0) = (0, 0, 0)$. Thus, Φ is injective. Finally, given any matrix $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$, Define $L : \mathbb{R}^3 \to \mathbb{R}^2$ is defined as $L(x, y, z) = (a_{11}x + a_{12}y + a_{13}z, a_{21}x + a_{22}y + a_{23}z)$. Then, $\Phi(L) = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$. Hence Φ is surjective. As Φ is linear and bijective, it is an isomorphism.

From Theorem 4.19 and Theorem 4.17, we have the following result

Theorem 4.23. *Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two finite dimensional vector spaces. If V and W are isomorphic, then $dim(V) = dim(W)$.*

Interestingly, its converse is also true

Theorem 4.24. *Let $(V, +, .)$ and $(W, \oplus, \otimes)$ be two finite dimensional vector spaces. If $dim(V) = dim(W)$, then V and W are isomorphic.*

Proof. Let $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_n)$ be a basis for V and W respectively. Define

$$L(\alpha_1 v_1 + \cdots + \alpha_n v_n) = (\alpha_1 \otimes w_1) \oplus \cdots \oplus (\alpha_n \otimes w_n).$$

Clearly L is linear. Injectivity of L follows from the linear indepdendence of $(w_1, \ldots, w_n)$ and surjectivity of L follows from the fact that $\text{Span}(w_1, \ldots, w_n) = W$. ☐

Exercise 4.25. Find an isomorphism from $\text{span}(\{(1,2,3),(4,5,6),(13,14,15)\})$ to $\mathbb{R}^2$ (and prove it is an isomorphism).

Exercise 4.26. Find an isomorphism from $\text{span}(\{(1,2,3),(4,5,6),(13,14,13)\})$ to $\mathbb{R}^3$ (and prove it is an isomorphism).

Exercise 4.27. A function $f : \mathbb{R}^2 \to \mathbb{R}$ is called a Lakhan functions if f is linear and $f(1,2) = 4$ and $f(4,2) = 1$. Show that there is a unique Lakhan function and find it.

Exercise 4.28. Let $\mathcal{P}(3)$ be the vector space of all polynomials of degree less than or equal to 3. Show that $\mathcal{P}(3)$ is isomorphic to $\mathbb{R}^4$ by explicitly constructing an isomorphism $L : \mathcal{P}(3) \to \mathbb{R}^4$.

Exercise 4.29. Let $U = \{f \in \mathcal{F}(\mathbb{N}, \mathbb{R}) \mid f(n) = 0 \ \forall n \geq 10\}$. Show that U is isomorphic to $\mathbb{R}^9$ by explicitly constructing an isomorphism $L : U \to \mathbb{R}^9$.

Exercise 4.30. Suppose V is a finite dimensional vector space and U its subspace. Let $f : U \to W$ be a linear map. Show that you can find a linear map $F : V \to W$ such that $F|_U = f$. Is the function F unique?

Exercise 4.31. Let $(V, +, .)$ be a vector space and U a subspace of V. Does there exist a vector space $(W, \oplus, \otimes)$ and a linear map $L : W \to V$ such that $Im(L) = U$?

4.3. Matrix associated to a linear map

Definition 4.32. Let V and W be finite dimensional vector spaces and let $B_1 = (v_1, \ldots, v_m)$ and $B_2 = (w_1, \ldots, w_n)$ be basis of V and W respectively. Further, let $L : V \to W$, be a linear map. Notice that as $L(v_j) \in W$, we can find constants a_{ij} for $i \in 1, \ldots, n$ such that $L(v_j) = \sum_{i=1}^{n} a_{ij} w_i$. The matrix

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ a_{n1} & a_{n_2} & \ldots & a_{nm} \end{bmatrix}$$

is called the matrix of the linear transformation L with respect to the basis B_1 on V and B_2 on W and is denoted by $[L]_{B_1}^{B_2}$.

To avoid the cumbersome notation, $\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ a_{n1} & a_{n_2} & \ldots & a_{nm} \end{bmatrix}$, we would often represent an $n \times m$ matrix whose ij-th entry is a_{ij} as $(a_{ij})_{n \times m}$. This allows us to express the product of two matrices also in a convenient form. Let A be an $m \times n$ matrix (a_{ij}) and B be an $n \times p$ matrix, then $AB = C$ is a $m \times p$ matrix whose ik-th entry is $c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$. As we saw earlier, the matrix multiplication is so defined to ensure the following theorem is true.

Example 4.33. Consider the basis $B_1 = (1, x, x^2, x^3)$ of $\mathcal{P}(3)$ and basis $B_2 = (1, x, x^2)$ of $\mathcal{P}(2)$. Consider the function $D : \mathcal{P}(3) \to \mathcal{P}(2)$ defined as $D(a_0 + a_1x + a_2x^2 + a_3x^3) = a_1 + 2a_2x + 3a_3x^2$. Then, $D(1) = 0$, $D(x) = 1$, $D(x^2) = 2x$, and $D(x^3) = 3x^2$. Therefore,

$$[D]_{B_1}^{B^2} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Example 4.34. Consider the basis $B_1 = (1, 1 + x, 1 + x + x^2, 1 + x + x^3)$ of $\mathcal{P}(3)$ and basis $B_2 = (1 + x, 1 - x, x^2)$ of $\mathcal{P}(2)$. Consider the function $D : \mathcal{P}(3) \to \mathcal{P}(2)$ defined as $D(a_0 + a_1x + a_2x^2 + a_3x^3) = a_1 + 2a_2x + 3a_3x^2$. Then, $D(1) = 0$, $D(1+x) = 0.5(1+x) + 0.5(1-x)$, $D(1+x+x^2) = 1+2x = 1.5(1+x) - 0.5(1-x)$, and $D(1+x+x^2+x^3) = 1+2x+3x^2 = 1.5(1+x) - 0.5(1-x) + 3x^2$. Therefore,

$$[D]_{B_1}^{B^2} = \begin{bmatrix} 0 & 0.5 & 1.5 & 1.5 \\ 0 & 0.5 & 0.5 & 0.5 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Theorem 4.35. *Let U, V, and W be finite-dimensional vector spaces and let $B_1 = (u_1, \ldots, u_m)$, $B_2 = (v_1, \ldots, v_n)$, and $B_3 = (w_1, \ldots, w_p)$ be basis of U, V, and W respectively. Further, let $S : U \to V$ and $T : V \to W$ be two linear maps. Then,*

$$[T \circ S]_{B_1}^{B_3} = [T]_{B_2}^{B_3}[S]_{B_1}^{B_2}.$$

Proof. Let $[T]_{B_2}^{B_3} = (a_{ij})$ and let $[S]_{B_1}^{B_2} = (b_{jk})$. Thus,

$$\begin{aligned} T \circ S(u_k) = T(S(u_k)) &= T\left(\sum_{j=1}^{n} b_{jk}v_j\right) \\ &= \sum_{j=1}^{n} b_{jk}T(v_j) = \sum_{j=1}^{n} b_{jk}\left(\sum_{i=1}^{p} a_{ij}w_i\right) \\ &= \sum_{i=1}^{p}\left(\sum_{k=1}^{n} a_{ij}b_{jk}\right)w_i \end{aligned}$$

Therefore if $[T \circ S]_{B_1}^{B_3} = (c_{ij})$, then $c_{ij} = \sum_{k=1}^{n} a_{ij}b_{kj}$, which is also the ij-th entry of $[T]_{B_2}^{B_3}[S]_{B_1}^{B_2}$. Hence, $[T \circ S]_{B_1}^{B_3} = [T]_{B_2}^{B_3}[S]_{B_1}^{B_2}$. □

Example 4.36. Let $U = \mathcal{P}(3)$, $V = M_2^3(\mathbb{R})$, and $W = \mathbb{R}^2$. Let A_{ij} be the matrix whose ij-th entry is 1 and all other entries are 0. Let $B_1 = (1, 1 + x, 1 + x + x^2)$, $B_2 = (A_{11}, A_{12}, A_{1,3}, A_{21}, A_{22}, A_{23})$, and $B_3 = ((1, 0), (0, 1))$. Finally, let $S : U \to V$ be defined as

$$S(a_0 + a_1x + a_2x^2) = \begin{bmatrix} a_0 & a_0 + a_1 & a_1 + a_2 \\ 0 & a_0 + a_1 + a_2 & 3a_0 \end{bmatrix}$$

and let $T : V \to W$ be defined as

$$T\left(\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}\right) = (a_{11} + a_{12} + a_{13}, a_{21} + a_{22} + a_{23}).$$

Note that $T(A_{11}) = (1, 0)$, $T(A_{12}) = (1, 0)$, $T(A_{13}) = (1, 0)$, $T(A_{21}) = (0, 1)$, $T(A_{22}) = (0, 1)$, and $T(A_{23}) = (0, 1)$. Therefore,

$$[T]_{B_2}^{B_3} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Note that $S(1) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix} = A_{11} + A_{12} + 0A_{13} + 0A_{21} + A_{22} + 3A_{23}$. Similarly, $S(1+x) = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 3 \end{bmatrix} = A_{11} + 2A_{12} + A_{13} + 0A_{21} + 2A_{22} + 3A_{23}$. Finally, $S(1+x+x^2) = \begin{bmatrix} 1 & 2 & 2 \\ 0 & 3 & 3 \end{bmatrix} = A_{11} + 2A_{12} + 2A_{13} + 0A_{21} + 3A_{22} + 3A_{23}$. Thus,

$$[S]_{B_1}^{B_2} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 1 & 2 & 3 \\ 3 & 3 & 3 \end{bmatrix}.$$

Then

$$[T]_{B_2}^{B_3}[S]_{B_1}^{B_2} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 1 & 2 & 3 \\ 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}.$$

On the other hand, $T(S(1)) = T\left(\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 3 \end{bmatrix}\right) = (2,4)$. Similarly, $T(S(1+x)) = T\left(\begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 3 \end{bmatrix}\right) = (4,5)$ and $T(S(1+x+x^2)) = T\left(\begin{bmatrix} 1 & 2 & 2 \\ 0 & 3 & 3 \end{bmatrix}\right) = (5,6)$. Thus,

$$[T \circ S]_{B_1}^{B_3} = \begin{bmatrix} 2 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix} = [T]_{B_2}^{B_3}[S]_{B_1}^{B_2}.$$

4.4. Change of coordinates - an example

Let R be the reflection about the line passing through $(0,0)$ and (a,b). Thus, $R(a,b) = (a,b)$ and $R(-b,a) = (b,-a) = -(-b,a)$. We feel intuitively that if we align our head along the vectors $(a,b), (-b,a)$, then the linear transformation would look like the map $L(x,y) = (x,-y)$. In this section, we make this idea precise. Consider two basis of $\mathbb{R}^2$ namely

$$B_1 = ((1,0),(0,1)) \text{ and } B_2 = ((a,b),(-b,a)).$$

Then, $(1,0) = x(a,b) + y(-b,a) = (xa - yb, xb + ya)$. That is, $xa - yb = 1$ and $xb = -ya$. That is $xab - yb^2 = b$ or $-ya^2 - yb^2 = b$. Thus,

$$(1,0) = \frac{a}{a^2+b^2}(a,b) + \frac{-b}{a^2+b^2}(-b,a).$$

Similarly, $(0,1) = (xa - yb, xa + yb)$ implies that $xa = yb$ and $xb + ya = 1$. Substituting, we get

$$(0,1) = \frac{b}{a^2+b^2}(a,b) + \frac{a}{a^2+b^2}(-b,a).$$

Thus, it is easy to see that $[I]_{B_1}^{B_2} = \begin{bmatrix} \frac{a}{a^2+b^2} & \frac{b}{a^2+b^2} \\ \frac{-b}{a^2+b^2} & \frac{a}{a^2+b^2} \end{bmatrix}$ and $I_{B_2}^{B_1} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ and they are inverses to each other. Further,

$$\begin{aligned} R(1,0) &= R\left(\frac{a}{a^2+b^2}(a,b) + \frac{-b}{a^2+b^2}(-b,a)\right) = R\left(\frac{a}{a^2+b^2}(a,b)\right) + R\left(\frac{-b}{a^2+b^2}(-b,a)\right) \\ &= \frac{a}{a^2+b^2}R\,(a,b) + \frac{-b}{a^2+b^2}R\,(-b,a) = \frac{a}{a^2+b^2}\,(a,b) + \frac{-b}{a^2+b^2}\,(b,-a) \\ &= \frac{a}{a^2+b^2}\,(a(1,0)+b(0,1)) + \frac{-b}{a^2+b^2}\,(b(1,0)+a(0,1)) \\ &= \frac{a^2-b^2}{a^2+b^2}(1,0) + \frac{2ab}{a^2+b^2}(0,1) \end{aligned}$$

Similarly,

$$\begin{aligned} R(0,1) &= R\left(\frac{b}{a^2+b^2}(a,b) + \frac{a}{a^2+b^2}(-b,a)\right) = R\left(\frac{b}{a^2+b^2}(a,b)\right) + R\left(\frac{a}{a^2+b^2}(-b,a)\right) \\ &= \frac{b}{a^2+b^2}R\,(a,b) + \frac{a}{a^2+b^2}R\,(-b,a) = \frac{b}{a^2+b^2}\,(a,b) + \frac{a}{a^2+b^2}\,(b,-a) \\ &= \frac{b}{a^2+b^2}\,(a(1,0)+b(0,1)) + \frac{a}{a^2+b^2}\,(b(1,0)-a(0,1)) \\ &= \frac{2ab}{a^2+b^2}(0,1) + \frac{b^2-a^2}{a^2+b^2}. \end{aligned}$$

Thus,

$$\begin{aligned} R(x,y) &= R(x(1,0)+y(0,1)) = xR(1,0) + yR(0,1) \\ &= x\left[\frac{a^2-b^2}{a^2+b^2}(1,0) + \frac{2ab}{a^2+b^2}(0,1)\right] + y\left[\frac{2ab}{a^2+b^2}(0,1) + \frac{b^2-a^2}{a^2+b^2}\right] \\ &= \left(\left(\frac{a^2-b^2}{a^2+b^2}\right)x + \left(\frac{2ab}{a^2+b^2}\right)y, \left(\frac{2ab}{a^2+b^2}\right)x + \left(\frac{b^2-a^2}{a^2+b^2}\right)y\right) \end{aligned}$$

Hence, the matrix of this linear transformation is $\begin{bmatrix} \frac{a^2-b^2}{a^2+b^2} & \frac{2ab}{a^2+b^2} \\ \frac{2ab}{a^2+b^2} & \frac{b^2-a^2}{a^2+b^2} \end{bmatrix}$. From now on, we would be a little bit more specific and say that this is the matrix of R with respect to the basis B_1 and denote it as $[R]_{B_1}$. Further notice that

$$[R]_{B_1} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \frac{a}{a^2+b^2} & \frac{b}{a^2+b^2} \\ \frac{-b}{a^2+b^2} & \frac{a}{a^2+b^2} \end{bmatrix}.$$

Take any point $(x,y) \in \mathbb{R}^2$, that is,

$$\begin{aligned} (x,y) &= x(0,1) + y(1,0) \\ &= x\left(\frac{a}{a^2+b^2}(a,b) + \frac{-b}{a^2+b^2}(-b,a)\right) + y\left(\frac{b}{a^2+b^2}(a,b) + \frac{a}{a^2+b^2}(-b,a)\right) \\ &= \left[\left(\frac{a}{a^2+b^2}\right)x + \left(\frac{b}{a^2+b^2}\right)y\right](a,b) + \left[\left(\frac{-b}{a^2+b^2}\right)x + \left(\frac{a}{a^2+b^2}\right)y\right](-b,a) \end{aligned}$$

As $\begin{bmatrix} \frac{xa+yb}{a^2+b^2} \\ \frac{-xb+ya}{a^2+b^2} \end{bmatrix}$ represents the coefficients when you express (x,y) as a linear combination of (a,b) and $(-b,a)$, we say it is the coordinates of (x,y) in basis B_2. And we write $[(x,y)]_{B_2} = \begin{bmatrix} \frac{xa+yb}{a^2+b^2} \\ \frac{-xb+ya}{a^2+b^2} \end{bmatrix}$.

We further notice that

$$\begin{bmatrix} \frac{xa+yb}{a^2+b^2} \\ \frac{-xb+ya}{a^2+b^2} \end{bmatrix} = \begin{bmatrix} \frac{a}{a^2+b^2} & \frac{b}{a^2+b^2} \\ \frac{-b}{a^2+b^2} & \frac{a}{a^2+b^2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Thus, the matrix $[I]_{B_1}^{B_2}$ allows us to convert the coordinates of (x,y) in terms of B_1 (written as $[(x,y)]_{B_1}$) to coordinates of (x,y) in terms of B_2 (written as $[(x,y)]_{B_2}$). Thus, the matrix is called a change of coordinate matrix. Further notice that

$$\begin{bmatrix} \frac{a}{a^2+b^2} & \frac{b}{a^2+b^2} \\ \frac{-b}{a^2+b^2} & \frac{a}{a^2+b^2} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} \frac{a}{a^2+b^2} & \frac{b}{a^2+b^2} \\ \frac{-b}{a^2+b^2} & \frac{a}{a^2+b^2} \end{bmatrix} \begin{bmatrix} -b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Hence, the matrix $[I]_{B_1}^{B_2}$ sends B_2 to B_1 while the matrix $[I]_{B^2}^{B_1}$ sends B_1 to B_2.

4.5. Change of coordinates

Definition 4.37. Let V be a finite-dimensional vector space and let $B = (v_1, \dots, v_n)$ be a basis of V. Given any vector $v \in V$. there exists $(\alpha_1, \dots, \alpha_n)$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Then, we say that the coordinates of v with respect to the basis B, denoted as $[v]_B$ is $(\alpha_1, \dots, \alpha_n)$ or $\begin{bmatrix} \alpha_1 \\ . \\ . \\ . \\ \alpha_n \end{bmatrix}$.

Example 4.38. Let $V = \mathbb{R}^2$, $B_1 = ((1,1),(1,-1))$, and $B_2 = ((1,2),(3,4))$. Given any element $(x,y) \in \mathbb{R}^2$, note that $(x,y) = (\frac{x+y}{2})(1,1) + (\frac{x-y}{2})(1,-1)$. Thus, $[(x,y)]_{B_1} = \begin{bmatrix} \frac{x+y}{2} \\ \frac{x-y}{2} \end{bmatrix}$. To find $[(x,y)]_{B_2}$, we solve $(x,y) = \alpha(1,2) + \beta(3,4) = (\alpha + 3\beta, 2\alpha + 4\beta)$. That is $2\beta = 2x - y$ or $\beta = x - \frac{y}{2}$. Thus, $\alpha + 3x - 3\frac{y}{2} = x$. Hence, $(\alpha, \beta) = (\frac{3y}{2} - 2x, x - \frac{y}{2})$. Hence,

$$[(x,y)]_{B_2} = \begin{bmatrix} \frac{3y}{2} - 2x \\ x - \frac{y}{2} \end{bmatrix}.$$

To check your answer is indeed correct, note that

$$[(1,2)]_{B_2} = \begin{bmatrix} \frac{3(2)}{2} - 2(1) \\ (1) - \frac{(2)}{2} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$[(3,4)]_{B_2} = \begin{bmatrix} \frac{3(4)}{2} - 2(3) \\ (3) - \frac{(4)}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Example 4.39. Let $V = \mathbb{R}^3$, $B_1 = ((0,1,0),(0,0,1),(1,0,0))$. Then given any vector $(x,y,z) \in \mathbb{R}^2$, $(x,y,z) = x(1,0,0) + y(0,1,0) + z(0,0,1) = y(0,1,0) + z(0,0,1) + x(1,0,0)$. Thus,

$$[(x,y,z)]_{B_1} = \begin{bmatrix} y \\ z \\ x \end{bmatrix}.$$

Let $B_2 = ((1,1,-1),(1,-1,1),(0,1,0))$. Given any element $(x,y,z) \in \mathbb{R}^3$, note that $(x,y,z) = \alpha(1,1,-1) + \beta(1,-1,1) + \gamma(0,1,0) = (\alpha + \beta, \alpha - \beta + \gamma, \beta - \alpha)$. Thus, $\beta = \frac{x+z}{2}$ and $\alpha = \frac{x-z}{2}$. Therefore, $\frac{x-z}{2} - \frac{x+z}{2} + \gamma = y$. That is, $\gamma = y + z$. Hence

$$[(x,y,z)]_{B_2} = \begin{bmatrix} \frac{x-z}{2} \\ \frac{x+z}{2} \\ y + z \end{bmatrix}.$$

To check your answer is indeed correct, note that

$$[(1,1,-1)]_{B_2} = \begin{bmatrix} \frac{(1)-(-1)}{2} \\ \frac{(1)+(-1)}{2} \\ (1)+(-1) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$[(1,-1,1)]_{B_2} = \begin{bmatrix} \frac{(1)-(1)}{2} \\ \frac{(1)+(1)}{2} \\ (-1)+(1) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$[(0,1,0)]_{B_2} = \begin{bmatrix} \frac{(0)-(0)}{2} \\ \frac{(0)+(0)}{2} \\ (1)+(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Theorem 4.40. *Let V be a finite-dimensional vector space and let B_1 and B_2 be two bases. Then $[v]_{B_2} = [I]_{B_1}^{B_2}[v]_{B_1}$.*

Proof. Let $B_1 = (v_1, \ldots, v_n)$ and $B_2 = (w_1, \ldots, w_n)$. Then,

$$[v]_{B_1} = (\alpha_1, \ldots, \alpha_n) \iff v = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

Let $v_i = a_{1i}w_1 + \cdots + a_{ni}w_n$. Thus,

$$\begin{aligned} v &= \alpha_1 v_1 + \cdots + \alpha_n v_n \\ &= \alpha_1(a_{11}w_1 + \cdots + a_{n1}w_n) + \ldots, \alpha_n(a_{1n}w_1 + \cdots + a_{nn}w_n) \\ &= (\alpha_1 a_{11} + \alpha_2 a_{12} + \ldots \alpha_n a_{1n})w_1 + \cdots + (\alpha_1 a_{n1} + \alpha_2 a_{n2} + \ldots \alpha_n a_{nn})w_n \end{aligned}$$

That is,

$$[v]_{B_2} = \begin{bmatrix} \alpha_1 a_{11} + \alpha_2 a_{12} + \ldots \alpha_n a_{1n} \\ \alpha_1 a_{21} + \alpha_2 a_{22} + \ldots \alpha_n a_{2n} \\ . \\ . \\ \alpha_1 a_{n1} + \alpha_2 a_{n2} + \ldots \alpha_n a_{nn} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ \alpha_n \end{bmatrix}$$

□

Example 4.41. Let $V = \mathbb{R}^2$, $B_1 = ((1,1),(1,-1))$, and $B_2 = ((1,2),(3,4))$. Then, $(1,1) = -\frac{1}{2}(1,2) + \frac{1}{2}(3,4)$ and $(1,-1) = -\frac{7}{2}(1,2) + \frac{3}{2}(3,4)$. Thus, $[I]_{B_1}^{B_2} = \frac{1}{2}\begin{bmatrix} -1 & -7 \\ 1 & 3 \end{bmatrix}$. And,

$$[I]_{B_1}^{B_2}[(x,y)]_{B_1} = \frac{1}{2}\begin{bmatrix} -1 & -7 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} \frac{x+y}{2} \\ \frac{x-y}{2} \end{bmatrix} = \begin{bmatrix} \frac{-x-y-7x+7y}{4} \\ \frac{x+y+3x-3y}{4} \end{bmatrix} = \begin{bmatrix} \frac{3y}{2} - 2x \\ x - \frac{y}{2} \end{bmatrix} = [(x,y)]_{B_2}.$$

Example 4.42. Let $V = \mathbb{R}^3$, $B_1 = ((0,1,0),(0,0,1),(1,0,0))$ and $B_2 = ((1,1,-1),(1,-1,1),(0,1,0))$. Then,

$$\begin{aligned} (0,1,0) &= 0(1,1,-1) + 0.(0,0,1) + 1(0,1,0) \\ (0,0,1) &= \frac{-1}{2}(1,1,-1) + \frac{1}{2}(1,-1,1) + 1(0,1,0) \\ (1,0,0) &= \frac{1}{2}(1,1,-1) + \frac{1}{2}(1,-1,1) + 0(0,1,0) \end{aligned}$$

Thus,

$$[I]_{B_1}^{B_2} = \begin{bmatrix} 0 & \frac{-1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 1 & 0 \end{bmatrix}.$$

And,

$$[I]_{B_1}^{B_2}[(x,y,z)]_{B_1} = \begin{bmatrix} 0 & \frac{-1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \\ x \end{bmatrix} = \begin{bmatrix} \frac{x-z}{2} \\ \frac{x+z}{2} \\ y+z \end{bmatrix} = [(x,y,z)]_{B_2}.$$

Definition 4.43. Let V and W be finite-dimensional vector spaces and let $B_1 = (v_1, \ldots, v_m)$ and $B_2 = (w_1, \ldots, w_m)$ be basis of V and W respectively. Finally, let $L : V \to W$ be a linear map. As $L(v_i) \in W$, there exist constants $(a_{1i}, \ldots, a_{ni})$ such that $L(v_i) = a_{1i}w_1 + \cdots + a_{ni}w_n$. Then, we say that the matrix associated to the linear map L with respect to B_1 and B_2,

$$[L]_{B_1}^{B_2} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}.$$

Remark 4.44. Let V be a finite-dimensional vector space and $B = (v_1, \ldots, v_m)$ be a basis of V. Given a linear map $L : V \to W$, we will often write $[L]_B$ instead of $[L]_B^B$.

Theorem 4.45. *Let V be finite-dimensional vector spaces and let $B_1 = (v_1, \ldots, v_n)$ and $B_2 = (w_1, \ldots, w_n)$ be two basis of V. Finally, let $L : V \to V$ be a linear map. Then,*

$$[L]_{B_2} = [I]_{B_1}^{B_2}[L]_{B_1}^{B_1}[I]_{B_2}^{B_1}.$$

Proof. Using Theorem 4.35, $[L]_{B_1}^{B_1}[I]_{B_2}^{B_1} = [L \circ I]_{B_2}^{B_1} = [L]_{B_2}^{B_1}$. Thus, $[I]_{B_1}^{B_2}[L]_{B_1}^{B_1}[I]_{B_2}^{B_1} = [I]_{B_1}^{B_2}[L]_{B_2}^{B_1}$. Once again, using Theorem 4.35, we can see that $[I]_{B_1}^{B_2}[L]_{B_2}^{B_1} = [I \circ L]_{B_2}^{B_2} = [L]_{B_2}^{B_2}$. □

Example 4.46. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be the function $L(x,y) = (4x+3y, 5x+7y)$. Further, let $B_1 = ((1,1),(1,-1))$ and $B_2 = ((1,2),(3,4))$. We will find $[L]_{B_1}^{B_1}$, $[L]_{B_1}^{B_2}$, $[L]_{B_2}^{B_1}$, and $[L]_{B_2}^{B_2}$.

Notice

$$L(1,1) = (4+3, 5+7) = (7,12) = \frac{19}{2}(1,1) + \frac{-5}{2}(1,-1)$$
$$L(1,-1) = (4-3, 5-7) = (1,-2) = \frac{-1}{2}(1,1) + \frac{3}{2}(1,-1)$$

Thus,

$$[L]_{B_1}^{B_1} = \frac{1}{2}\begin{bmatrix} 19 & -1 \\ -5 & 3 \end{bmatrix}.$$

Similarly, as

$$L(1,1) = (4+3, 5+7) = (7,12) = 4(1,2) + 1(3,4)$$
$$L(1,-1) = (4-3, 5-7) = (1,-2) = -5(1,2) + 2(3,4)$$

we have

$$[L]_{B_1}^{B_2} = \begin{bmatrix} 4 & -5 \\ 1 & 2 \end{bmatrix}.$$

Similarly, as

$$L(1,2) = (4+6, 5+14) = (10,19) = \frac{29}{2}(1,1) + \frac{-9}{2}(1,-1)$$
$$L(3,4) = (12+12, 15+28) = (24,43) = \frac{67}{2}(1,1) + \frac{-19}{2}(1,-1)$$

we have

$$[L]_{B_2}^{B_1} = \frac{1}{2}\begin{bmatrix} 29 & 67 \\ -9 & -19 \end{bmatrix}.$$

Similarly, as

$$L(1,2) = (4+6, 5+14) = (10,19) = \frac{17}{2}(1,2) + \frac{1}{2}(3,4)$$
$$L(3,4) = (12+12, 15+28) = (24,43) = \frac{33}{2}(1,2) + \frac{5}{2}(3,4)$$

we have

$$[L]_{B_2}^{B_2} = \frac{1}{2}\begin{bmatrix} 17 & 1 \\ 33 & 5 \end{bmatrix}.$$

Exercise 4.47. Let $B_1 = ((1,0),(0,1))$ and $B_2 = ((1,1),(1,2))$. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map $L(x,y) = (6x+2y, 3y-x)$.

(1) Find $[(x,y)]_{B_2}$ and $[I]_{B_1}^{B_2}$.

(2) Check that $[(x,y)]_{B_2} = [I]_{B_1}^{B_2}[(x,y)]_{B_1}$.

(3) Find the matrix $[L]_{B_2}^{B_2}$

(4) Verify that $[L]_{B_2}^{B_2} = [I]_{B_1}^{B_2}[L]_{B_1}^{B_1}[I]_{B_2}^{B_1}$.

Exercise 4.48. Let $B_1 = ((1,0,0),(0,1,0),(0,0,1))$ and $B_2 = ((1,1,0),(1,0,1),(0,1,1))$. Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear map $L(x,y,z) = (3x+z, 3y-x, z)$.

(1) Find $[(x,y,z)]_{B_2}$ and $[I]_{B_1}^{B_2}$.

(2) Check that $[(x,y,z)]_{B_2} = [I]_{B_1}^{B_2}[(x,y,z)]_{B_1}$.

(3) Find the matrix $[L]_{B_2}^{B_2}$

(4) Verify that $[L]_{B_2}^{B_2} = [I]_{B_1}^{B_2}[L]_{B_1}^{B_1}[I]_{B_2}^{B_1}$.

Exercise 4.49. Find the change of coordinates matrix $[I]_B^C$ for the ordered basis B and C of the vector space V.

(1) $V = \mathcal{P}(1)$, $B = (7-4x, 5x)$, and $C = (1-2x, 2+x)$

(2) $V = \mathcal{P}(2)$, $B = (5-3x, 1, 1+2x^2)$, and $C = (1-x+x^2, 1-x, 4)$

(3) $V = M_2(\mathbb{R})$ (2×2 matrices), $B = \left(\begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}, \begin{bmatrix} 3 & 5 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 3 & 0 \end{bmatrix}, \begin{bmatrix} -2 & -4 \\ 0 & 0 \end{bmatrix}\right)$, and $C = \left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right)$

Exercise 4.50. Let (v_1, v_2, v_3) be a basis of $\mathbb{R}^3$ and $L : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear map such that $[L]_{(v_1,v_2,v_3)} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$. Find

(1) $[L]^{(v_1,v_2,v_3)}_{(v_2,v_1,v_3)}$.

(2) $[L]^{(v_2,v_1,v_3)}_{(v_1,v_2,v_3)}$.

(3) $[L]^{(v_2,v_1,v_3)}_{(v_2,v_1,v_3)}$.

(4) $[L]^{(v_1,v_2,v_3)}_{(v_1+v_2,v_2,v_3)}$

(5) $[L]^{(v_1+v_2,v_2,v_3)}_{(v_1,v_2,v_3)}$

(6) $[L]^{(v_1+v_2,v_2,v_3)}_{(v_1+v_2,v_2,v_3)}$

Eigenvalues and Eigenvectors

5.1. Diagonalisation

Earlier we observed that for the linear map $R : \mathbb{R}^2 \to \mathbb{R}^2$ defined as

$$R(x, y) = \left(\left(\frac{a^2 - b^2}{a^2 + b^2} \right) x + \left(\frac{2ab}{a^2 + b^2} \right) y, \left(\frac{2ab}{a^2 + b^2} \right) x + \left(\frac{b^2 - a^2}{a^2 + b^2} \right) y \right)$$

we could choose a basis (namely $B_2 = ((a, b), (-b, a))$) such that $[L]_{B_2}$ is diagonal. It would be awesome if we could as diagonal matrices are easy to work with - all computations are easy.

Let V be a finite-dimensional vector space and $L : V \to V$ a linear map. Assume there exists a basis $B = (v_1, \ldots, v_n)$ with respect to which the matrix of the linear transformation is diagonal, say

$$[L]_B = \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 & 0 \\ . & . & . & \ldots & . & . \\ . & . & . & \ldots & . & . \\ 0 & 0 & 0 & \ldots & 0 & \lambda_n \end{bmatrix}$$

Then by the definition of a matrix associated to a linear map, we have $L(v_i) = \lambda_i v_i$.

Definition 5.1. Let V be a finite-dimensional vector space and $L : V \to V$ be a linear map. A non-zero vector $v \in V$ is said to be an eigenvector iff there exists a $\lambda \in \mathbb{R}$ such that $L(v) = \lambda v$. The value λ is called an eigenvalue of L.

Example 5.2. For the linear map $R : \mathbb{R}^2 \to \mathbb{R}^2$ defined as

$$R(x, y) = \left(\left(\frac{a^2 - b^2}{a^2 + b^2} \right) x + \left(\frac{2ab}{a^2 + b^2} \right) y, \left(\frac{2ab}{a^2 + b^2} \right) x + \left(\frac{b^2 - a^2}{a^2 + b^2} \right) y \right)$$

(a, b) is an eigenvector corresponding to the eigenvalue 1 and $(-b, a)$ is an eigenvector corresponding to the eigenvalue -1.

Theorem 5.3. *Let V be a finite-dimensional vector space and $L : V \to V$ be a linear map. Let $v_1, \ldots, v_k$ be nonzero vectors such that $L(v_i) = \lambda_i v_i$ with $\lambda_i \neq \lambda_j$ if $i \neq j$. Then $(v_1, \ldots, v_k)$*

are linearly independent. This is often informally stated as: eigenvalues corresponding to distinct eigenvalues are linearly independent.

Proof. We will prove this statement using induction. Notice that P(1) is obvious as any nonzero vector is linearly independent. Assume that P(k) is true. Now we will prove P(k+1). Let $v_1, \ldots, v_{k+1}$ be nonzero vectors such that $L(v_i) = \lambda_i v_i$ with $\lambda_i \neq \lambda_j$ if $i \neq j$. Assume $\alpha_1 v_1 + \cdots + \alpha_{k+1} v_{k+1} = 0$. Thus, $\lambda_{k+1}\alpha_1 v_1 + \cdots + \lambda_{k+1}\alpha_{k+1}v_{k+1} = 0$ and $0 = L(0) = L(\alpha_1 v_1 + \cdots + \alpha_{k+1}v_{k+1}) = \lambda_1\alpha_1 v_1 + \cdots + \lambda_{k+1}\alpha_{k+1}v_{k+1}$. Subtracting the two equations, we have

$$\alpha_1(\lambda_1 - \lambda_{k+1})v_1 + \cdots + \alpha_k(\lambda_k - \lambda_{k+1})v_k = 0.$$

But, by the induction hypothesis, $(v_1, \ldots, v_k)$ is linearly independent. Thus, $\alpha_i(\lambda_i - \lambda_{k+1}) = 0$ for all $i \in \{1, \ldots, k\}$. But, $\lambda_i - \lambda_{k+1} \neq 0$ so, $\alpha_i = 0$ for all $i \in \{1, \ldots, k\}$. Thus, $0 = \alpha_1 v_1 + \cdots + \alpha_{k+1}v_{k+1} = \alpha_{k+1}v_{k+1}$. As, $v_{k+1} \neq 0$, $\alpha_{k+1} = 0$. Thus, $(\alpha_1, \ldots, \alpha_{k+1}) = 0$ and hence $(v_1, \ldots, v_{k+1})$ is linearly independent. □

Corollary 5.4. *Let V be a finite-dimensional vector space of dimension n and $L : V \to V$ be a linear map. Then L has at most n distinct eigenvalues.*

Thus, the problem of diagonalisation is closely tied with the problem of finding eigenvalues and eigenvectors. Notice that $L(v) = \lambda v$ iff $(L - \lambda I)(v) = 0$ (where I is the transformation $I(v) = v$). Thus, λ is an eigenvalue iff $Ker(L - \lambda I) \neq \{0\}$. Thus, we need a quick method to find if $L - \lambda I$ is injective. Recall that if $L : \mathbb{R}^2 \to \mathbb{R}^2$ is the linear map $L(x, y) = (ax + by, cx + dy)$, then L is injective iff $ad - bc \neq 0$. If $B = (e_1, e_2)$ is the standard basis $[L]_B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and the quantity $ad - bc$ is called the determinant of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Let $B = (e_1, e_2, e_3)$ is the standard basis on $\mathbb{R}^3$ We can similarly show that a linear map $L : \mathbb{R}^3 \to \mathbb{R}^3$ with $[L]_B = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ is injective iff determinant of $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ defined as $a(ei - fh) - b(di - fg) + c(dh - eg) \neq 0$.

—- Geometrically determinant captures the factor by which area or volume gets scaled: no time to complete this right now! Sorry!—

The determinant allows us to easily compute eigenvalues and hence eigenvectors. Let us look at some examples.

Example 5.5. Let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map $L(x, y) = (x + 2y, 3x + 4y)$ and $B = (e_1, e_2)$. Then, $[L]_B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. To find the eigenvalues, we look at

$$det\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = det\left(\begin{bmatrix} 1-\lambda & 2 \\ 3 & 4-\lambda \end{bmatrix}\right) = \lambda^2 - 5\lambda - 2.$$

Thus, the determinant is 0 iff

$$\lambda = \frac{5 \pm \sqrt{25+8}}{2} = \frac{5 \pm \sqrt{33}}{2}.$$

Let us now find the corresponding eigenvectors

$$\begin{bmatrix} \frac{5+\sqrt{33}}{2}x \\ \frac{5+\sqrt{33}}{2}y \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + 2y \\ 3x + 4y \end{bmatrix}$$

Thus, we get $2y = \frac{5+\sqrt{33}}{2}x - x = \frac{3+\sqrt{33}}{2}x$. Thus, $(4, 3+\sqrt{33})$ is an eigenvector. Let us confirm our answer.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 3+\sqrt{33} \end{bmatrix} = \begin{bmatrix} 10+\sqrt{33} \\ 24+\sqrt{33} \end{bmatrix} = \frac{5+\sqrt{33}}{2} \begin{bmatrix} 4 \\ 3+\sqrt{33} \end{bmatrix}.$$

Similarly, we can check that $(4, 3-\sqrt{33})$ is an eigenvector corresponding to $5-\sqrt{33}$. It is easy to see that, $B' = ((4, 3+\sqrt{33}), (4, 3-\sqrt{33}))$ is a basis of $\mathbb{R}^2$ and $[L]_{B'} = \begin{bmatrix} \frac{5+\sqrt{33}}{2} & 0 \\ 0 & \frac{5-\sqrt{33}}{2} \end{bmatrix}$

Example 5.6. Not all matrices have eigenvalues. To see an example, look at $L(x, y) = (y, -x)$. Then, $[L]_B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ and

$$det\left(\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = det\left(\begin{bmatrix} -\lambda & 1 \\ -1 & -\lambda \end{bmatrix}\right) = \lambda^2 + 1.$$

Thus, the determinant is not zero for any real number λ.

Example 5.7. Let $B = (e_1, e_2, e_3)$ be the standard basis of $\mathbb{R}^3$. Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear map such that $[L]_B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$. Then,

$$\begin{aligned} det\left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) &= det\left(\begin{bmatrix} 1-\lambda & 2 & 3 \\ 4 & 5-\lambda & 6 \\ 7 & 8 & 9-\lambda \end{bmatrix}\right) \\ &= (1-\lambda)(\lambda^2 - 14\lambda - 3) - 2(-6-4\lambda) + 3(-3+7\lambda) \\ &= -\lambda^3 + 15\lambda^2 + 18\lambda \\ &= (-\lambda)(\lambda^2 - 15\lambda - 18) \end{aligned}$$

Thus, the eigenvalues are 0, $\frac{15\pm\sqrt{297}}{2}$. Solving

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

we get $y = -2x$ and $z = x$. Thus, $(1, -2, 1)$ is an eigenvector corresponding to 0. The computations for the other eigenvectors is really tedious, so let us not do it.

Bibliography

[Axler] Axler, Sheldon. *Linear Algebra done right*. Springer, 2015.

[Bryant] Victor Bryant. *Reducing Classical Axioms*. The Mathematical Gazette, vol. 55, no. 391, 1971, pp. 38–40.

[Rigby-Wiegold] Rigby, J. F., and James Wiegold. *Independent Axioms for Vector Spaces*. The Mathematical Gazette 57, no. 399 (1973): 56–62.

[user7530] user7530 (https://math.stackexchange.com/users/7530/user7530). *Answer to "Is every axiom in the definition of a vector space necessary?"*, https://math.stackexchange.com/q/1412923

[Hartshorne] Hartshorne, Robin. *Geometry: Euclid and beyond*. Springer Science and Business Media, 2013.

[Needham] Needham, Tristan. *Visual complex analysis*. Oxford University Press, 2023.

[Hughes] John Hughes (https://math.stackexchange.com/users/114036/john-hughes), *Bijection* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *preserves collinearity* $\iff f(x) = Ax + b$, https://math.stackexchange.com/q/553151

[Lyndon] Lyndon, Roger C. *Groups and geometry*. Vol. 101. Cambridge University Press, 1985.

Geometry and Linear Algebra

Study of numbers, shapes and space are the most natural areas in mathematics as human beings interact with these. These field in turn give rise to most fields within mathematics. Modern mathematics gives emphasis to the the formalistic viewpoint and proof through deductive reasoning takes a central position. Euclidean geometry is historically one of the first instance of a rigorous deductive system of mathemaitcs and is thus historically very important. Probably, due to this importance, Euclidean geometry also forms a large part of the school curriculum at least till the tenth standard.

Euclid leaves several notions like the point and plane undefined and is taken to be self evident. But, there is no appeal to intuition as every result is a consequence of one of the axioms/postulates. In other words, these undefined words are assumed to be defined by their properties. In the formalistic viewpoint, sets are taken to be fundamental objects - defined by its properties - and everything else is defined via sets. Let us define the Cartesian plane as $\mathbb{R}^2 = \{(x, y) : x \text{ and } y \text{ are real numbers}\}$. The Cartesian plane forms (in a sense) **the** model of Euclidean geometry. The proof of the correspondence between the Euclidean (or the modern Hilbert style) view of geometry and the Cartesian view of geometry is deep and beyond the scope of this book. The interested reader can look at [**Hartshorne**] for a detailed exposition. We would just look at the Cartesian definitions of some of the most important notions and prove a few theorems.

A.1. Euclidean notions in the Cartesian plane

Definition A.1 (Point)**.** Any element of the Cartesian plane $\mathbb{R}^2$ is called a point on the plane.

Definition A.2 (Line)**.** Given real numbers $a \neq 0, b \neq 0, c, d$, a set of the form

$$\Lambda_{a,b,c,d} = \{\alpha(a, b) + (c, d) | \alpha \in \mathbb{R}\}$$

is called a line.

Exercise A.3. What is $\Lambda_{0,0,c,d}$?

Definition A.4 (Parallel)**.** Two lines are called parallel if either they are the same or are disjoint.

Some may prefer to define lines using the equation of a line. That is, a line is described by a set $L_{p,q,r} = \{(x,y) : px + qy = r\}$ where $(p,q) \neq (0,0)$. This gives us two equivalent ways of describing lines.

Lemma A.5.

$$\{L_{p,q,r} : p,q,r \in \mathbb{R}^2 \text{ and } (p,q) \neq (0,0))\} = \{\Lambda_{a,b,c,d} : a,b,c,d \in \mathbb{R}^2 \text{ and } (a,b) \neq (0,0)\}$$

Proof. Assuming $a \neq 0$,

$$\begin{aligned}\Lambda_{a,b,c,d} &= \{(\alpha a + c, \alpha b + d) : \alpha \in \mathbb{R}\} = \left\{\left(\alpha a + c, \frac{b}{a}(\alpha a + c) - \frac{bc}{a} + d\right) : \alpha \in \mathbb{R}\right\} \\ &= \left\{\left(x, \frac{b}{a}x + \left(d - \frac{bc}{a}\right)\right) : x \in \mathbb{R}\right\} = \{(x,y) \in \mathbb{R}^2 : y = \frac{b}{a}x + \left(d - \frac{bc}{a}\right)\} \\ &= \{(x,y) \in \mathbb{R}^2 : ay = bx + (ad - bc)\} = \{(x,y) \in \mathbb{R}^2 : bx - ay = bc - ad\} \\ &= L_{b,-a,bc-ad}.\end{aligned}$$

Similarly, if $b \neq 0$,

$$\begin{aligned}\Lambda_{a,b,c,d} &= \{(\alpha a + c, \alpha b + d) : \alpha \in \mathbb{R}\} = \left\{\left(\frac{a}{b}(\alpha b + d) - \frac{ad}{b} + c, \alpha b + d\right) : \alpha \in \mathbb{R}\right\} \\ &= \left\{\left(\frac{a}{b}y + \left(c - \frac{ad}{b}\right), y\right) : y \in \mathbb{R}\right\} = \{(x,y) \in \mathbb{R}^2 : x = \frac{a}{b}y + \left(c - \frac{ad}{b}\right)\} \\ &= \{(x,y) \in \mathbb{R}^2 : bx - ay = bc - ad\} = L_{b,-a,bc-ad}.\end{aligned}$$

Thus, we have proved that $\Lambda_{a,b,c,d} = L_{b,-c,bc-ad}$. Hence, we have showed that the first set in the statement of the lemma is a subset of the second. Now, let us show the other way. Assuming $q \neq 0$,

$$\begin{aligned}L_{p,q,r} &= \{(x,y) : px + qy = r\} = \left\{\left(x, \frac{-p}{q}x + \frac{r}{q}\right) : x \in \mathbb{R}\right\} \\ &= \left\{x\left(1, \frac{-p}{q}\right) + \left(0, \frac{r}{q}\right) : x \in \mathbb{R}\right\} = \Lambda_{1,\frac{-p}{q},0,\frac{r}{q}}.\end{aligned}$$

On the other hand,

$$L_{p,0,r} = \left\{(x,y) : x = \frac{r}{p}\right\} = \left\{\alpha(0,1) + \left(\frac{r}{p}, 0\right) : \alpha \in \mathbb{R}\right\} = \Lambda_{0,1,\frac{r}{p},0}.$$

□

Theorem A.6. *Given any two points on the Cartesian, there exists a line passing through the two points.*

Proof. Let (p,q) and (r,s) be two arbitrary points on the Cartesian plane. Then, the line $\mathcal{L}((p,q),(r,s)) := \{(x,y) : (s-q)x + (p-r)y + (qr - ps) = 0\}$ is the required line. To check this, we just need to verify that (p,q) and (r,s) satisfy the equation, which can be verified as follows:

$$(s-q)p + (p-r)q + qr - ps = ps - pq + pq - qr + qr - ps = 0$$

and

$$(s-q)r + (p-r)s + qr - ps = rs - qr + ps - rs + qr - ps = 0.$$

□

Theorem A.7. *Given a line $\{(x,y) : ax + by + c = 0\}$ and a point (p,q), there exists a line passing through (p,q) and parallel to $\{(x,y) : ax + by + c = 0\}$.*

Proof. Consider the line $\{(x,y) : ax+by+(-ap-bq)=0\}$. Clearly, the line passes through (p,q). Moreover, a point (x,y) belongs to both $\{(x,y) : ax+by+(-ap-bq)=0\}$ and $\{(x,y) : ax+by+c=0\}$ iff $ax+by+(-ap-bq)=0=ax+by+c$. That is, $-ap-bq=c$. Thus, the two lines are either same or disjoint. □

Definition A.8 (Circle)**.** Any set of the form $\{(x,y) : (x-a)^2+(y-b)^2=r^2\}$ is called a circle.

Theorem A.9. *Given any point (a,b) and any radius r, there exists a circle with centre (a,b) and radius r.*

Proof. This is obvious as the required circle is $\{(x,y) : (x-a)^2+(y-b)^2=r^2\}$ □

A.1.1. Linear maps.

Definition A.10. A function $f : \mathbb{R}^2 \to \mathbb{R}^2$ is called a linear map if $f(v+w)=f(v)+f(w)$ and $f(\alpha v)=\alpha f(v)$.

Lemma A.11. *If f is linear, then $f(0,0)=(0,0)$.*

Proof. $f(0,0)=f((0,0))=f(0(x,y))=0f(x,y)=(0,0)$. □

Let us now fix some notation, say $e_1=(1,0)$ and $e_2=(0,1)$. Notice that given any $(x,y)\in\mathbb{R}^2$, we can write $(x,y)=xe_1+ye_2$. An expression of the form xe_1+ye_2 is called a linear combination of e_1 and e_2. Thus, we can describe our observation as

Lemma A.12. *Every element $v\in\mathbb{R}^2$ can be expressed as a linear combination of e_1 and e_2.*

Lemma A.13. *If f is linear, then there exists $a,b,c,d\in\mathbb{R}$ such that $f(x,y)=(ax+by,cx+dy)$.*

Proof. Let $(a,c)=f(1,0)$ and $(b,d)=f(0,1)$. Given any point $(x,y)\in\mathbb{R}^2$,

$$\begin{aligned} f(x,y) &= f(x(1,0)+y(0,1)) = f(x(1,0))+f(y(0,1)) \\ &= xf(1,0)+yf(0,1) = x(a,c)+y(b,d) \\ &= (ax,cx)+(by,dy) = (ax+by,cx+dy) \end{aligned}$$

□

Conversely,

Lemma A.14. *The function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $f(x,y)=(ax+by,cx+dy)$ is linear for any $a,b,c,d\in\mathbb{R}^2$.*

Thus, there is a one-one corresponding between linear functions and arrays/matrices of the form $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Suppose and f_i corresponds to $\begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix}$. Then

$$\begin{aligned} f_2\circ f_1(x,y) &= f_2(a_1x+b_1y,c_1x+d_1y) \\ &= (a_2a_1x+a_2b_1y+b_2c_1x+b_2d_1y, c_2a_1x+c_2b_1y+d_2c_1x+d_2d_1y) \\ &= ((a_2a_1+b_2c_1)x+(a_2b_1+b_2d_1)y,(c_2a_1+d_2c_1)x+(c_2b_1+d_2d_1)y). \end{aligned}$$

Thus, $f_2\circ f_1$ corresponds to the matrix $\begin{bmatrix} a_2a_1+b_2c_1 & a_2b_1+b_2d_1 \\ c_2a_1+d_2c_1 & c_2b_1+d_2d_1 \end{bmatrix}$. This allows us to define an operation on matrices, namely, $\begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} = \begin{bmatrix} a_2a_1+b_2c_1 & a_2b_1+b_2d_1 \\ c_2a_1+d_2c_1 & c_2b_1+d_2d_1 \end{bmatrix}$. Notice that this matches with the familiar matrix multiplication you might have encountered in school. Thankfully, the mystery of why matrix multiplication was defined in this strange manner is now resolved!

Definition A.15 (Line-preserving functions)**.** We say a function $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a line preserving function if for every a, b, c, d with $(a, b) \neq (0, 0)$, $f(\Lambda_{a,b,c,d}) = \Lambda_{a',b',c',d'}$ for some $a', b'c', d$ such that $(a', b') \neq (0, 0)$.

Lemma A.16. *Let f be a linear map. Then given $a, b, c, d \in \mathbb{R}$, there exists $a', b', c', d' \in \mathbb{R}$ such that $f(\Lambda_{a,b,c,d}) = \Lambda_{a',b',c',d'}$.*

Proof. Let $f(a, b) = (a', b')$ and let $f(c, d) = (c', d')$. Then

$$\begin{aligned} f(\Lambda_{a,b,c,d}) &= f\left(\{\alpha(a, b) + (c, d) : \alpha \in \mathbb{R}\}\right) \\ &= \{f\left(\alpha(a, b) + (c, d)\right) : \alpha \in \mathbb{R}\} \\ &= \{f\left(\alpha(a, b)\right) + f\left((c, d)\right) : \alpha \in \mathbb{R}\} \\ &= \{\alpha f\left((a, b)\right) + f\left((c, d)\right) : \alpha \in \mathbb{R}\} \\ &= \{\alpha(a', b') + (c', d') : \alpha \in \mathbb{R}\} \\ &= \Lambda_{a',b',c',d'} \end{aligned}$$

□

However, $\Lambda_{a',b',c',d'}$ is a line iff $(a', b') \neq 0$. Thus,

Lemma A.17. *A linear map f is line-preserving $\iff$ $(a, b) \neq (0, 0)$ implies $f(a, b) \neq (0, 0)$ $\iff$ $f(a, b) = (0, 0)$ implies $(a, b) = (0, 0)$.*

Lemma A.18. *A linear map f is injective $\iff$ $f(a, b) = (0, 0)$ implies $(a, b) = (0, 0)$.*

Proof. Assume f is injective and $f(a, b) = (0, 0)$. Notice that earlier we had shown that $f(0, 0) = (0, 0)$ as well. Thus, by injectivity of f, $(a, b) = (0, 0)$.

Now assume $f(a, b) = (0, 0)$ implies $(a, b) = (0, 0)$. Let (x_1, y_1) and (x_2, y_2) be any two points such that $f(x_1, y_1) = f(x_2, y_2)$. Thus, $f(x_2 - x_1, y_2 - y_1) = f(x_2, y_2) - f(x_1, y_1) = (0, 0)$ Thus, by our assumption, $(x_2 - x_1, y_2 - y_1) = (0, 0)$, that is, $(x_1, y_1) = (x_2, y_2)$. As, (x_i, y_i) were arbitrary, f is injective. □

Lemma A.19. *A linear map $f : \mathbb{R}^2 \to \mathbb{R}^2$ is injective iff it is bijective.*

We have proved that every bijective linear map is a bijective line-preserving map. Is the converse true? Unfortunately, the following exercise shows that the converse is not true.

Exercise A.20. Show that $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $f(x, y) = (x + a, y + b)$ is line-preserving but is not linear.

Thus, the obstruction is the existence of translations. Or more precisely, the obstruction is the existence of functions such that $f(0, 0) \neq (0, 0)$.

A.2. A line-preserving bijection that preserves origin is a linear map

The proof is a mixture of ideas from [Hughes] and Chapter 6 of [Lyndon]. From now on, when there is no ambiguity, we will denote $(0, 0)$ as just 0.

Lemma A.21. *Given two vectors $v, w \in \mathbb{R}^2$, the line containing v, w is given by $\{w + \alpha(v - w) : \alpha \in \mathbb{R}\}$.*

Proof. Let $v = (v_1, v_2)$ and $w = (w_1, w_2)$. Then,

$$\{w + \alpha(v - w) : \alpha \in \mathbb{R}\} = \{(w_1, w_2) + \alpha(v_1 - w_1, v_2 - w_2) : \alpha \in \mathbb{R}\} = \Lambda_{v_1 - w_1, v_2 - w_2, w_1, w_2}.$$

Thus, the set describes a line. Moreover, when $\alpha = 0$ we have $w + \alpha(v - w) = w$ and when $\alpha = 1$ we have $w + \alpha(v - w) = v$. Thus, the line contains v and w. □

Now, $w + \alpha(v - w) = \alpha v + (1 - \alpha)w$. Thus, $\{w + \alpha(v - w) : \alpha \in \mathbb{R}\} = \{av + bw : a + b = 1\}$. In particular, we have,

Lemma A.22. *Let $a, b \in \mathbb{R}$ are such that $a + b = 1$. Given any two $v, w \in \mathbb{R}^2$, $av + bw$ lies on the line passing through v and w.*

Lemma A.23. *Let f be line-preserving bijection such that $f(0) = 0$, then $f(e_2) \neq c.f(e_1)$ for any $c \in \mathbb{R}$.*

Proof. Assume the contrary. Let $f(e_i) = v_i$, our assumptions imply $v_2 = cv_1$. Let $v = (a, b)$ be an arbitrary point in $\mathbb{R}^2$. Then, we can write v as $ae_1 + be_2$. Assume $a + b \neq 0$. Then,

$$\begin{aligned} v = ae_1 + be_2 &= (a + b)\left(\frac{a}{a+b}e_1 + \frac{b}{a+b}e_2\right) \\ &= (1 - (a + b))0 + (a + b)\left(\frac{a}{a+b}e_1 + \frac{b}{a+b}e_2\right) \end{aligned}$$

Let us use the notation $w = \frac{a}{a+b}e_1 + \frac{b}{a+b}e_2$. Then w lies on the line containing e_1 and e_2. Therefore, if $f(w)$ lies on the line joining v_1 and $v_2 = cv_1$. Therefore, $w = dv_1$ for some $d \in \mathbb{R}$. Now, v lies on the line joining 0 and w. Thus, $f(v)$ lies on the line joining $f(0) = 0$ and $f(w) = dv_1$. Thus, $f(v) = \lambda v_1$ for some $\lambda \in \mathbb{R}$.

Thus, the image of $\mathbb{R}^2 \setminus \{a(1, -1) : a \in \mathbb{R}\}$ is a line. Moreover, the set $\{a(1, -1) : a \in \mathbb{R}\}$ is also a line. Thus, the image of $\mathbb{R}^2$ can atmost be the union of two distinct lines and therefore cannot be $\mathbb{R}^2$. Thus, f is not surjective, which is a contradiction. Therefore, our assumption that $v_2 = cv_1$ has to be wrong. □

Exercise A.24. Two points (x_1, y_1) and (x_2, y_2) lie on the same line passing through origin iff $x_1y_2 - x_2y_1 = 0$.

As $v_2 \neq cv_1$ for any $c \in \mathbb{R}$ they don't lie on the same line passing through origin. Thus, if $v_i = (x_i, y_i)$, then $x_1y_2 - x_2y_1 \neq 0$. Thus, define T from $\mathbb{R}^2 \to \mathbb{R}^2$ to be the function that corresponds to the matrix $\begin{bmatrix} \frac{y_2}{x_1y_2 - x_2y_1} & \frac{-x_2}{x_1y_2 - x_2y_1} \\ \frac{-y_1}{x_1y_2 - x_2y_1} & \frac{x_1}{x_1y_2 - x_2y_1} \end{bmatrix}$. Then $T(v_i) = e_i$. That is, $T \circ f(e_i) = e_i$ and $T \circ f(0) = 0$. Let us denote $T \circ f$ by h. As T and f maps lines to lines, h also maps lines to lines. Moreover,

Lemma A.25. *If l and m are distinct parallel lines, then $h(l)$ and $h(m)$ are also parallel.*

Proof. If $h(l)$ and $h(m)$ are not parallel, they intersect at some point p. As $p \in h(l)$, there exists some point v in l such that $h(v) = p$. As $p \in h(m)$, there exists some point $w \in m$ such that $h(w) = p$. But, h is a bijection, so $v = w$. But, this means l and m intersect. A contradition to the assumption that they are distinct parallel lines. □

On the other hand, if l and m were the same line, then clearly, $h(l)$ and $h(m)$ are the same and thus parallel. Therefore,

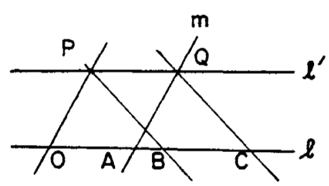
Lemma A.26. *If l and m are parallel, then $h(l)$ and $h(m)$ are parallel.*

Further,

1. The line passing through 0 and e_1 (that is the x-axis), is mapped to a line by h. But, this line contains 0 and e_1 so it has to be the x-axis.
2. The line passing through 0 and e_2 (that is the y-axis), is mapped to a line by h. But, this line contains 0 and e_2 so it has to be the y-axis.
3. Thus, by previous lemma, h sends horizontal lines to horizontal lines and vertical lines to vertical lines.
4. But $h(1,0) = (1,0)$, so $h(\{(x,y) : x = 1\}) = \{(x,y) : x = 1\}$. Similarly, $h(\{(x,y) : y = 1\}) = \{(x,y) : y = 1\}$.

Our aim is to now prove that $h(x,0) = (x,0)$ and $h(0,y) = (0,y)$ for all $x, y \in \mathbb{R}$. We prove this by showing that if f a line preserving bijection such that $f(0) = 0$ fixes a line and two points on it, then it fixes each point on the line.

Given a line l and a point O on l. We can define a construction that takes two points A and B on l and gives a point C on l such that $|OA| + |OB| = |OC|$. Take an arbitrary point P not on l. Draw lines OP, BP and the line l' parallel to l passing through P. Then draw a line passing through A and parallel to OP, let this hit l' at Q. Now, draw a line parallel to PB passing through PB. Then, $|OA| = |PQ| = |BC|$. Thus, $|OA| + |OB| = |BC| + |OB| = |OC|$.

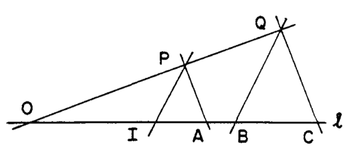


Note that this construction depends on O but does not depend on P. We denote $C = A \oplus B$.

Lemma A.27. *Let g be a line preserving bijection such that $g(0) = 0$. If $g(l) = l$ and $g(O) = O$, then $g(A \oplus B) = g(A) \oplus g(B)$.*

Proof. As distinct parallel lines are sent to distinct parallel lines by g, the figure used to construct $C = A \oplus B$ gets transformed to the figure used to construct $g(C) = g(A) \times g(B)$. □

Given a line l and two points O and I on L we now construct an operation that takes two points A and B and returns a point C such that $|OC||OI| = |OA||OB|$. Choose an arbitrary point P not on l. Then draw lines OP, IP, and AP. Now draw a line parallel to IP passing through B, let this line intersect OP at Q. And draw a line parallel to AP passing through Q, let this line intersect l at C. Note that $\triangle OAP$ is similar to $\triangle OCQ$. Thus, $\frac{|OC|}{|OA|} = \frac{|OQ|}{|OP|}$. And $\triangle OIP$ is similar to $\triangle OBQ$. Thus, $\frac{|OQ|}{|OP|} = \frac{OB}{OI}$. Thus, $\frac{|OC|}{|OA|} = \frac{|OB|}{|OI|}$[1].



This construction depends on both O and I but does not depend on P. We denote $C = A \otimes B$. Then, as before, we have

[1] Note to self: The two images in this section are copied from the book.

Lemma A.28. *Let g be a line preserving bijection such that $g(0) = 0$. If $g(l) = l$, $g(O) = O$ and $g(I) = I$, then $g(A \otimes B) = g(A) \otimes g(B)$.*

Let l be the x-axis, $O = (0, 0)$ and $I = (1, 0)$. Then, as $|OI| = 1$, $(x, 0) \oplus (y, 0) = (x + y, 0)$ and $(x, 0) \otimes (y, 0) = (xy, 0)$. Further, note that $h|_{\{(x,0):x\in\mathbb{R}\}}$ is also a bijection. Thus, h induces a function $g : \mathbb{R} \to \mathbb{R}$, given by the relation $h(x, 0) = (g(x), 0)$.

$$\begin{aligned}(g(x + y), 0) &= h(x + y, 0) = h((x, 0) \oplus (y, 0)) = h(x, 0) \oplus h(y, 0)\\ &= (g(x), 0) \oplus (g(y), 0)) = (g(x) + g(y), 0)\end{aligned}$$

Thus, $g(x + y) = g(x) + g(y)$. Similarly,

$$\begin{aligned}(g(xy), 0) &= h(xy, 0) = h((x, 0) \otimes (y, 0)) = h(x, 0) \otimes h(y, 0)\\ &= (g(x), 0) \otimes (g(y), 0)) = (g(x)g(y), 0)\end{aligned}$$

Thus, $g(xy) = g(x)g(y)$. We will now prove that the only function that satisfies these properties is the identity function.

Lemma A.29. *If $g : \mathbb{R} \to \mathbb{R}$ is a bijection such that $g(x + y) = g(x) + g(y)$ and $g(xy) = g(x)g(y)$, then $g|_{\mathbb{Q}}$ is the identity.*

Proof. First observe that $g(0) = g(0 + 0) = g(0) + g(0)$. Thus, $g(0) = 0$. And $g(1) = g(1.1) = g(1)g(1)$. As $1 \neq 0$ and g is bijection, $g(1) \neq g(0) = 0$. Thus, $g(1) = 1$. Now, $g(2) = g(1 + 1) = 1+1 = 2$. Using induction, you can now prove that $g(n) = n$. Also, $0 = g(0) = g(n+(-n)) = g(n)+g(-n) = n + g(-n)$. Thus, $g(-n) = -n$. Finally, $m = g(m) = g\left(\frac{m}{n}.n\right) = g\left(\frac{m}{n}\right) g(n) = g\left(\frac{m}{n}\right) n$. Thus, $g\left(\frac{m}{n}\right) = \frac{m}{n}$. □

Lemma A.30. *If $g : \mathbb{R} \to \mathbb{R}$ is a bijection such that $g(x + y) = g(x) + g(y)$ and $g(xy) = g(x)g(y)$, then $x \leq y \iff g(x) \leq g(y)$.*

Proof.

$$\begin{aligned}0 \leq x &\iff x = y^2 \text{ for some } y \in \mathbb{R}\\ &\iff g(x) = g(y^2) = g(y)g(y) = g(y)^2\\ &\iff 0 \leq g(x)\end{aligned}$$

Now,

$$x \leq y \iff 0 \leq y - x \iff 0 \leq g(y - x) = g(y) - g(x) \iff g(x) \leq g(y)$$

□

Theorem A.31. *If $g : \mathbb{R} \to \mathbb{R}$ is a bijection such that $g(x+y) = g(x)+g(y)$ and $g(xy) = g(x)g(y)$, then g is the identity.*

Proof. Define $L_x = \{y \in \mathbb{Q} : y \leq x\}$. Then,

$$\begin{aligned}L_{g(x)} &= \{y \in \mathbb{Q} : y \leq g(x)\}\\ &= \{y \in \mathbb{Q} : g(y) \leq g(x)\} \text{ (note } g \text{ is identity on } \mathbb{Q})\\ &= \{y \in \mathbb{Q} : y \leq x\} \text{ (by Lemma A.30)}\\ &= L_x\end{aligned}$$

But given any two distint real numbers x and y, there exists a rational number z such that $x < y < z$. Therefore $L_x = Ly$ iff $x = y$. Thus, $g(x) = x$. As x was arbitrary, we have proved that g is the identity. □

Thus, $h(x,0) = (g(x),0) = (x,0)$ for all $x \in \mathbb{R}$. A very similar argument will show that $h(0,y) = (0,y)$ for all $y \in \mathbb{R}$. Now, we already saw that $h(\{(x,y) : y = b\})$ is a horizontal line. But, as $h(0,b) = (0,b)$, the image contains the point $(0,b)$. Thus, $h(\{(x,y) : y = b\}) = \{(x,y) : y = b\}$. Similarly, $h(\{(x,y) : x = a\}) = \{(x,y) : x = a\}$. Then,$h(a,b) \in h(\{(x,y) : x = a\}) = \{(x,y) : x = a\}$ and $h(a,b) \in h(\{(x,y) : y = b\}) = \{(x,y) : y = b\}$. Thus, $h(a,b) \in \{(x,y) : x = a\} \cap \{(x,y) : y = b\} = \{(a,b)\}$. Thus, $h(a,b) = (a,b)$. Therefore, by the uniqueness of inverse (of T), f is the function corresponding to $\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}$. And thus, f is linear. Hence, we have finally proved,

Theorem A.32. *Let f be a line-preserving bijection such that $f(0,0) = (0,0)$, then f is linear.*

Let us summarise what we have seen so far. First, we showed that given a linear map f, there exists real numbers a, b, c, d such that $f(x,y) = (ax + by, cx + dy)$. We associated a matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ to this function. The collection of all 2×2 matrices is often denoted as $M(2)$. Later, to ensure lines go to lines, we decided to focus only on bijections. These functions therefore correspond to invertible matrices.

Definition A.33. The collection of 2×2 invertible matrices is denoted by $GL(2)$. It is often called the general linear group.

The structure of linear maps allowed us to prove that they map lines to lines. We then observed that the converse is not true as linear maps fix the origin, but line-preserving maps do not necessarily fix the origin. With immense effort, we managed to prove that all line-preserving bijections that that preserve the origin are linear. This discovery allows us to understand all line-preserving bijections.

Theorem A.34. *If f is a line-preserving bijection, then there exists real numbers a, b, c, d, e, f such that $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible and $f(x,y) = (ax + by + e, cx + dy + f)$.*

Proof. Given a line-preserving bijection, consider the function $g : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $g(x,y) = f(x,y) - f(0,0)$. Then, $g(0,0) = (0,0)$. Thus, g is a linear bijection and hence $g(x,y) = (ax + by, cx + dy)$ for some a, b, c, d such that $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible. Define $(e,f) := f(0,0)$. Then, $f(x,y) = g(x,y) + f(0,0) = ax + by, cx + dy) + (e,f) = (ax + by + e, cx + dy + f)$. ☐

A.3. Rotations, Reflections and Translations

We already encountered translations as examples of functions that are line-preserving but not linear. Translation is an important example of transformation of the plane and the other two classes of functions are reflections and rotations.

Definition A.35 (Translation)**.** A function $T : \mathbb{R}^2 \to \mathbb{R}^2$ of the form $T(x,y) = (x + a, y + b)$ is called a translation.

Definition A.36. Two lines $L_{p,q,r}$ and $L_{p',q',r'}$ are perpendicular iff $pp' = -qq'$.

Definition A.37 (Reflection)**.** A function $R : \mathbb{R}^2 \to \mathbb{R}^2$ is a reflection about a line l if given any point (x,y),

(1) the line l is perpendicular to the line passing through $R(x,y)$ and (x,y).
(2) The mid-point of $R(x,y)$ and (x,y), that is $\frac{1}{2}(R(x,y) + (x,y))$, lies on l.

Remark A.38. The two conditions in the above definition together mean that l is the perpendicular bisector of the line segment joining $R(x, y)$ and (x, y).

Theorem A.39. *The reflection $R_{p,0,r}$ about the line $L_{p,0,r}$ is given by the map*

$$R_{p,0,r}(x, y) = \left(-x - \frac{2r}{p}, y\right).$$

Proof. Note that, then $p \neq 0$ and the line is parallel to the y-axis. Let (a, b) be an arbitrary point and let $R_{p,0,r}(a, b) = (c, d)$. Then, from Theorem A.6, the line passing through (a, b) and (c, d) is $L_{(d-b),(a-c),(bc-ad)}$. As $L_{p,0,r}$ is perpendicular to $L_{(d-b),(a-c),(bc-ad)}$,

$$p(d-b) = 0(a-c) \iff d = b. \tag{A.1}$$

Moreover, the point $\left(\frac{a+c}{2}, \frac{b+d}{2}\right)$ lies on $L_{p,0,r}$. That is,

$$p(a+c) + 0(b+d) + 2r = 0 \iff c = -a - \frac{2r}{p}. \tag{A.2}$$

That is, $R_{p,0,r}(a, b) = (a, -b - \frac{2r}{p})$. As, (a, b) was arbitrary, we have the result. □

Theorem A.40. *The reflection about a line parallel to y-axis can be decomposed into a translation that takes the line to the y-axis, reflection about the y-axis, and the translation back to the original line. More precisely,*

$$R_{p,0,r} = T_{-\frac{r}{p},0} \circ R_{p,0,0} \circ T_{\frac{r}{p},0}$$

Proof.

$$\begin{aligned} T_{-\frac{r}{p},0} \circ R_{p,0,0} \circ T_{\frac{r}{p},0}(x, y) &= T_{-\frac{r}{p},0} \circ R_{p,0,0}\left(x + \frac{r}{p}, y\right) \\ &= T_{-\frac{r}{p},0}\left(-x - \frac{r}{p}, y\right) \\ &= \left(-x - \frac{2r}{p}, y\right) \\ &= R_{p,0,r}(x, y). \end{aligned}$$

□

Theorem A.41. *Given a line $L_{p,q,r}$ where $q \neq 0$, define $m = \frac{-p}{q}$. Then, the reflection $R_{p,q,r}$ about the line $L_{p,q,r}$ is given by the map*

$$R_{p,q,r}(x, y) = \begin{bmatrix} \frac{1-m^2}{1+m^2} & \frac{2m}{1+m^2} \\ \frac{2m}{1+m^2} & \frac{m^2-1}{1+m^2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Proof. Let (a, b) be an arbitrary point and let $R(a, b) = (c, d)$. Then, from Theorem A.6, the line passing through (a, b) and (c, d) is $L_{(d-b),(a-c),(bc-ad)}$. As $L_{p,q,r}$ is perpendicular to the line $L_{(d-b),(a-c),(bc-ad)}$,

$$p(d-b) = -q(a-c). \tag{A.3}$$

Moreover, the point $\left(\frac{a+c}{2}, \frac{b+d}{2}\right)$ lies on $L_{p,q,0}$. That is,

$$p(a+c) + q(b+d) + r = 0. \tag{A.4}$$

From Equation A.3, we get

$$c = a - md + mb \tag{A.5}$$

and from Equation A.4, we get

$$d = ma + mc - b - \frac{r}{q} \tag{A.6}$$

Substituting this value of d back in Equation A.5, we get

$$c = a - m^2 a - m^2 c + mb + \frac{mr}{q} + mb.$$

That is,

$$c = \left(\frac{1-m^2}{1+m^2}\right) a + \left(\frac{2m}{1+m^2}\right) b + \frac{mr}{(1+m^2)q}. \tag{A.7}$$

Substituting this value of c back in Equation A.5, we get

$$d = \left(\frac{2m}{1+m^2}\right) a + \left(\frac{m^2-1}{1+m^2}\right) b - \frac{r}{(1+m^2)q}. \tag{A.8}$$

As $(a,b) \in \mathbb{R}^2$ was arbitrary, we can say

$$R_{p,q,r}(x,y) = \begin{bmatrix} \frac{1-m^2}{1+m^2} & \frac{2m}{1+m^2} \\ \frac{2m}{1+m^2} & \frac{m^2-1}{1+m^2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \frac{mr}{(1+m^2)q} \\ \frac{-r}{(1+m^2)q} \end{bmatrix}$$

□

In particular, when $r = 0$, we have

Corollary A.42. *Given a line $L_{p,q,0}$ where $p \neq 0$, define $m = \frac{-q}{p}$. Then, the reflection $R_{p,q,0}$ about the line $L_{p,q,0}$ is given by the map*

$$R_{p,q,0}(x,y) = \begin{bmatrix} \frac{1-m^2}{1+m^2} & \frac{2m}{1+m^2} \\ \frac{2m}{1+m^2} & \frac{m^2-1}{1+m^2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Theorem A.43. *The reflection about an arbitrary line (not parallel to the y-axis) can be written as a composition of a Translation that takes it to a line passing through the origin, reflection about this line, and translation that takes the line back to original line. More precisely,*

$$R_{p,q,r} = T_{0,-\frac{r}{q}} \circ Rp, q, 0 \circ T_{0,\frac{r}{q}}$$

Proof.

$$\begin{aligned} T_{0,-\frac{r}{q}} \circ R_{p,q,0} \circ T_{0,\frac{r}{q}}(x,y) &= T_{0,-\frac{r}{q}} \circ R_{p,q,0}\left(x, y + \frac{r}{q}\right) \\ &= T_{0,-\frac{r}{q}}\left(-x - \frac{r}{p}, y\right) \\ &= \left(-x - \frac{2r}{p}, y\right) \\ &= R_{p,0,r}(x,y). \end{aligned}$$

□

Theorem A.44. *Let l be a line passing through the origin making an angle $\theta \neq \frac{\pi}{2}$ with the x-axis. Then, $l = L_{-\sin(\theta),\cos(\theta),0}$.*

Proof. Notice that $(\cos(\theta), \sin(\theta))$ and $(0,0)$ satisfy the equation $-\sin(\theta)x + \cos(\theta)y + 0 = 0$. □

Theorem A.45. *If $\theta \neq 0$, then reflection about $L_{-\sin(\theta),\cos(\theta),0}$ is given by*

$$\begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}.$$

Proof. Here $m = \frac{\sin(\theta)}{\cos(\theta)}$. Thus,

$$\frac{1-m^2}{1+m^2} = \frac{\cos^2(\theta) - \sin^2(\theta)}{\sin^2(\theta) + \cos^2(\theta)} = cos(2\theta).$$

and

$$\frac{2m}{1+m^2} = 2\sin(\theta)\cos(\theta) = \sin(2\theta).$$

□

Theorem A.46. *The rotation R_α about the origin by an angle α is given by the map*

$$R_\alpha(x, y) = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Proof. The easiest way to prove this result is by using the polar coordinates. Recall that a point $(x, y) = r(\cos(\theta), \sin(\theta)) = (r\cos(\theta), r\sin(\theta))$. When rotated by an angle α, you get the point

$$\begin{aligned} (r\cos(\theta+\alpha), r\sin(\theta+\alpha)) &= (r\cos(\theta)\cos(\alpha) - r\sin(\theta)\sin(\alpha), r\sin(\theta)\cos(\alpha) + r\cos(\theta)\sin(\alpha)) \\ &= (x\cos(\alpha) - y\sin(\alpha), y\cos(\alpha) + x\sin(\alpha)) \\ &= (x\cos(\alpha) - y\sin(\alpha), x\sin(\alpha)) + y\cos(\alpha) \\ &= \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \end{aligned}$$

□

A.4. Klein's approach to geometry

In school, the notion of congruence is not usually defined precisely. It is first expressed intuitively as "two triangles are congruent if you can pick one and place it on the other in such a way that they match perfectly". Later, we talk about conditions/criteria for checking congruence, but the equivalence of the definition and these conditions are not discussed. The equivalence of the various conditions are generally discussed on the other hand.

To make the intuitive notion precise, we need to define a collection of functions that do the job of "picking a triangle and placing it on another". For example, we are allowed to translate, rotate, reflect etc. And each of these "moves" correspond to some functions from $f : \mathbb{R}^2 \to \mathbb{R}^2$. The similarity of triangles can also be defined in a similar manner. Thus, Felix Klein observed that Euclidean geometry is the study of properties that are preserved a collection (let us call it E for Euclidean) of functions $f : \mathbb{R}^2 \to \mathbb{R}^2$. He further observed that this collection satisfies some nice properties[2]:

(1) The collection contains the identity function.
(2) If f belongs to the collection, then f is invertible and f^{-1} also belongs to the collection.
(3) If f and g belong to the collection, then so does $f \circ g$.

[2]We first encountered this presentation in [Needham]

He further realised that these properties were no accident. This collection of functions defines a notion of "sameness" on triangles. And any sensible notion of "sameness" should be an equivalence relation. And these properties help establish the reflexivity, symmetry and transitivity of the relationship.

Our goal is to understand Euclidean geometry from this Kleinean perspective. We may call functions that can be expressed as composition of rotations, reflections, and translations as rigid maps. We may call functions that preserve distances as isometries. Interestingly every rigid map is an isometry and every isometry is a rigid map. This fact can be proved using high school geometry or can be proved using concepts from linear algebra - needless to say that linear algebra makes our life a lot easier. This can be a motivation to study linear algebra.

Definition A.47 (Isometry)**.** We say a function $f : \mathbb{R}^2 \to \mathbb{R}^2$ is an isometry if for all $v, w \in \mathbb{R}^2$ $d(v, y) = d(f(v), f(w))$.

Lemma A.48. *If f is an isometry and $f(0) = 0$, then $\|f(v)\| = \|v\|$ for all $v \in \mathbb{R}^2$.*

Proof. $\|v\| = d(v, 0) = d(f(v), f(0)) = d(f(v), 0) = \|f(v)\|$. □

Exercise A.49. Show that if $L : \mathbb{R}^2 \to \mathbb{R}^2$ is an isometry, then the map $T_{(L(0,0))} \circ L$ is an isometry that preserves origin, that is $T_{L(0,0)} \circ L(0, 0) = (0, 0)$.

Exercise A.50. Show that if $f : \mathbb{R}^2 \to \mathbb{R}^2$ is an isometry that preserves origin that is $f(0, 0) = (0, 0)$, then f preserves dot product, that is, $f(v) \cdot f(w) = v \cdot w$.
Hint: Use $(f(v) - f(w)) \cdot (f(v) - f(w)) = \|f(v) - f(w)\| = \|v - w\| = (v - w) \cdot (v - w)$.

Theorem A.51. *If $f : \mathbb{R}^2 \to \mathbb{R}^2$ is an isometry that preserves origin (that is $f(0, 0) = (0, 0)$), then f is linear.*

Proof. To show that f is linear, we need to prove that $f(v+w) = f(v)+f(w)$ and $f(\alpha v) = \alpha f(v)$. As the only vector whose norm is 0 is the 0-vector, it is enough to show that $\|f(v + w) - f(v) - f(w)\| = 0$ and $\|f(\alpha v) - \alpha f(v)\| = 0$. Both these follow from a simple computation using the observation that f preserves inner product. Thus, these computations are left as exercise. □

Exercise A.52. Show that if $f : \mathbb{R}^2 \to \mathbb{R}^2$ is an isometry that preserves origin, then $\|f(v + w) - f(v) - f(w)\| = 0$ and $\|f(\alpha v) - \alpha f(v)\| = 0$.

Lemma A.53. *If f is an isometry and $f(0) = 0$ then f is a bijection.*

Proof. Consider two points $v, w \in \mathbb{R}^2$ such that $f(v) = f(w)$. Then, $d(f(v), f(w)) = 0$. But, $d(v, w) = d(f(v), f(w))$. Therefore $d(v, w) = 0$ and thus, $v = w$. Thus, f is injective.

Let $v \in \mathbb{R}^2$ be arbitrary. We will prove f is surjective by finding a preimage for v. If $f(v) = v$, then we are done. So, assume $f(v) \neq v$, let $f(v) = u$. Once again, if $f(u) = v$, we are done. So, assume $f(u) \neq v$. Let $\|v\| = r$. Then, $\|u\| = \|v\| = r$. Now, there are two possibilities, $d(u, v) < 2r$ or $d(u, v) = 2r$.

Case 1 - $d(u, v) = 2r$: Notice that v is the only element with norm r such that $d(u, v) = 2r$. But, $d(u, v) = d(f(u), f(v)) = d(f(u), u) = d(u, f(u))$. Thus, $f(u) = v$. Hence we have found a preimage for v.

Case 2 - $d(u, v) < 2r$: Now there is exists exactly one $w \neq u$ such that $\|w\| = r$ and $d(w, v) = d(u, v)$. Thus, $d(u, f(w)) = d(f(w), u) = d(f(w), f(v)) = d(f(u), f(v)) = d(f(u), u) = d(f(u), f(v)) = d(u, v)$ and $\|f(w)\| = r$. That is, $d(u, v) = d(u, f(u)) = d(u, f(w))$. Thus, either $f(u) = f(w)$ or $f(w) = v$. But, as f is injective and $w \neq u$, $f(w) \neq f(u)$. Therefore, $f(w) = v$. Hence we have found a preimage for v. □

Theorem A.54. *If f is an isometry, then f is a bijection.*

Proof. Consider $g : \mathbb{R}^2 \to \mathbb{R}^2$ such that $g(v) = f(v) - f(0)$. Then, $g \in G_3$ and $g(0) = 0$, so g is a bijection. $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $T(v) = v + f(0)$ is also a bijection. Thus, $T \circ g$ is a bijection. But, $T \circ g(v) = T(f(v) - f(0)) = f(v) - f(0) + f(0) = f(v)$. That is, $T \circ g = f$. Thus, f is a bijection. □

Theorem A.55. *Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map f and let A be the matrix associated to f. If f is an isometry then $AA^T = I = A^T A$.*

Proof. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ or in other words $f(x, y) = (ax + by, cx + dy)$. Then $f(1, 0) = (a, c)$ and $f(0, 1) = (b, d)$. But, as f preserves norm

$$1 = \|(1, 0)\| = \|f(1, 0)\| = \|(a, c)\| = a^2 + c^2.$$

Similarly,

$$1 = \|(0, 1)\| = \|f(0, 1)\| = \|(b, d)\| = b^2 + d^2.$$

Moreover,

$$0 = \langle (1, 0), (0, 1)\rangle = \langle f(1, 0), f(0, 1)\rangle = \langle (a, c), (b, d)\rangle = ab + cd.$$

Thus,

$$A^T A = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

But, notice that

But, if A is associated with an isometry, it is invertible and its inverse $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ is also associated with an isometry. Thus, if $B = A^{-1}$, then

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = B^T B = \frac{1}{(ad - bc)^2} \begin{bmatrix} d & -c \\ -b & a \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} d^2 + c^2 & -(bd + ac) \\ -(bd + ac) & a^2 + b^2 \end{bmatrix}.$$

Thus, $ac + bd = 0$ and $d^2 + c^2 = (ad - bc)^2 = a^2 + b^2$. Further notice that

$$2 = 1 + 1 = (a^2 + c^2) + (b^2 + d^2) = (a^2 + b^2) + (c^2 + d^2) = 2(ad - bc)^2$$

Therefore, $(ad - bc)^2 = 1$. That is, $d^2 + c^2 = 1 = a^2 + b^2$. Hence,

$$AA^T = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} a^2 + b^2 & ac + bd \\ ac + bc & c^2 + d^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

□

Theorem A.56. *Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map f and let A be the matrix associated to f. If $AA^T = I = A^T A$, then f is an isometry.*

Proof. Notice that $\langle v, w\rangle = v^T w = \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$. Therefore,

$$\|Av\|^2 = \langle Av, Av\rangle = (Av)^T (Aw) = \left(v^T A^T\right)(Av) = v^T \left(A^T A\right) v = v^T v = \langle v, v\rangle = \|v\|^2.$$

Thus, A preserves norm and hence distance. □

Definition A.57. A matrix A is called an orthogonal matrix if $AA^T = I$. The collection of all orthogonal matrices is denoted as $O(2)$. Notice that $1 = \det(I) = \det(AA^T) = \det(A)\det(A^T) = \det(A)^2$. Thus, $\det(A) = \pm 1$. The collection of all elements of $O(2)$ with determinant 1 is denoted as $SO(2)$.

The set of rigid motions (or isometries) under function composition also form great examples of groups. Given any set X, the collection of bijections $f : X \to X$ under composition forms a group. These were one of the original motivations to study group theory and Cayley's theorem says that in fact every group can be thought of as a collection of bijections from a set to itself. Having played with concrete examples will help students when they start learning group theory.

Exercise A.58. Show that $O(2)$, the set of all orthogonal matrices, is a group under matrix multiplication.

Exercise A.59. Show that $A \in SO(2)$ iff there exits some $\theta \in [0, 2\pi]$ such that

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

Let $R = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. If $A \in O(2)$, then either $A \in SO(2)$ or $RA \in SO(2)$. If $RA \in SO(2)$, then there exists $\theta \in [0, 2\pi]$ such that

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

That is,

$$\begin{aligned} A &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(\theta) & -sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} -\cos(\theta) & \sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \\ &= \begin{bmatrix} \cos(\pi-\theta) & \sin(\pi-\theta) \\ \sin(\pi-\theta) & -\cos(\pi-\theta) \end{bmatrix} = \begin{bmatrix} \cos\left(2\left(\frac{\pi}{2}-\frac{\theta}{2}\right)\right) & \sin\left(2\left(\frac{\pi}{2}-\frac{\theta}{2}\right)\right) \\ \sin\left(2\left(\frac{\pi}{2}-\frac{\theta}{2}\right)\right) & -\cos\left(2\left(\frac{\pi}{2}-\frac{\theta}{2}\right)\right) \end{bmatrix} \end{aligned}$$

Thus, we have

Theorem A.60. *If $A \in O(2)$, then there exists $\theta \in [0, 2\pi]$ such that $A = \begin{bmatrix} \cos(\theta) & sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ or there exists $\theta \in [0, 2\pi]$ such that $A = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}$. In other words, A is either a rotation about the origin or is a reflection about a line passing through origin.*

An even more exciting result says that actually every rigid motions (and hence every isometry) can be expressed as a composition of atmost three reflections. In the language of group theory, this means that the collection of reflections generate the group of rigid motions. Thus, reflections gain importance among other collection of functions.

Having characterised isometries or rigid motions, one can prove that the image of any triangle under an isomtery is another triangle satisfying SSS criterion. Moreover, given two triangles satisfying the SSS criterion, there exists an isometry that takes the first triangle to the second. This shows the equivalence of the intuitive notion with the criteria we study in school and thus gives a better understanding of Eucliedan geometry. So, in summary, we can think of Euclidean geometry as the study of reflections.