Contents

Lastly, do not assume that I am always right. Have a health scepticism and check out the statements. In case you find an typographical or mathematical mistakes, make sure that you ar right. This will give you immense confidence. (I enjoyed this aspect o reading a book when I was a student.) I also request you to send m your list of my mistakes and other suggestions for improvement i future editions.

Notation

We use the following notation.

$\mathbb{N}$ stands for the set of positive integers.

$\mathbb{Z}$ stands for the set of all integers.

$\mathbb{Q}$ stands for the set of all rational numbers.

$\mathbb{R}$ stands for the set of real numbers.

$\mathbb{C}$ stands for the set of all complex numbers.

If X is any set, then X^n stands for the cartesian product of X with itself n-times: $X^n \equiv \underbrace{X \times \ldots \times X}_{n \text{ times}}$

The notation := signifies that the left side object is defined by the right side. For example,

$$\mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$$

defines the set of positive real numbers. If we write $X := \mathbb{R}^n$, then it means that the set X is taken to be the set $\mathbb{R}^n$.

1. Systems of Linear Equations

1.1 A Motivating Example

Before beginning the theory of vector spaces, let us start with an example which will serve as a motivation as well as a precursor to what is to follow.

A word of caution: Do not be overly concerned with where m, n and x are from. If you wish you may assume that they are rational numbers.

Suppose in my neighbourhood there is an eccentric shopkeeper. He is convinced that north Indians eat more wheat than rice and south Indians eat more rice than wheat. So he offers only two standard packets. The first packet, call it N, has 5 kilograms of wheat and 2 kilograms of rice, whereas the second packet, call it S, has 2 kilograms of wheat and 5 kilograms of rice. Let us invent a shorthand. Whenever we write (m, n) we mean m kg of wheat and n kg of rice. Now if I buy 3 packets of N, it means that I am buying 15 kg of wheat and 6 kg of rice, that is, $3N = 3(5, 2) = (15, 6)$. Similarly, 2 packets of S means 4 kg of wheat and 10 kg of rice, that is, $2S = 2(2, 5) = (4, 10)$.

If I buy one of each of the packets, then I would have bought 7 kg of wheat and 7 kg of rice. That is,

$$N + S = (5, 2) + (2, 5) = (5 + 2, 2 + 5) = (7, 7).$$

Thus if I need m packets of N or n of S or both, there is no problem. Suppose I need 19 kg of wheat and 16 kg of rice. What do I do? I need to buy x packets of N and y packets of S so that $x(5, 2) + y(2, 5) = (19, 16)$. That is, $(5x, 2x) + (2y, 5y) = (19, 16)$ or $(5x + 2y, 2x + 5y) = (19, 16)$. Thus I end up solving a system of linear equations

$$\begin{aligned} 5x + 2y &= 19 \\ 2x + 5y &= 16. \end{aligned}$$

I solve this system using the methods we learnt in high school, I see that I need to buy 3 packets of N and 2 packets of S.

Suppose I need 34 kg of wheat and 1 kg of rice. Then I find I must buy 8 packets of N and -3 packets of S. What does this mean? I buy 8 packets of N and from these I make three packets of S and give them back to the shopkeeper. How nice of him to accept these!

A little more thinking would convince you that if you want to buy m kg of wheat and n kg of rice, you can always find x and y such that buying x packets of N and y packets of S does the job.

Of course if the shopkeeper is simple minded, he would be selling a packet e_1 containing 1 kg of wheat and 0 kg of rice and another packet e_2 containing 0 kg of wheat and 1 kg of rice, thereby making *our* life easier!

One instructive exercise in the same vein.

Exercise 1.1.1 Let us assume that the shopkeeper sells packets P_1 of 1 kg of wheat and 1 kg of rice and 1 kg of turdal, and P_2 containing 1 kg of wheat, 0 kg of rice and 1 kg of turdal and P_3 comprising 0 kg of wheat, 1 kg of rice and 1 kg of turdal. Is it possible for me to buy only one kilogram of turdal?

We also note the following. In the first example of $(5,2)$ and $(2,5)$, if I want to buy (m,n), there is exactly one way of doing this: Buy x packets of N and y packets of S where x and y are rational numbers.

On the contrary, consider the case when there is another packet $P = (1,1)$. One then easily checks that there are many ways: For example, $(7,7) = 1N + 1S = 7P$.

One last remark. Let the packet $P_1 = (5,2)$ be priced at Rs. R_1 and P_2 at Rs. R_2. Then if f is the price of m packets of P_1 and n packets of P_2, we see that $f(mP_1 + nP_2) = mf(P_1) + nf(P_2) = mR_1 + nR_2$.

1.2 Systems of Linear Equations

The recurring themes in Linear Algebra are:

(1) solutions of linear equations, and

(2) their geometric interpretation.

say that the point $(x, y) \in \mathbb{R}^2$ satisfies the equation or is a *solution* of the equation. The geometric interpretation of the equation is that the set of all points satisfying the equation forms a straight line in the plane through the point $(c/b, 0)$ and with slope $-a/b$.

The collection of linear equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\quad \vdots \quad &= \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

is called a system of m linear equations in n unknowns $x_1, \ldots, x_n$. Here a_{ij}, $b_i \in \mathbb{R}$ are given. We shall write this in a short form as

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \qquad 1 \le i \le m. \tag{1.2.1}$$

To solve this system is to find real numbers $x_1, \ldots, x_n$ which satisfy the system. Any n-tuple $x := (x_1, \ldots, x_n)$ which satisfies the system is called a *solution* of the system.

For example, consider the case when $m = 2$ and $n = 3$. Then,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2. \end{aligned}$$

Each of the above equations determines a plane, assuming at least one of $a_{1j} \neq 0$ and $a_{2j} \neq 0$. Hence, the set of solutions of the system is the set of all points which lie on both the planes, that is, the set of solutions is the intersection of the planes.

Consider the system

$$\begin{aligned} 2x - 3y &= 0 \\ -8x + 12y &= 1. \end{aligned}$$

We see that this system has no solution. Geometrically, each equation is a line and these are two distinct but parallel lines. Hence, their intersection is empty.

If each b_i in the above system (1.2.1) is zero, then the system is said to be *homogeneous*. The homogeneous system in n variables (unknowns) is

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= 0 \\ \vdots \qquad\quad \vdots \quad &= \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

Again, we adopt a shorthand notation

$$\sum_{j=1}^{n} a_{ij}x_j = 0, \qquad 1 \leq i \leq m. \tag{1.2.2}$$

Note that $0 = (0, \ldots, 0)$ is always a solution for the homogeneous system. This solution is called the *trivial* solution. We say $x = (x_1, \ldots, x_n)$ is a *non-trivial* solution if $(x_1, \ldots, x_n) \neq (0, \ldots, 0)$. That is, if there exists at least one i such that $x_i \neq 0$.

A non-trivial solution need not always exist. For example, consider the system of two linear equations in two unknowns: $x = 0$ and $y = 0$.

Let $(a_1, \ldots, a_n)$ be a solution of the homogeneous system (1.2.2). Then we see that $(\alpha a_1, \ldots, \alpha a_n)$ is again a solution of (1.2.2) for any $\alpha \in \mathbb{R}$. This has the following geometric interpretation in the case of the three dimensional space $\mathbb{R}^3$.

Let (r, s, t) be a solution of the system $ax + by + cz = 0$. That is, $ar + bs + ct = 0$. Then the solution set is a plane through the origin. Now the plane contains the two points $(0, 0, 0)$ and (r, s, t) and hence all the points on the line joining them. But any point on the line is of the form $\alpha(r, s, t)$ for some $\alpha \in \mathbb{R}$. For, recall the equation joining these two points is given by

$$\frac{x-0}{0-r} = \frac{y-0}{0-s} = \frac{z-0}{0-t}$$

say, a constant $-\alpha$. It is then immediate that $x = \alpha r$, $y = \alpha s$ and $z = \alpha t$.

Also if $(b_1, \ldots, b_n)$ is another solution of (1.2.2), then

$$(a_1 + b_1, \ldots, a_n + b_n)$$

is again a solution of (1.2.2). We describe this as *the set of solutions of a homogeneous system of linear equations is closed under addition and scalar multiplication.*

This suggests the following definition of "addition" and "scalar multiplication" on $\mathbb{R}^n$, the set of n-tuples of real numbers. If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ and $\alpha \in \mathbb{R}$, define $x + y := (x_1 + y_1, \ldots, x_n + y_n)$ and $\alpha x := (\alpha x_1, \ldots, \alpha x_n)$.

However, the set of solutions of a non-homogeneous system of linear equations need not be closed under addition and scalar multiplication. For example, consider the equation $3x - 2y = 1$. $a = (1, 1)$ is a solution but $\alpha(1, 1) = (\alpha, \alpha)$ is not a solution if $\alpha \neq 1$. Also, $b = (2, 2\frac{1}{2})$ is another solution but $(1, 1) + (2, 2\frac{1}{2}) = (3, 3\frac{1}{2})$ is not a solution. Note that $(0, 0)$ is not a solution.

The homogeneous system given by $\sum_{j=1}^{n} a_{ij}x_j = 0$, $1 \leq i \leq m$ is called the *associated homogeneous system* of (1.2.1).

Let S be the set of solutions of the non-homogeneous system and S_h be the set of solutions of the associated homogeneous system of equations. Assume $S \neq \emptyset$. S_h is always non-empty, as the trivial solution $(0, \ldots, 0) \in S_h$.

Let $x \in S$ and $y \in S_h$. We claim that for any $\alpha \in \mathbb{R}$, $x + \alpha y \in S$. Since $x \in S$ we have $\sum_{j=1}^{n} a_{ij}x_j = b_i$. Similarly, $\sum_{j=1}^{n} a_{ij}y_j = 0$ for $1 \leq i \leq m$. For $\alpha \in \mathbb{R}$ and $1 \leq i \leq m$, we have

$$\begin{aligned} \sum_{j=1}^{n} a_{ij}(x_j + \alpha y_j) &= \sum_{j=1}^{n} a_{ij}x_j + \alpha \sum_{j=1}^{n} a_{ij}y_j \\ &= \sum_{j=1}^{n} a_{ij}x_j \\ &= b_i \qquad \text{for } 1 \leq i \leq m. \end{aligned}$$

This proves our claim.

Let us try to understand the geometry underlying this observation. Let us look at a simple equation

$$x + y = 2. \tag{1.2.3}$$

This obviously represents a line in the plane $\mathbb{R}^2$ passing through the point $(1, 1)$ (see Figure 1.2.1). That is, the solution set S is a line passing through $(1, 1)$. The solution set S_h of the associated system $x + y = 0$ is the line passing through the origin whose "slope" is -1. Note that the line S_h is parallel to the line S so that to get the solution set S we need to take the line parallel to the solution set S_h of the associated homogeneous system and passing through any point in S (for example, in this case, $(1, 1)$). Similar geometric considerations apply to a more general equation of the type $ax + by = c$.

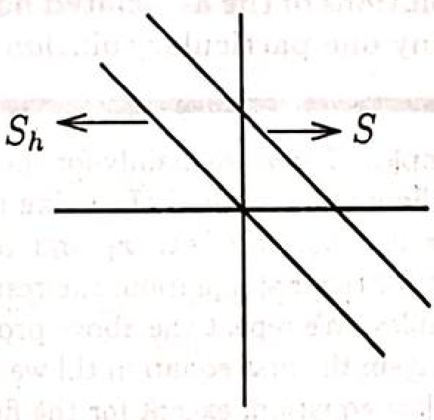


Figure 1.2.1 Solution of a linear system.

Exercise 1.2.1 Extend the geometric interpretation to

(1) an equation of the type $ax + by + cz = d$, and

(2) a system of two equations in three variables.

If $z := (z_1, \ldots, z_n) \in S$ and $x := (x_1, \ldots, x_n) \in S$, then $\sum_{j=1}^n a_{ij}z_j = b_i$ and $\sum_{j=1}^n a_{ij}x_j = b_i$. Therefore,

$$\sum_{j=1}^n a_{ij}(x_j - z_j) = \sum_{j=1}^n a_{ij}x_j - \sum_{j=1}^n a_{ij}z_j = b_i - b_i = 0.$$

That is, if x and z are any two solutions of the non-homogeneous system then $x - z$ is a solution of the homogeneous system. That is, $x - z \in S_h$.

Do you see the geometric meaning underlying this? For instance, in the case of Equation (1.2.3), it says that if we take *any* two points $p = (x, y)$ and $q = (x', y')$ on the line S, their "difference" $p - q := (x - x', y - y')$ lies on the parallel line through the origin.

Exercise 1.2.2 What is the analogue of the geometric meaning of the cases in Exercise 1.2.1?

We put the above observations into a single fact: Fix $x \in S$. Then if we define $x + S_h := \{x + y \mid y \in S_h\}$ what we saw above is that $x + S_h \subset S$. Also, for all $z \in S$, $z = x + (z - x) \in x + S_h$. This implies $S \subset x + S_h$. Therefore, $S = x + S_h$. This x is called a *particular solution* of (1.2.1). So we have the following fact:

To find all the solutions of (1.2.1) it is enough to find all the solutions of the associated homogeneous system and any one particular solution of (1.2.1).

We look at some examples. These are mainly for the purpose of reviewing the so-called Gaussian elimination method of solving linear equations. The idea is to eliminate the first variable, say x_1 and reduce the system to another in which, except for the first equation, the rest are linear equations in fewer number of variables. We repeat the above process with the system so obtained by deleting again the first equation till we are finally left with a single equation. In this last equation, except for the first x_i terms, the rest of the variables are treated as "free" and assigned arbitrary real numbers. The best way is to go through some of the examples below and work some more on your own.

Example 1.2.1 Consider the system

$$\begin{aligned} L_1 &:= & x + 2y + 3z &= 3 \\ L_2 &:= & 2x + 3y + 8z &= 4 \\ L_3 &:= & 3x + 2y + 17z &= 1. \end{aligned}$$

We let $\alpha L_i + \beta L_j$ stand for the addition of α times the ith equation with β times the jth equation. Now, $2L_1 - L_2$ and $-3L_1 + L_3$ are

$$\begin{aligned} 2x + 4y + 6z &= 6 \\ -2x - 3y - 8z &= -4 \end{aligned} \qquad \text{and} \qquad \begin{aligned} -3x - 6y - 9z &= -9 \\ 3x + 2y + 17z &= 1. \end{aligned}$$

We thus get a system of two linear equations in two variables y and z

$$\begin{aligned} y - 2z &= 2 \\ -4y + 8z &= -8. \end{aligned}$$

The last two equations are essentially the same as the second is -4 times the first. Thus, effectively we have only one equation in two variables: $y - 2z = 2$. We think of z as a "free" variable and assign to it any arbitrary value, say t. Then $y = 2 + 2t$. Substituting these values of y and z in the first equation of the given system, we get $x = -1 - 7t$. Thus the solution set S of the given system is

$$\begin{aligned} S &= \{(-1 - 7t, 2 + 2t, t) \mid t \in \mathbb{R}\} \\ &= (-1, 2, 0) + \{t(-7, 2, 1) \mid t \in \mathbb{R}\}. \end{aligned}$$

Note that $(-1, 2, 0)$ is a particular solution of the original system and $(-7t, 2t, t)$ is a solution of the homogeneous system for any $t \in \mathbb{R}$. We shall be brief in the rest of the examples.

Example 1.2.2 Consider the system

$$\begin{aligned} L_1 &:= & x + y + z &= 1 \\ L_2 &:= & 2x - y + z &= 2. \end{aligned}$$

The obvious thing to do is to consider $L_1 + L_2$ thereby eliminating the y-variable and getting the equation $3x + 2z = 3$. We treat z as the free variable and assign the value t to z: $z := t$. We then get $x = 1 - (2/3)t$. Substituting this value in the first of the given equations, we get $y = -t/3$. Thus the solution set S is given by

$$\begin{aligned} S &= \left\{\left(1 - \frac{2}{3}t, -\frac{1}{3}t, t\right) \;\middle|\; t \in \mathbb{R}\right\} \\ &= (1, 0, 0) + \left\{t\left(-\frac{2}{3}, -\frac{1}{3}, 1\right) \;\middle|\; t \in \mathbb{R}\right\} \end{aligned}$$

Note that $(1,0,0)$, a particular solution of the given system — a point of the three dimensional space — lies simultaneously on both the planes defined by the equation of the system. And $(-\frac{2}{3}, -\frac{1}{3}, 1)$ is a point of $\mathbb{R}^3$ which lies on the planes through the origin corresponding to the associated homogeneous system and hence all the points on the line joining it and the origin also lie on the planes through the origin.

Example 1.2.3 Consider the system

$$\begin{aligned} L_1 &:= x_1 - 2x_2 + x_3 + x_4 &= 1 \\ L_2 &:= x_1 - 2x_2 + x_3 - x_4 &= -1 \\ L_3 &:= x_1 - 2x_2 + x_3 + 5x_4 &= 5. \end{aligned}$$

Here $L_1 - L_2$ yields $2x_4 = 2$ and $L_1 - L_3$ yields $-4x_4 = -4$. Both these equations are equivalent and we see that $x_4 = 1$. Substitution of this value of x_4 in the above equations yields a single equation $x_1 - 2x_2 + x_3 = 0$. This is a linear equation in three variables and we may think of x_2 and x_3 as free variables. So, we let $x_2 = s$ and $x_3 = t$ so that $x_1 = 2s - t$. Hence the solution set is

$$\begin{aligned} S &:= \{(2s-t, s, t, 1) \mid s \in \mathbb{R}, t \in \mathbb{R}\} \\ &= \{(0,0,0,1) + (2s, s, 0, 0) + (-t, 0, t, 0) \mid s \in \mathbb{R},\ t \in \mathbb{R}\} \\ &= (0,0,0,1) + \{s(2,1,0,0) + t(-1,0,1,0) \mid s \in \mathbb{R},\ t \in \mathbb{R}\}. \end{aligned}$$

Example 1.2.4 Solve the system

$$\begin{aligned} x_1 + 0x_2 + 4x_3 - x_4 &= 7 \\ 0x_1 + x_2 - 2x_3 - 3x_4 &= 8. \end{aligned}$$

A look at the second equation shows that the first variable x_1 is already eliminated and we are left with a system of one equation in three variables x_2, x_3 and x_4. So we treat x_3 and x_4 as free variables. Let $x_3 := s$ and $x_4 := t$. Then $x_2 = 8 + (2s + 3t)$. Using this in the first equation, we get $x_1 = 7 - 4s + t$. Let S be the set of solutions for the above system. Then, $S = \{(7,8,0,0) + s(-4,2,1,0) + t(1,3,0,1) \mid s, t \in \mathbb{R}\}$. Note that $(7,8,0,0)$ is a solution of the system. But $(-4,2,1,0)$ and $(1,3,0,1)$ are not. However $(-4,2,1,0)$ and $(1,3,0,1)$ are solutions of the associated homogeneous system.

Example 1.2.5 Solve the system

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

Let S be the set of solutions for the above system. Then,

$$S = \{\alpha(3, -2, 1) \mid \alpha \in \mathbb{R}\}.$$

Consider the homogeneous system $\sum_{j=1}^{n} a_{ij}x_j = 0$ for $1 \leq i \leq m$. We would like to know when this system has non-trivial solutions. Before we prove a result in this direction, we look at some more examples.

Example 1.2.6 Consider the system of two equations in two unknowns: $3x + 4y = 0$ and $x + y = 0$. The set of solutions is $S = \{(0, 0)\}$. Thus this system has no non-trivial solutions.

Example 1.2.7 For the system $3x + 4y + z = 0$ and $x + y + z = 0$, we have $S = \{\alpha(-3, 2, 1) \mid \alpha \in \mathbb{R}\}$.

Example 1.2.8 For the system $x - y + 4z = 4$ and $2x + 6z = -2$, we have $S = \{(-1, -5, 0) + \alpha(-3, 1, 1) \mid \alpha \in \mathbb{R}\}$.

We see that a homogeneous system need not always have non-trivial solutions. However, we observe that if the number of unknowns is more than the number of equations then the system always has a non-trivial solution.

This is intuitively clear if we look at the geometric interpretation of the set of solutions. For example, the solutions of $ax + by = 0$ are all points lying on the line determined by the given equation.

Again, the system

$$\begin{aligned} a_1x + b_1y + c_1z &= 0 \\ a_2x + b_2y + c_2z &= 0 \end{aligned}$$

always has non-trivial solutions. These are two planes (or a single plane) passing through the origin, hence they intersect and the intersection is a line (or a plane) passing through the origin.

Theorem 1.2.1 *The system* $\sum_{j=1}^{n} a_{ij}x_j = 0$ *for* $1 \leq i \leq m$ *always has a non-trivial solution if* $m < n$.

Proof We first prove the result for $m = 1$ and $n > 1$:

$$a_{11}x_1 + \cdots + a_{1n}x_n = 0.$$

If each $a_i = 0$ then the equation is $0 = 0$. Any n-tuple $(x_1, \ldots, x_n)$ with $x_i \in \mathbb{R}$ is a solution. Assume therefore $a_{1j} \neq 0$ for some j. Then we can write

$$x_j = -a_{1j}^{-1}(a_{11}x_1 + \cdots + a_{1j-1}x_{j-1} + a_{1j+1}x_{j+1} + \cdots + a_{1n}x_n).$$

Hence, if we arbitrarily chose $\alpha_i \in \mathbb{R}$, for all $i \neq j$ and take

$$\alpha_j = -a_{1j}^{-1}(a_{11}\alpha_1 + \cdots + a_{1j-1}\alpha_{j-1} + a_{1j+1}\alpha_{j+1} + \cdots + a_{1n}\alpha_n)$$

then $(\alpha_1, \ldots, \alpha_n)$ is a solution of $\sum_{j=1}^{n} a_{ij}x_j = 0$. Thus for $m = 1$ and $n > 1$ we have a non-trivial solution.

We prove the result by induction on m. As induction hypothesis, we assume that if we are given a system of $(m-1)$ equations in k variables with $(m-1) < k$, there exists a non-trivial solution. We prove the result for m and n with $m < n$.

Let $\sum_{j=1}^{n} a_{ij}x_j = 0$ for $1 \leq i \leq m$ be a system of m equations in n unknowns with $m < n$. If each $a_{ij} = 0$, as before the system is $0 = 0$ for all i and so any n-tuple $(x_1, \ldots, x_n)$ is a solution. So assume that there exist (i, j) such that $a_{ij} \neq 0$. Let

$$\begin{array}{rcccc} L_1 & := & a_{11}x_1 + \cdots + a_{1n}x_n & = & 0 \\ L_2 & := & a_{21}x_1 + \cdots + a_{2n}x_n & = & 0 \\ \vdots & & \vdots \quad\quad \vdots & & \vdots \\ L_i & := & a_{i1}x_1 + \cdots + a_{in}x_n & = & 0 \\ \vdots & & \vdots \quad\quad \vdots & & \vdots \\ L_m & := & a_{m1}x_1 + \cdots + a_{mn}x_n & = & 0 \end{array}$$

Digression

We work out the case $(i, j) = (1, 1)$ and $m = 2$, $n = 3$ so as to understand the ideas and not be overwhelmed by the complicated notation below. If $a_{11} \neq 0$, then we can write $x_1 = -a_{11}^{-1}(a_{12}x_2 + a_{13}x_3)$. We substitute this value of x_1 in L_2 to get

$$-a_{21}(a_{11}^{-1})(a_{12}x_2 + a_{13}x_3) + a_{22}x_2 + a_{23}x_3 = 0.$$

Rearranging the terms, we get

$$[a_{22} + a_{21}(-a_{11}^{-1})a_{12}]x_2 + [a_{23} + a_{21}(-a_{11}^{-1})a_{13}]x_3 = 0.$$

This is a homogeneous system of one equation in two unknowns x_2 and x_3 and has a non-trivial solution by induction hypothesis. So, if $(x_2, x_3) \neq (0, 0)$ is a non-trivial solution, it is easily verified that x_1 defined above along with x_2 and x_3 satisfies the original system.

End of Digression

Since $a_{i1}x_1 + \cdots + a_{in}x_n = 0$ and $a_{ij} \neq 0$, we have

$$x_j = -a_{ij}^{-1}(a_{i1}x_1 + \cdots + a_{i,j-1}x_{j-1} + a_{i,j+1}x_{j+1} + \cdots + a_{in}x_n).$$

We substitute this value of x_j in the equation $L_k = 0$ for $k \neq i$. We get a new system of $(m-1)$ equations L'_k in $(n-1)$ variables

$$x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n$$

as follows: For $1 \leq k \leq m$, $k \neq i$,

$$L'_k = \sum_{r \neq j} [a_{kr} + a_{kj}(-a_{ij}^{-1})a_{ir}]x_r = 0.$$

Since $m - 1 < n - 1$, by induction hypothesis on m, we get a non-trivial solution $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n$ of this system. In particular, $x_k \neq 0$ for some $k \neq j$. We take $\alpha_j := x_j$ as above:

$$\alpha_j = -a_{ij}^{-1}\Big(\sum_{r \neq j} a_{ir}\alpha_r\Big).$$

We claim $\alpha_1, \dots, \alpha_{j-1}, \alpha_j, \dots, \alpha_n$ is a non-trivial solution. For $1 \leq k \leq m$,

$$\begin{aligned} L_k &= \sum_{r \neq j} a_{kr}x_r + a_{kj}\alpha_j \\ &= \sum_{r \neq j} a_{kr}\alpha_r + a_{kj}(-a_{ij}^{-1}) \sum_{s \neq j} a_{is}\alpha_s \\ &= \sum_{r \neq j} [a_{kr} + a_{kj}(-a_{ij}^{-1})a_{ir}]\alpha_r \\ &= L'_k = 0 \qquad \text{for } k \neq i. \end{aligned}$$

By our very definition of α_j, $(\alpha_1, \dots, \alpha_n)$ is a solution of L_i:

$$\begin{aligned} \sum_r a_{ir}\alpha_r &= \sum_{r \neq j} a_{ir}\alpha_r + (a_{ij})\Big(-a_{ij}^{-1}\Big(\sum_{r \neq j} a_{ir}\alpha_r\Big)\Big) \\ &= \sum_{r \neq j} (a_{ir} - a_{ir})\alpha_r = 0. \end{aligned}$$

$(\alpha_1, \dots, \alpha_n)$ is non-trivial since $\alpha_k \neq 0$ for some $k \neq j$ (given by the induction hypothesis). Thus $\alpha_1, \dots, \alpha_n$ is a non-trivial solution of the original system.

$\square$

Exercise 1.2.3 Show that the solution set of the equation $x + y - z = 0$ can be written as $\{s(-1,1,0) + t(1,0,1) \mid s, t \in \mathbb{R}\}$.

Exercise 1.2.4 Find the points of intersection of

$$\begin{aligned} x^2 + y^2 &= 13 \quad \text{and} \\ 3x^2 + 4y^2 &= 48. \end{aligned}$$

Exercise 1.2.5 Let $P(x) = a_0 + a_1 x + a_2 x^2$. Choose a_i such that $P(1) =$ $P(2) = b_2$, $P(3) = b_3$. Is this choice unique?

Exercise 1.2.6 Give a system of linear equations having

(1) $(1, 0, 0)$ as "the" only solution.

(2) $(1, 0, 0)$ and $(0, 1, 0)$ as solutions.

(3) $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ as solutions.

2. Vector Spaces

2.1 Definition and Examples

Do you recall the "addition" and "scalar multiplication" we defined on the space of solutions of a homogeneous system of linear equations? We define a vector space to be a set on which similar operations are defined. More precisely, we have

Definition 2.1.1 A non-empty set V is said to be a *vector space* **over** $\mathbb{R}$ (or a real vector space) if there exist maps $+\colon V \times V \to V$, defined by $(x, y) \mapsto x+y$, called *addition*, and $\cdot : \mathbb{R} \times V \to V$, defined by $(\alpha, x) \mapsto \alpha \cdot x$, called *scalar multiplication*, satisfying the following properties:

(i) $x + y = y + x$ (commutativity of addition).

(ii) $(x + y) + z = x + (y + z)$ (associativity of addition).

(iii) There exists $\mathbf{0} \in V$ such that $x + \mathbf{0} = x = \mathbf{0} + x$ (existence of additive identity).

(iv) For every $x \in V$ there exists $y \in V$ such that $x + y = \mathbf{0} = y + x$. This y is denoted by $-x$ (existence of additive inverse).

(v) $\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y$.

(vi) $(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x$.

(vii) $(\alpha\beta) \cdot x = \alpha(\beta \cdot x)$.

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

We adopt the following standard notation: $x + (-y)$ is written as $x - y$ for all $x, y \in V$ and for $\alpha \in \mathbb{R}$ and $x \in V$ we write αx for $\alpha \cdot x$. Hereafter, by a vector space we mean a vector space over $\mathbb{R}$. Elements of a vector space V are called *vectors* of V. The addition in a vector space is referred to as the *vector addition*. $\mathbf{0}$ is called the zero vector. $x \in V$ is a *nonzero* vector if $x \neq \mathbf{0}$. A vector αv is called a *scalar multiple* of $v \in V$.

Theorem 2.1.1 *In a vector space V, we have*

(1) $0 \cdot x = 0$ *for all* $x \in V$.

(2) *There is a unique additive identity. That is, if* 0 *and* $0'$ *are such that* $x + 0 = x$ *and* $x + 0' = x$ *for all* $x \in V$, *then* $0 = 0'$.

(3) *The additive inverse is unique. That is, if for a given* x, *there are* $y, y' \in V$ *such that* $x + y = 0$ *and* $x + y' = 0$, *then* $y = y'$.

(4) $(-1) \cdot x = -x$, *the negative element such that* $x + (-x) = 0$ *for* $x \in V$.

(5) $\alpha \cdot 0 = 0$ *for all* $\alpha \in \mathbb{R}$ *and* $0 \in V$.

(6) *If* $\alpha \cdot x = 0$ *for* $\alpha \in \mathbb{R}$ *and* $x \in V$, *then either* $\alpha = 0$ *or* $x = 0$.

Proof (1) We have

$$0 \cdot x = (0 + 0) \cdot x = 0 \cdot x + 0 \cdot x \tag{2.1.1}$$

by property (vi) in Definition 2.1.1. Adding $-0 \cdot x$, an additive inverse of $0 \cdot x$, which exists according to (iv), to both sides of Equation (2.1.1), we get

$$\begin{aligned} 0 &= 0 \cdot x + (-0 \cdot x) && \text{by (iv)} \\ &= (0 \cdot x + 0 \cdot x) + (-0 \cdot x) && \text{by Equation (2.1.1)} \\ &= 0 \cdot x + (0 \cdot x + (-0 \cdot x)) && \text{by (ii)} \\ &= 0 \cdot x + 0 && \text{by (iv)} \\ &= 0 \cdot x && \text{by (iii).} \end{aligned}$$

(2) Assume that $x + 0 = x = x + 0'$ for all $x \in V$. In particular $0 + 0' = 0$ since $0'$ is an additive identity. Similarly, $0 + 0' = 0'$ since 0 is an additive identity. Hence $0 = 0'$.

(3) Assume that $x + y = 0 = x + y'$. We add y' to all the sides of this equation, use commutativity and associativity of the addition:

$$\begin{aligned} & x + y &&= 0 \\ \Longrightarrow\ & y' + (x + y) &&= y' + 0 \\ \Longrightarrow\ & (y' + x) + y &&= y' + 0 \\ \Longrightarrow\ & (x + y') + y &&= y' \\ \Longrightarrow\ & 0 + y &&= y' \\ \Longrightarrow\ & y &&= y'. \end{aligned}$$

(4) $(-1) \cdot x + x = (-1) \cdot x + 1 \cdot x = (-1 + 1) \cdot x = 0 \cdot x = 0$ so that $(-1) \cdot x = -x$.

(5) We have $\mathbf{0} + \mathbf{0} = \mathbf{0}$ from (iii). Hence $\alpha \cdot \mathbf{0} = \alpha \cdot (\mathbf{0} + \mathbf{0}) = \alpha \cdot \mathbf{0} + \alpha \cdot \mathbf{0}$ by (v). Adding $-\alpha \cdot \mathbf{0}$ to both sides and using (iii) and (ii), we get

$$\begin{aligned} \mathbf{0} &= \alpha \cdot \mathbf{0} + (-\alpha \cdot \mathbf{0}) \\ &= (\alpha \cdot \mathbf{0} + \alpha \cdot \mathbf{0}) + (-\alpha \cdot \mathbf{0}) \\ &= \alpha \cdot \mathbf{0} + (\alpha \cdot \mathbf{0} + (-\alpha \cdot \mathbf{0})) \\ &= \alpha \cdot \mathbf{0} + \mathbf{0} \\ &= \alpha \cdot \mathbf{0}. \end{aligned}$$

(6) If $\alpha x = \mathbf{0}$ and $\alpha \neq 0$, then we multiply both sides of $\alpha x = \mathbf{0}$ by α^{-1} to get

$$\alpha^{-1}(\alpha x) = \alpha^{-1} \cdot \mathbf{0} = (\alpha^{-1}\alpha) \cdot x = \alpha^{-1} \cdot \mathbf{0}.$$

The extreme left term of this equation is x by (vii) and (viii), and the extreme right term is $\mathbf{0}$ by (5).

□

Remark 2.1.1 For every $x \in V$ there is a unique $y \in V$ such that $x+y = \mathbf{0}$. This y is given by $y = (-1)x$ so that $-x = (-1)x$.

Remark 2.1.2 We shall denote the additive identity $\mathbf{0}$ by 0 henceforth. The context will make it clear whether 0 denotes the real number zero or the additive identity $\mathbf{0}$. The reader is urged to go through the above proof using the same symbol 0 for both and try to understand the proof.

We now look at some examples of vector spaces.

Example 2.1.1 Let X be a non-empty set. Let

$$V = \mathcal{F}(X, \mathbb{R}) = \{f : X \to \mathbb{R}\}$$

be the set of real valued functions on the set X. For $f, g \in V$, we wish to define $f + g \in V$. Thus $f + g$ must be a function from X to $\mathbb{R}$. So to define $f + g$ it is enough if we say what its value at any arbitrary point $x \in X$ is. We define $(f + g)(x) := f(x) + g(x)$. Similar considerations suggest the definition of $\alpha f \in V$, for $\alpha \in \mathbb{R}$ and $f \in V$, as the function whose value at x is given by $(\alpha f)(x) = \alpha f(x)$. Then V is a vector space over $\mathbb{R}$.

Example 2.1.2 Let X be any nonempty set and let $F_0(X, \mathbb{R})$ denote the set of functions from X to $\mathbb{R}$ such that the set $\{x \in X \mid f(x) \neq 0\}$ is finite (this set may depend on f). Thus, $f \in F_0(X, \mathbb{R})$ if and only if $f(x) = 0$ except for finitely many $x \in X$. Clearly, $F_0(X, \mathbb{R})$ is a subset of $\mathcal{F}$ of Example 2.1.1. We define addition and scalar multiplication as earlier: $f + g$ and αf are elements of $F_0(X, \mathbb{R})$ whose values at $x \in X$ are given by $(f + g)(x) := f(x) + g(x)$ and $(\alpha f)(x) = \alpha f(x)$. Note that $f + g$ and αf lie in $F_0(X, \mathbb{R})$, that is, the sets $\{x \in X \mid (f + g)(x) \neq 0\}$ and $\{x \in X \mid \alpha f(x) \neq 0\}$ are finite subsets of X.

Example 2.1.3 Let $V = \mathbb{R}^n = \{(x_1, \dots, x_n) \mid x_i \in \mathbb{R}\}$. For

$$x = (x_1, \dots, x_n),\ y = (y_1, \dots, y_n)$$

define

$$x + y := (x_1 + y_1, \dots, x_n + y_n)$$

and

$$\alpha x = (\alpha x_1, \dots, \alpha x_n)$$

for $\alpha \in \mathbb{R}$. Then V is a vector space under these operations. In particular, $\mathbb{R}^1 = \mathbb{R}$ is a vector space over $\mathbb{R}$. Note that the addition and scalar multiplication on $\mathbb{R}^n$ are the same as the ones we had defined for solutions of a homogeneous system of linear equations.

This example is special case of Example 2.1.1: Take $X = \{1, \dots, n\}$ and define $f : X \to \mathbb{R}$ by $x_i = f(i)$ for $1 \le i \le n$. Then the map

$$T\colon f \mapsto (f(1), \dots, f(n))$$

is a bijection of $F(X, \mathbb{R})$ and $\mathbb{R}^n$. What is interesting is the fact that the way addition (respectively scalar multiplication) on $F(X, \mathbb{R})$ corresponds to that on $\mathbb{R}^n$ under this bijection.

Geometric interpretation of vector addition in $\mathbb{R}^2$

Most often we look at $\mathbb{R}$, $\mathbb{R}^2$ and $\mathbb{R}^3$ to understand the geometric meaning underlying the concepts. In such attempts, we shall assume some very basic knowledge of analytic geometry. The first in this direction is the geometric interpretation of addition of two vectors in $\mathbb{R}^2$. Let $x, y \in \mathbb{R}^2$, where $x = (x_1, x_2)$, $y = (y_1, y_2)$. Then $x + y = (x_1 + y_1, x_2 + y_2)$. In Figure 2.1.1, $A = (x_1, x_2)$, $B = (y_1, y_2)$. Form the parallelogram with sides OA and OB. Let the fourth vertex be C with coordinates (α, β).

Let M be the point of intersection of the diagonals of the parallelogram $OACB$. Then M is the midpoint of BA as well as OC. Since M is the midpoint of BA,

$$M = \frac{1}{2}(B + A) = \left(\frac{x_1 + y_1}{2}, \frac{x_2 + y_2}{2}\right).$$

Since M is the midpoint of OC, we have $M = \frac{1}{2}[(0,0) + (\alpha, \beta)]$. Comparing coordinates we get $\frac{x_1+y_1}{2} = \frac{\alpha}{2}$ and $\frac{x_2+y_2}{2} = \frac{\beta}{2}$ or $\alpha = (x_1 + y_1)$ and $\beta = x_2 + y_2$, that is, $C = A + B$. Thus $x + y$ is the vector represented by the diagonal of the parallelogram spanned by x and y.

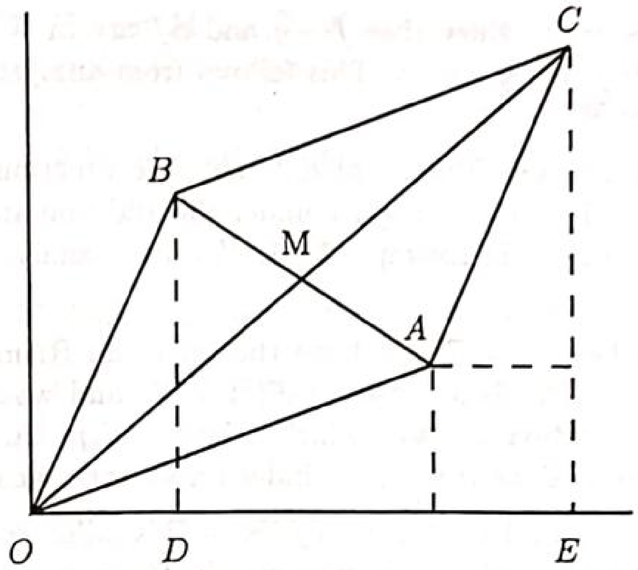


Figure 2.1.1 Addition of vectors.

Exercise 2.1.1 Let $V := \mathcal{S} := \{(x_n) \mid x_n \in \mathbb{R}\}$ be the set of all real sequences. Then V is a vector space over $\mathbb{R}$ under the following operations:

$$\begin{aligned} (x_n) + (y_n) &:= (x_n + y_n) \\ \alpha(x_n) &:= (\alpha x_n). \end{aligned}$$

Note that again this is a special case of Example 2.1.1 if we take $X = \mathbb{N}$.

Exercise 2.1.2 Let $\mathbf{C}$ be the set of all convergent real sequences. Note that $\mathbf{C}$ is a subset of $\mathcal{S}$ of Exercise 2.1.1. We define the addition and scalar multiplication as in Exercise 2.1.1. Then $\mathbf{C}$ is a vector space. The subtle point of this assertion is that we have to show that if $(x_n) \in \mathbf{C}$ and $(y_n) \in \mathbf{C}$ then $(x_n) + (y_n)$ lies in $\mathbf{C}$ — in other words, $(x_n + y_n)$ is convergent if (x_n) and (y_n) are so. A similar fact is needed for $\alpha(x_n)$. These are well-known facts from Analysis. Thus to show that $\mathbf{C}$ is a vector space we need results from analysis!

Exercise 2.1.3 Let $\mathbf{C}_0$ be the set of null sequences, that is,

$$\{(x_n) \mid \lim x_n = 0\}.$$

Note that $\mathbf{C}_0 \subseteq \mathbf{C} \subseteq \mathcal{S}$. Then $\mathbf{C}_0$ is a vector space under the same operations as in Exercise 2.1.1.

Exercise 2.1.4 Let $V := \mathcal{C}([a, b])$, be the set of all real valued continuous functions on $[a, b]$. Note that this is a subset of $\mathcal{F}([a, b], \mathbb{R})$ of Example 2.1.1. We define the addition and scalar multiplication as in Example 2.1.1. Again the crucial steps are to show that $f+g$ and αf are in V. That is, they are continuous if f and g are so. This follows from analysis. Then V is a vector space over $\mathbb{R}$.

Exercise 2.1.5 The set $\mathcal{D}([0,1])$ of differentiable functions on $[0,1]$ is a subset of $\mathcal{C}([0,1])$. It is a vector space under the addition and scalar multiplication of functions as in Example 2.1.1. You need analysis to prove this assertion.

Exercise 2.1.6 Let $V := \mathcal{R}([a,b])$, be the set of all Riemann integrable functions on $[a,b]$. This is a subset of $\mathcal{F}([a,b],\mathbb{R})$ and we define addition and scalar multiplication in a way which is by now familiar to you. Again you need analysis to show that this is indeed a vector space.

Exercise 2.1.7 Recall that a function $f\colon \mathbb{R}\to\mathbb{R}$ is called *even* (respectively *odd*) if $f(-x)=f(x)$ for all $x\in\mathbb{R}$ (respectively $f(-x)=-f(x)$ for $x\in\mathbb{R}$). Let $\mathcal{F}_+(\mathbb{R},\mathbb{R})$ (respectively, $\mathcal{F}_-(\mathbb{R},\mathbb{R})$) denote the set of even (respectively odd) functions from $\mathbb{R}$ to $\mathbb{R}$. Are they vector spaces under the obvious definitions?

Exercise 2.1.8 Let $V=\mathcal{P} := \left\{\sum_{i=0}^n a_iX^i \mid a_i\in\mathbb{R},\ n\in\mathbb{N}\right\}$, be the set of all polynomials in one variable with real coefficients. The addition and scalar multiplication are the usual ones:

$$\left(\sum_{i=0}^m a_iX^i\right)+\left(\sum_{j=0}^n b_jX^j\right) := \sum_r (a_r+b_r)X^r$$

where $a_r=0$ if $r>m$ and $b_r=0$ if $r>n$. Also,

$$\alpha\left(\sum_i a_iX^i\right) := \sum_i \alpha a_iX^i.$$

Then V is a vector space over $\mathbb{R}$.

Exercise 2.1.9 Let $V=\mathcal{P}_n := \left\{\sum_{i=0}^n a_iX^i \mid a_i\in\mathbb{R}\right\}$, be the set of all polynomials of degree $\le n$ with real coefficients. This is a subset of $\mathcal{P}$ of Exercise 2.1.8. So we define the operations as in $\mathcal{P}$. Then V is a vector space over $\mathbb{R}$.

Exercise 2.1.10 Let V denote the set of all polynomials exactly of degree n. Is it a vector space under the usual addition and scalar multiplication of polynomials?

Exercise 2.1.11 Let V be the set of all solutions of a system of m homogeneous linear equations in n variables with real coefficients. Then V is a vector space over $\mathbb{R}$ under obvious operations. Can you realize this as a subset of $\mathbb{R}^n$?

Exercise 2.1.12 The set $M_{n\times m}(\mathbb{R})$ of all $n \times m$ matrices with real entries is a vector space over $\mathbb{R}$ with the operation of addition of matrices and scalar multiplication of matrices: $(a_{ij}) + (b_{ij}) := (a_{ij} + b_{ij})$ and $\alpha(a_{ij}) := (\alpha a_{ij})$.

Can you set up a bijection from $M_{n\times m}(\mathbb{R})$ onto $\mathcal{F}(X, \mathbb{R})$ where

$$X := \{1, \ldots, n\} \times \{1, \ldots, m\}?$$

We let $M(n, \mathbb{R})$ denote the set of $n \times n$ matrices with real entries.

Exercise 2.1.13 Recall that a real $n \times n$ matrix $A = (a_{ij})$ is said to be *symmetric* if $a_{ij} = a_{ji}$ for all $1 \le i, j \le n$. Let $\mathbf{S}_n$ denote the set of all $n \times n$ symmetric real matrices. Then under the operations of matrix addition and scalar multiplication as in Exercise 2.1.12, $\mathbf{S}_n$ is a vector space.

Exercise 2.1.14 A real matrix $A = (a_{ij})$, $1 \le i, j \le n$ is *skew-symmetric* if $a_{ij} = -a_{ji}$ for all $1 \le i, j \le n$. If $\mathbf{A}_n$ denotes the set of all skew-symmetric matrices, then $\mathbf{A}_n$ is a vector space under obvious addition and scalar multiplication. Note that both $\mathbf{S}_n$ and $\mathbf{A}_n$ are subsets of $M(n, \mathbb{R})$.

Exercise 2.1.15 Let V, W be vector spaces. Let us form the Cartesian product $V \times W$. Define addition and scalar multiplication on $V \times W$ as follows:

$$\begin{aligned} (v_1, w_1) + (v_2, w_2) &= (v_1 + v_2, w_1 + w_2), \quad (v_i, w_i) \in V \times W, \quad i = 1, 2 \\ \alpha(v, w) &= (\alpha v, \alpha w), \qquad \alpha \in \mathbb{R}, \quad (v, w) \in V \times W. \end{aligned}$$

Then $V \times W$ is a vector space. This vector space is usually denoted by $V \oplus W$ and called *direct sum* of V and W.

Exercise 2.1.16 Extend the construction in Exercise 2.1.15 to define the direct sum $V_1 \oplus \cdots \oplus V_n$ of n vector spaces V_i, $1 \le i \le n$. Do you recognize $\mathbb{R} \oplus \cdots \oplus \mathbb{R}$?

Exercise 2.1.17 Let V be a vector space. On $V \times V$, define $+$, and $\cdot$ as follows:

(1)
$$\begin{aligned} (v_1, w_1) + (v_2, w_2) &= (v_1 + w_2, w_1 + v_2) \\ \alpha(v, w) &= (\alpha v, \alpha w) \quad \alpha \in \mathbb{R}, (v, w) \in V \times V \end{aligned}$$

(2)
$$\begin{aligned} (v_1, w_1) + (v_2, w_2) &= (v_1 + w_1, v_2 + w_2) \\ \alpha(v, w) &= (\alpha v, \alpha w). \end{aligned}$$

Then $V \times V$ is *not* a vector space as the addition violates some of the conditions (i) – (iv) in Definition 2.1.1.

Exercise 2.1.18 Let V be a vector space and X be any (non-empty) set. Let W be the set of functions $f : X \to V$. On W define addition and scalar multiplication as follows:

$$\begin{aligned} (f+g)(x) &= f(x)+g(x), \quad f,g \in W, \quad x \in X \\ (\alpha \cdot f)(x) &= \alpha f(x), \quad \alpha \in \mathbb{R}, \quad x \in X. \end{aligned}$$

Then W is a vector space.

Note the similarity of this exercise with that of Example 2.1.1.

Exercise 2.1.19 Let $X := \{\star\}$ be a singleton set and let V be a vector space. Let $W = \{\star\} \times V$. We can turn W into a vector space as follows:

$$\begin{aligned} (\star, v_1) + (\star, v_2) &= (\star, v_1 + v_2), \quad v_1, v_2 \in V \\ \alpha(\star, v) &= (\star, \alpha v), \quad \alpha \in \mathbb{R}, \quad v \in V. \end{aligned}$$

Exercise 2.1.20 Let $V := \mathbb{Q}$. On $\mathbb{Q}$ we have a natural addition, namely, the addition of rational numbers. However, if $\alpha \in \mathbb{R}$ is irrational and $r \in \mathbb{Q}$ then $\alpha r \in \mathbb{R}$ but not in $\mathbb{Q}$. Then V is not a vector space over $\mathbb{R}$.

Exercise 2.1.21 Let $\mathbb{C}$ denote the set of complex numbers:

$$\mathbb{C} := \{z := x + iy \mid x, y \in \mathbb{R}\}.$$

We identify $\mathbb{R}$ as a subset of $\mathbb{C}$ consisting of the complex numbers whose imaginary part is zero. Recall the addition of complex numbers:

$$(x+iy) + (u+iv) := (x+u) + i(y+v)$$

and the multiplication of complex numbers:

$$(a+ib)(x+iy) := (ax-by) + i(ay+bx).$$

We turn $\mathbb{C}$ into a vector space over $\mathbb{R}$ by declaring vector addition the same as addition of complex numbers as above and the scalar multiplication $\alpha \cdot z$ is the multiplication of complex numbers α and z: $\alpha \cdot z := \alpha x + i\alpha y$ where $z := x + iy$. Under these operations, $\mathbb{C}$ becomes a vector space over $\mathbb{R}$. This follows from the commutativity, associativity and distributivity of the operations in $\mathbb{C}$.

Exercise 2.1.22 Let P_i, $1 \le i \le n$ be continuous functions on $[a,b] \subset \mathbb{R}$. Let V be the set of n-times continuously differentiable solutions f on $[a,b]$ of a linear differential equation

$$y^{(n)} + P_1(x)y^{(n-1)} + \cdots + P_n(x)y = 0.$$

If f and g are solutions of the differential equation, we let

$$(f+g)(x) = f(x) + g(x) \text{ and } \alpha \cdot f(x) = \alpha f(x)$$

for $x \in [a,b]$ and $f \in V$. Then V is a vector space.

The rest of the section may be omitted in the first reading.

Remark 2.1.3 One can define vector spaces over $\mathbb{C}$: Scalar multiplication now will involve complex numbers. In Definition 2.1.1, if we replace $\mathbb{R}$ by $\mathbb{C}$ then what we get is called a vector space over $\mathbb{C}$ or a complex vector space. Examples are obtained from some of our earlier examples (and exercises), by replacing $\mathbb{R}$ with $\mathbb{C}$. For instance, if we consider the set of complex valued functions $\mathcal{F}(X,\mathbb{C})$ on a nonempty set X with the operations

$$\begin{array}{lll} (f+g)(x) & := & f(x)+g(x), \quad x \in X, \quad f,g \in \mathcal{F}(X,\mathbb{C}) \\ (\alpha f)(x) & := & \alpha f(x), \qquad\qquad x \in X, \quad f \in \mathcal{F}(X,\mathbb{C}),\ \alpha \in \mathbb{C}, \end{array}$$

then $\mathcal{F}(X,\mathbb{C})$ is a complex vector space. One can similarly construct complex vector spaces $\mathbb{C}^n$, $M(n,\mathbb{C})$ etc.

Remark 2.1.4 One can similarly replace $\mathbb{R}$ by $\mathbb{Q}$ in Definition 2.1.1 and get vector spaces over $\mathbb{Q}$. Can you think of vector spaces over $\mathbb{Q}$? Do you see that any vector space over $\mathbb{R}$ is a vector space over $\mathbb{Q}$? Can you think of a vector space over $\mathbb{Q}$ which is a not a vector space over $\mathbb{R}$? Replace the pair $(\mathbb{Q},\mathbb{R})$ by the pair $(\mathbb{R},\mathbb{C})$ in the above questions and answer them. Could we have replaced $\mathbb{Q}$ by $\mathbb{C}$ in the original set of questions?

2.2 Vector Subspaces

An astute reader would have noticed something very striking in our list of exercises of vector spaces in Section 2.1. We seem to have basically a few vector spaces and the rest were subsets of them. For example, the solution sets of a homogeneous system of m equations (with real coefficients) in n variables is a subset of $\mathbb{R}^n$ (see Exercise 2.1.11). Similarly, the vector spaces $\mathcal{C}([a,b],\mathbb{R})$ of Exercise 2.1.4, $\mathcal{R}([a,b],\mathbb{R})$ of Exercise 2.1.6 and $\mathcal{D}[0,1]$ of Exercise 2.1.5 are subsets of the same vector space $\mathcal{F}([a,b],\mathbb{R})$. Furthermore, from analysis, we know that $\mathcal{C}([a,b],\mathbb{R}) \subset \mathcal{R}([a,b],\mathbb{R})$ and $\mathcal{D}([0,1],\mathbb{R}) \subset \mathcal{C}([0,1],\mathbb{R})$. Again, $\mathbf{S}_n$ of Exercise 2.1.13 and $\mathbf{A}_n$ of Exercise 2.1.14 are subsets of $M(n,\mathbb{R})$. Moreover, addition and scalar multiplications on the subsets were the same as the ones on the bigger set. To put it differently, these subsets enjoy the property that whenever we add any two elements of the subset we again get an element of the subset and a scalar multiple of an element of the subset lies again in the subset. One says that *the subset is closed under the vector addition and scalar multiplication.* These observations suggest the following definition.

A *vector subspace* W of a vector space V over $\mathbb{R}$ is any non-empty subset $W \subset V$ which is closed under the addition and scalar multiplication on V.

More precisely, we have the following definition:

Definition 2.2.1 Let W be a non-empty subset of a vector space V. Then W is said to be a *vector subspace* (or simply a subspace) of V if W itself is a vector space under the operations induced from V. That is,

(i) $0 \in W$.

(ii) If $w_1, w_2 \in W$, then $w_1 + w_2 \in W$.

(iii) $\alpha \in \mathbb{R}$ and $w \in W$, then $\alpha w \in W$.

Exercise 2.2.1 Show that the following subsets W are vector subspaces of V:

(1) $W = \mathcal{C}([0,1])$ and $V = \mathcal{F}([0,1],\mathbb{R})$.

(2) $W = \mathcal{C}([0,1])$ and $V = \mathcal{R}([0,1],\mathbb{R})$.

(3) $W = \mathcal{D}([0,1])$ and $V = \mathcal{C}([0,1],\mathbb{R})$.

(4) $W = \mathbf{C}$ and $V = \mathcal{S}$.

(5) $W = \mathbf{S}_n$ and $V = M(n,\mathbb{R})$.

Exercise 2.2.2 Show that the set $\{0\}$ consisting of the zero vector in any vector space is a vector subspace.

Exercise 2.2.3 Find some more examples of vector spaces and vector subspaces from our list in the last section.

Exercise 2.2.4 Fix $x_0 \in X$. Let $S = \{f : X \to \mathbb{R} \mid f(x_0) = 0\}$. Then S is a vector subspace of $\mathcal{F}(X,\mathbb{R})$.

We now want to address the following problem: How does one "create" vector subspaces out of a given vector space? We start with the simplest case.

Let V be a vector space. Let $v \in V$. We want a vector subspace V_0 of V which contains v. We can take $V_0 = V$! So what we want is a vector subspace V_0 containing v which is as "small" as possible. If such a V_0 exists, since $v \in V_0$, all scalar multiples $\alpha v \in V_0$ for any $\alpha \in \mathbb{R}$. As

$$\alpha v + \beta v = (\alpha + \beta)v, \ -v = (-1)v$$

we see that if we take $V_0 = \{\alpha v \mid \alpha \in \mathbb{R}\}$, then V_0 is a vector subspace containing v. Let us make sure that we understand this. If $x = \alpha v \in V_0$ and $y = \beta v \in V_0$, then, what we are supposed to show is that $x + y \in V_0$. But $x + y = \alpha v + \beta v = (\alpha + \beta)v = \gamma v \in V_0$ where $\gamma = \alpha + \beta$. Similarly we show that if $v \in V_0$, $\alpha v \in V_0$ for any $\alpha \in \mathbb{R}$. Also, V_0 is the smallest in the sense that if W is a vector subspace containing v, then $W \supseteq V_0$. This is clear: Whenever $v \in W$, since W is a vector subspace, $\alpha v \in W$ for any $\alpha \in \mathbb{R}$. Thus any arbitrary element of V_0 lies in W. Hence $V_0 \subset W$. V_0 is usually denoted by $\mathbb{R}v$.

It is worth going through the last paragraph once again as it forms the heart of the matter to come. Before we go any further let us look at the geometric meaning of V_0. Let $v := (r, s) \in \mathbb{R}^2$ be a nonzero vector in $\mathbb{R}^2$. Then the smallest subspace V_0 containing v is the set

$$\{tv \mid t \in \mathbb{R}\} = \{(tr, ts) \mid t \in \mathbb{R}\}.$$

This is nothing but the line through the origin and the point (r, s) in $\mathbb{R}^2$. Do you see this? The line joining the origin and (r, s) is given by the equation

$$\frac{x-0}{0-r} = \frac{y-0}{0-s}.$$

Hence any point of this line is given by (tr, ts) for some $t \in \mathbb{R}$.

Exercise 2.2.5 What is the geometric object corresponding to the smallest subspace V_0 containing a nonzero vector $v = (r, s, t) \in \mathbb{R}^3$?

Now we let $v, w \in V$ and ask for the smallest vector subspace V_0 containing v and w. As earlier, $\alpha v \in V_0$ and $\beta w \in V_0$ for any $\alpha, \beta \in \mathbb{R}$. Since V_0 is a vector subspace, $\alpha v + \beta w \in V_0$. This suggests taking

$$V_0 = \{\alpha v + \beta w \mid \alpha \in \mathbb{R}, \beta \in \mathbb{R}\}.$$

Notice that for $\alpha = 1$, $\beta = 0$, $1v + 0w = v \in V_0$. Similarly,

$$w = 0 \cdot v + 1 \cdot w \in V_0.$$

One easily shows that V_0 is a vector subspace of V containing v and w. Moreover it is the smallest vector subspace containing both v and w.

Exercise 2.2.6 Prove the last two assertions of the last paragraph.

Exercise 2.2.7 Let $v = (r, s, t) \in \mathbb{R}^3$ and $w = (a, b, c) \in \mathbb{R}^3$ be two nonzero vectors. Show that the smallest vector subspace of $\mathbb{R}^3$ containing v and w is either

(i) a plane containing these points and the origin, or

(ii) a line passing through these points and the origin.

In the latter case, one of them is a scalar multiple of the other: Either $v = \alpha w$ or $w = \beta v$ for some $\alpha, \beta \in \mathbb{R}$.

More generally, we have the following exercise:

Exercise 2.2.8 If $S = \{v_1, \dots, v_k\}$ is a subset of a vector space V, arguing as above, show that

$$V_0 = \{\alpha_1 v_1 + \cdots + \alpha_k v_n \mid \alpha_1, \dots, \alpha_k \in \mathbb{R}\}$$

is the smallest vector subspace containing S, that is, a vector subspace containing S and if W is a vector subspace containing S, then $V_0 \subset W$.

This suggests the following definition:

Definition 2.2.2 Given $\{v_i\}_{i=1}^k$, a *finite linear combination* of v_i is a vector of the form $\sum_{i=1}^k \alpha_i v_i$, with $\alpha_i \in \mathbb{R}$.

Thus Exercise 2.2.8 can be reformulated using this definition as follows: The set of *all* finite linear combinations of $v_1, \dots, v_k$ is the smallest vector subspace containing $v_1, \dots, v_k$.

Now if S is any subset of V, we let $L(S)$ be the set of all finite linear combinations of elements of S. Thus

$$L(S) = \left\{ \sum_{i=1}^k \alpha_i v_i \;\middle|\; k \in \mathbb{N}, v_i \in S, \alpha_i \in \mathbb{R} \right\}.$$

Note here that k, α_i, v_i are all arbitrarily chosen from their respective domains. Then $L(S)$ is the smallest vector subspace of V containing the given set S. $L(S)$ is also called the *linear span* of S and denoted by Span (S).

Example 2.2.1 Let $S = \{v\} \subseteq V$ for some $v \in V$. Then

$$L(S) = \mathbb{R}v := \{\alpha v \mid \alpha \in \mathbb{R}\}.$$

What is $L(S)$ if $S := \{e_1 - e_2, e_1 + e_2\}$, where $e_1 = (1,0)$ and $e_2 = (0,1)$ in $V = \mathbb{R}^2$?

What is $L(S)$ if $S = \{e_1, e_2, e_1 + e_2\}$ in $\mathbb{R}^3$ where $e_1 = (1,0,0)$ and $e_2 = (0,1,0)$?

Definition 2.2.3 We say a vector subspace V_0 of V is *generated* by the subset $S \subset V$ if the smallest vector subspace containing S is V_0, that is, a vector subspace V_0 of V such that

(i) $S \subset V_0$, and

(ii) if W is any vector subspace such that $S \subset W$ then $V_0 \subset W$.

In such a case, we denote V_0 by $< S >$.

Exercise 2.2.9 If $S = \{v_1, \dots, v_k\} \subset V$, then the vector subspace $< S >$ generated by S is precisely $L(S)$.

Remark 2.2.1 Thus for a subset $S \subset V$, we have

$$L(S) = \text{Span}\,(S) =< S > .$$

We shall use these interchangeably.

Let us address another problem. Can it happen that $L(S') = L(S)$ for subsets $S' \subset S \subset V$? Let us again look at the simplest case. Let $S' = \{v\}$ and $S = \{v, w\}$. The question, therefore, is: When is $L(\{v\}) = L(\{v, w\})$? If equality holds, then, $w \in L(\{v\})$. But we know that any element of this latter set is of the form αv for some $\alpha \in \mathbb{R}$. Hence we see that $w = \alpha v$ for some $\alpha \in \mathbb{R}$. Conversely, if $w = \alpha v$, then $L(\{v\}) = L(\{v, w\})$. In the same vein, we can solve the following exercise:

Exercise 2.2.10 Let v and $\{v_i\}_{i=1}^n$, be vectors in a vector space V. Let $S' = \{v_i\}_{i=1}^n$ and $S = \{v\} \cup S'$. Then $L(S') = L(S)$ if and only if there exist scalars $\alpha_i \in \mathbb{R}$, $1 \le i \le n$, such that $v = \sum_{i=1}^n \alpha_i v_i$.

In particular, we see that $v \in L(\{v_1, \dots, v_k\})$ if and only if

$$L(\{v_1, \dots, v_k\}) = L(\{v, v_1, \dots, v_k\}).$$

Exercise 2.2.10 motivates the following definition:

Definition 2.2.4 Let v and $\{v_i\}_{i=1}^k$ be vectors in V. $v \in V$ is *linearly dependent* on $v_1, \dots, v_k$ if and only if there exist $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ such that $v = \sum_{i=1}^k \alpha_i v_i$. We express this by saying that v is a *linear combination* of $v_1, \dots, v_k$.

We want to look at this from a slightly different point of view. Suppose we are given that v is linearly dependent on $v_1, \dots, v_k$. Let $\{y_1, \dots, y_{k+1}\}$ be a different labelling of $v, v_1, \dots, v_k$. We want to say $\{y_1, \dots, y_{k+1}\}$ is a linearly dependent set. What do we mean by this? We mean that there exists one element among $y_1, \dots, y_{k+1}$, say y_j, which is a linear combination of $y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_{k+1}$. That is,

$$y_j = \sum_{\substack{j=1 \\ j \ne i}}^{k+1} \alpha_i y_i .$$

This suggests the following definition:

Definition 2.2.5 We say $\{v_1, \dots, v_n\}$ are *linearly dependent* if there exists α_i, $1 \le i \le n$, *not* all zero such that $\sum_{i=1}^n \alpha_i v_i = 0$.

Exercise 2.2.11 Definition 2.2.5 is equivalent to the following one (with which we started): $\{x_1, \ldots, x_n\}$ is linearly dependent if and only if there exists an x_j which is a linear combination of x_i, for $i \neq j$, that is, which is a linear combination of the other elements. For, if $\alpha_j \neq 0$, then

$$\alpha_j x_j = -\sum_{i \neq j} \alpha_i x_i$$

so that

$$x_j = -\alpha_j^{-1} \sum_{i \neq j} \alpha_i x_i.$$

What is the geometric meaning underlying linear dependence of vectors? Let us look at $\mathbb{R}^2$ and $\mathbb{R}^3$ to answer this question. If $\{v, w\}$ is a linearly dependent set in $\mathbb{R}^2$, then one of them is a linear combination of the other — here it simply means that one is a linear multiple of the other, say, $w = \alpha v$. Thus w lies on the line joining 0 and v. (Do you see the genesis of the word "linear"?) If $\{u, v, w\}$ is a linearly dependent set in $\mathbb{R}^3$, then, say w is a linear combination of u and v: $w = \alpha u + \beta v$. I claim that this means that w lies on any plane containing the points u, v and the origin. Now any plane passing through the origin is given by an equation of the form

$$ax + by + cz = 0, \qquad (a, b, c) \neq (0, 0, 0). \tag{2.2.1}$$

Let $u = (x_1, y_1, z_1)$ and $v = (x_2, y_2, z_2)$. They lie on the plane given by Equation (2.2.1) if and only if $ax_i + by_i + cz_i = 0$ for $i = 1, 2$. But this is a homogeneous system of two equations in three unknowns a, b and c. Hence by Theorem 1.2.1, it has a non-trivial solution which we again denote by (a, b, c). Now, if (x_i, y_i, z_i) is a solution of the homogeneous Equation (2.2.1), so are $\alpha(x_1, y_1, z_1)$ and $\beta(x_2, y_2, z_2)$. Hence

$$\alpha(x_1, y_1, z_1) + \beta(x_2, y_2, z_2)$$

is a solution of Equation (2.2.1). But this simply says that $\alpha u + \beta v$ lies on the plane given by Equation (2.2.1).

Exercise 2.2.12 Find whether v is a linear combination of v_i's in the following cases.

(1) $v = (a, b)$, $v_1 = (1, 0)$ and $v_2 = (1, 1)$ in $\mathbb{R}^2$.

(2) $v = (0, 0, 1)$, $v_1 = (1, 0, 1)$ and $v_2 = (0, 1, 1)$.

(3) $v = (x_1, x_2, x_3, 1)$, $x_i \in \mathbb{R}$ arbitrary, $v_1 = (1, 2, 3, 0)$, $v_2 = (2, 3, 1, 0)$ and $v_3 = (3, 2, 1, 0)$.

Can you generalize this?

Exercise 2.2.13 Let $V = \mathcal{P}_n$, be the vector space of polynomials of degree less than or equal to n. Describe the set $L(\{x^2+x+1, x\})$ and $L(\{x^2+1, x\})$. Show that they are the same. Find which of the following polynomials lie in $L(\{x^2 + x + 1, 1\})$:

(1) $\frac{22}{7}x^2 + 5x + \pi$.

(2) $10^2x^2 + 10x + 10$.

(3) $ex^2 + 0 \cdot x + e$, where e is the usual base of the natural logarithm.

Exercise 2.2.14 Let $v \in \mathbb{R}^3$ and $S = \{(1,0,1),(1,2,-1)\}$. Give a geometric description of $L(S)$.

Exercise 2.2.15 If $V = \mathbb{R}^3$ and $S = \{v, w\}$, find the set of linear equations which define the geometric object $L(S)$.

Exercise 2.2.16 Let $V = \mathbb{R}^n$ and $S = \{e_1, \dots, e_k\}$, $1 \le k \le n$, where

$$e_j = (0, \dots, 0, 1, 0, \dots, 0), \qquad 1 \le j \le k,$$

(1 at the jth place). What is $L(S)$?

Exercise 2.2.17 If S is a vector subspace of V, what is $L(S)$?

Exercise 2.2.18 The set of solutions S of $ax + by + cz = 0$ for $(a, b, c) \neq 0$ is a vector subspace of $\mathbb{R}^3$. S is the plane through the origin with normal (a, b, c).

Definition 2.2.6 If A and B are nonempty subsets of a vector space V, we denote by $A + B$ the subset $\{a + b \mid a \in A, b \in B\}$.

Exercise 2.2.19 Let the notation be as in Definition 2.2.6. Show that $L(\{v, w\}) = \mathbb{R}v + \mathbb{R}w$. More generally, show that $L(S) = \mathbb{R}v_1 + \cdots + \mathbb{R}v_k$ if $S = \{v_1, \dots, v_k\}$.

Exercise 2.2.20 Given W_1, W_2 vector subspaces of V, does there exist any smallest vector subspace W_3 containing W_1 and W_2?

Exercise 2.2.21 Let W be a vector subspace of V. What is $w + W$ if $w \in W$? What is $W + W$? Is it true that $w + W = W$ if and only if $w \in W$?

Exercise 2.2.22 If W_1, W_2 are vector subspaces of a vector space V, then $W_1 + W_2$ is a vector subspace of V. What is $W_1 + W_2$ if $W_1 = W_2$? More generally, what is $W_1 + W_2$ if $W_1 \subseteq W_2$?

Exercise 2.2.23 Let W_i, $1 \le i \le 2$, be vector subspaces of a vector space V. When is $W_1 \cup W_2$ a subspace of V?

Exercise 2.2.24 Let V be the set of Cauchy sequences in $\mathbb{R}$. Let W be the set of convergent sequences in $\mathbb{R}$. For $x = (x_n), y = (y_n) \in V$, we define $x + y := (x_n + y_n)$ and $\alpha x = (\alpha x_n)$. Then V is a vector space and W is a vector subspace of V over $\mathbb{R}$. Is W a proper subset of V? (You need analysis to answer this!)

Exercise 2.2.25 What are all the vector subspaces of $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$? What are their geometric descriptions?

Exercise 2.2.26 If W_i, $1 \leq i \leq 2$, are vector subspaces of a vector space V, then $W_1 \cap W_2$ is a vector subspace.

Exercise 2.2.27 If $\{W_i\}_{i \in I}$ is a family of vector subspaces of V indexed by a set I, then $\cap_{i \in I} W_i$ is a vector subspace.

Exercise 2.2.28 If S is an arbitrary subset of a vector space V, then $L(S) = < S > =$ Span(S) is the intersection of all vector subspaces containing S.

A **meta exercise** Can you identify the arguments in the foregoing places where our themes appeared?

2.3 Basis and Dimension of a Vector Space

The beginning of this section is a repetition of what we have seen in Section 2.2. These definitions are repeated here since they introduce two of the most important concepts in linear algebra.

Definition 2.3.1 A vector $v \in V$ is said to be a *linear combination* of vectors $v_1, \ldots, v_k$ if there exists $\alpha_i \in \mathbb{R}$ such that $v = \sum_{i=1}^{k} \alpha_i v_i$.

Definition 2.3.2 A set $S = \{v_1, \ldots, v_n\}$ in a vector space V is said to be *linearly dependent* if there exists $\alpha_i \in \mathbb{R}$, $1 \leq i \leq n$, not all α_i's zero, such that $\sum_{i=0}^{n} \alpha_i v_i = 0$.

S is said to be *linearly independent* if it is not linearly dependent. In other words, if $\sum_{i=1}^{n} \alpha_i v_i = 0$ then $\alpha_i = 0$ for all $1 \leq i \leq n$. (Can you convince yourself of this? Most often this is the formulation which is used.)

Exercise 2.3.1 If $\{v_1, \ldots, v_n\}$ is linearly dependent, then there exists j such that $v_j = \sum_{i \neq j} \alpha_i v_i$, for some $\alpha_i \in \mathbb{R}$.

Exercise 2.3.2 If $0 \in S$ then S is linearly dependent.

Exercise 2.3.3 $\{v, w\} \subseteq V$ is linearly dependent if and only if one is a scalar multiple of the other.

Exercise 2.3.4 The vectors (a, b) and (c, d) in $\mathbb{R}^2$ are linearly independent if and only if $ad - bc \neq 0$.

Exercise 2.3.5 Show that if v and w are linearly independent vectors in V, then so are $v + w$ and $v - w$.

Exercise 2.3.6 Let S_1 be a linearly dependent subset of a vector space V and S_2 be such that $S_2 \subseteq S_1$. Then prove that S_2 is linearly dependent. State and prove a similar property for linear independence.

Theorem 2.3.1 *Let V be a vector space. Then $\{v_1, \ldots, v_n\}$ is linearly dependent if and only if one of the v_i's is a linear combination of the other v_j's.*

Proof Since $\{v_1, \ldots, v_n\}$ is linearly dependent there exists $\alpha_i \in \mathbb{R}$, not all zero, such that $\sum_{i=1}^{n} \alpha_i v_i = 0$. Suppose $\alpha_i \neq 0$ for some i. Then $\alpha_1 v_1 + \cdots + \alpha_i v_i + \cdots + \alpha_n v_n = 0$. Hence $v_i = -\alpha_i^{-1}(\sum_{j \neq i} \alpha_j v_j)$ and therefore $v_i = \sum_{j \neq i} \beta_j v_j$ where $\beta_j = -\alpha_i^{-1}\alpha_j \in \mathbb{R}$. Hence v_i is a linear combination of the other v_j's.

We now prove the converse. That is, if for some i, v_i is expressed as a linear combination of v_j, $j \neq i$, then $\{v_1, \ldots, v_n\}$ is a linearly dependent set. Suppose $v_i = \sum_{j \neq i} \alpha_j v_j$. Then we have

$$\alpha_1 v_1 + \cdots + (-1)v_i + \alpha_{i+1} v_{i+1} + \cdots + \alpha_n v_n = 0.$$

That is, there exists $\alpha_1, \ldots, \alpha_n$, with $\alpha_i = -1 \neq 0$ such that $\sum_{i=1}^{n} \alpha_i v_i = 0$. Hence $\{v_1, \ldots, v_n\}$ is linearly dependent.

□

Remark 2.3.1 The above theorem does *not* say the following. If $\{v_i\}_{i=1}^{n}$ is linearly dependent, then *any* v_i is a linear combination of other v_j's, $j \neq i$. For instance, let us consider $V = \mathbb{R}^2$ and $e_1 := (1, 0)$ and $e_2 := (0, 1)$. Then $\{e_1, e_2, 2e_1\}$ is obviously a linearly dependent set as

$$-2e_1 + 0e_2 + 2e_1 = 0.$$

But e_2 is not a linear combination of $\{e_1, 2e_1\}$. (Prove this. Or, see Exercise 2.2.12 (3).)

Definition 2.3.3 A set $B = \{e_1, \ldots, e_n\}$ in a vector space V is said to be a *basis* of V if every vector $v \in V$ can be expressed uniquely as $v = \sum_{i=1}^{n} \alpha_i e_i$, where $\alpha_i \in \mathbb{R}$. That is, if $v = \sum_{i=1}^{n} \alpha_i e_i = \sum_{i=1}^{n} \beta_i e_i$ then $\alpha_i = \beta_i$ for $1 \leq i \leq n$.

We now look at some elementary examples. We shall have a lot more examples in the form of exercises later.

Example 2.3.1 Let $V := \mathbb{R}^n$. If $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, we call x_j, th jth coordinate of x. Let $e_i := (0, \dots, 0, 1, 0, \dots, 0)$ be the vector whose jt coordinate is zero unless $j = i$ in which case it is 1. It is easy to show tha $\{e_i \mid 1 \le i \le n\}$ is a basis of V. This is called the *standard basis* of $\mathbb{R}^n$ In the sequel, when we write $e_i \in \mathbb{R}^n$ it refers to the basis vector as abov

Example 2.3.2 Let $V := M(n, \mathbb{R})$. Let E_{ij} be the element of V whos (i, j)th entry is one and the rest are zero. In notation, if $E_{ij} = (x_{rs})$, w have

$$x_{rs} = \begin{cases} 0, & \text{if } r \neq i \text{ and } s \neq j \\ 1, & \text{if } r = i \text{ and } s = j. \end{cases}$$

Then $\{E_{ij} \mid 1 \le i, j \le n\}$ is a basis for $M(n, \mathbb{R})$.

Example 2.3.3 Let V be the vector space of polynomials in the variable X of degree less than or equal to n, $n \in \mathbb{N}$. Then the set $\{X^k \mid 0 \le k \le n$ is a basis of V.

Example 2.3.4 Let $V := \mathbb{R}^2$. Let

$$\begin{aligned} v_1 &:= e_1 + e_2 = (1, 1), \quad \text{and} \\ v_2 &:= e_1 - e_2 = (1, -1). \end{aligned}$$

We claim that $\{v_1, v_2\}$ is a basis of $\mathbb{R}^2$. We are supposed to establish th following:

(i) any vector $v = (x, y)$ is a linear combination of v_i, and

(ii) this linear combination is unique.

To prove (i), let us assume that $v = (x, y) = \alpha v_1 + \beta v_2$. Then we have

$$(x, y) = \alpha(1, 1) + \beta(1, -1) = (\alpha, \alpha) + (\beta, -\beta) = (\alpha + \beta, \alpha - \beta).$$

Consequently, we see that $x = \alpha + \beta$ and $y = \alpha - \beta$. Hence $\alpha = (x + y)/$ and $\beta = (x - y)/2$. Thus, we have

$$(x, y) = \frac{x + y}{2} v_1 + \frac{x - y}{2} v_2.$$

The above argument also shows the uniqueness of this expression.

Exercise 2.3.7 Show that $\{(1, 2), (4, 3)\}$ is a basis of $\mathbb{R}^2$.

Exercise 2.3.8 When is $\{v, w\}$ a basis of $\mathbb{R}^2$?

Exercise 2.3.9 Show that $\{X, 3X^2, 5 + X\}$ is a basis of P_2. What abou $\{2X, X^2 - 3X, 2X^2\}$?

Exercise 2.3.10 Show that $\{1, (X-a), (X-a)^2, \ldots, (X-a)^n\}$ is a basis of $\mathcal{P}_n$ for all $a \in \mathbb{R}^n$. *Hints (for three different proofs):* (i) Expand $X^k = ((X-a)+a)^k$ by binomial theorem. (ii) Use induction to show that X^k is a linear combination of $(X-a)^j$ for $0 \le j \le k$. (iii) Use Taylor expansion of the function $f(X) = X^n$ at the point $X = a$. This allows you to directly exhibit the coefficients in the linear combination.

Remark 2.3.2 In view of Example 2.3.1, Example 2.3.4 and Exercise 2.3.8, it should be clear that the basis is not unique for a vector space.

Exercise 2.3.11 Fix a basis $\{v_1, \ldots, v_n\}$ of V. Let $v = \sum_{i=1}^n x_i v_i$ and $w = \sum_{i=1}^n y_i v_i$. Show that $v+w = \sum_{i=1}^n (x_i + y_i) v_i$ and $\alpha v = \sum_{i=1}^n (\alpha x_i) \cdot v_i$ are the unique expressions of $v + w$ and αv in terms of the basic vectors v_i. While writing a formal proof, see which of the properties in Definition 2.1.1 of a vector space are used.

Lemma 2.3.2 *Any basis $\{e_i\}$ of V is a linearly independent set.*

Proof If a basis $\{e_i\}$ is linearly dependent, then there exist scalars $a_i \in \mathbb{R}$, not all zero, such that $\sum \alpha_i e_i = 0$. However, we have $0 = 0 \cdot e_1 + \cdots + 0 \cdot e_n$. Since $\{e_i\}$ is a basis, by uniqueness of coefficients, we deduce $\alpha_i = 0$ for all i. This contradicts our assumption on α_i's.

□

Theorem 2.3.3 *If a vector space V has a basis of n elements then any set of $n+1$ vectors is linearly dependent.*

Proof Let $\{e_1, \ldots, e_n\}$ be a basis of V and let $\{v_1, \ldots, v_{n+1}\}$ be any set of $n+1$ vectors. Since $\{e_1, \ldots, e_n\}$ is a basis of V there exists $\alpha_{ij} \in \mathbb{R}$ such that $v_i = \sum_{j=1}^n \alpha_{ij} e_j$, $1 \le i \le n+1$. To show that $\{v_i\}_{i=1}^{n+1}$ is linearly dependent, we need to find $\beta_i \in \mathbb{R}$, not all zero, such that

$$\sum_{i=1}^{n+1} \beta_i v_i = 0. \tag{2.3.1}$$

That is, $\sum_{i=1}^{n+1} \beta_i (\sum_{j=1}^n \alpha_{ij} e_j) = 0$. This means that we need to find $\{\beta_i\}_{i=1}^{n+1}$ such that

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{n+1} \beta_i \alpha_{ij} \right) e_j = 0, \tag{2.3.2}$$

is the zero vector. Since $\{e_i\}_{i=1}^n$ is a basis, $0 = 0e_1 + \cdots + 0e_n$ is the unique representation of 0. Hence equating the coefficients of e_i, we get $\sum_{i=1}^{n+1} \beta_i \alpha_{ij} = 0$, $1 \le j \le n$. This is a system of n equations in $n+1$ unknowns β_i. By Theorem 1.2.1, there exists a non-trivial solution, which we denote by $\{\beta_1, \dots, \beta_{n+1}\}$. Thus there exist scalars β_i, $1 \leq i \leq n+1$ such that Equation (2.3.2) is satisfied. Consequently, $\{v_1, \dots, v_{n+1}\}$ is linearly dependent.

□

Corollary 2.3.4 *Any basis $\{e_i\}_{i=1}^n$ of V is a maximal linearly independent set, that is, if S is a subset of V and $\{e_i\}$ is a proper subset of S, then S is linearly dependent.*

Proof Let $v \in S \setminus \{e_i\}_{i=1}^n$. By Theorem 2.3.3 $\{e_1, \dots, e_n, v\}$ is linearly dependent. Thus there exist real numbers $\alpha_i \in \mathbb{R}$, $1 \leq i \leq n+1$, at least one of them non-zero such that $\alpha_1 e_1 + \cdots + \alpha_n e_n + \alpha_{n+1} v = 0$. Since $\{v, e_1, \dots, e_n\}$ is linearly dependent and is a subset of S, S is linearly dependent by Exercise 2.3.6.

□

Definition 2.3.4 A vector space V is *finite dimensional* if there exists a finite subset S of V such that $L(S) = V$. That is, V is finite dimensional if and only if there exists a finite set $\{v_i\}_{i=1}^n$ such that any $v \in V$ is a linear combination of v_i's.

Exercise 2.3.12 If a vector space V has a basis with a finite number of elements then it is finite dimensional.

Remark 2.3.3 The converse of Exercise 2.3.12 is true. See Theorem 2.3.8 and its corollary.

Theorem 2.3.5 *Let V be a vector space. Assume V has a basis consisting of n elements. Then any linearly independent set of n vectors in V is a basis of V.*

Proof Let $\{e_1, \dots, e_n\}$ be a linearly independent set and let $v \in V$. Then $\{v, e_1, \dots, e_n\}$ is a set of $n+1$ vectors and hence linearly dependent by Theorem 2.3.3. By definition, there exist $\alpha_i \in \mathbb{R}$, not all zero, such that $\alpha_0 v + \alpha_1 e_1 + \cdots + \alpha_n e_n = 0$. If $\alpha_0 = 0$, then by assumption there exists a k, $1 \leq k \leq n$, such that $\alpha_k \neq 0$. Since $\alpha_0 v + \sum_{i=1}^n \alpha_i e_i = 0$ and $\alpha_0 = 0$, we see that $\alpha_1 e_1 + \cdots + \alpha_n e_n = 0$. But then $\{e_i\}_{i=1}^n$ is linearly independent so that $\alpha_i = 0$ for $1 \leq i \leq n$. This contradicts our observation that $\alpha_k \neq 0$ for some $1 \leq k \leq n$. Therefore we are forced to conclude that $\alpha_0 \neq 0$. Hence $v = -\alpha_0^{-1}(\alpha_1 e_1 + \cdots + \alpha_n e_n)$. Moreover, this expression is unique. For, if $v = \sum_{i=1}^n \alpha_i e_i = \sum_{i=1}^n \beta_i e_i$, then $\sum_{i=1}^n (\alpha_i - \beta_i) e_i = 0$. Since $\{e_i\}_{i=1}^n$ is linearly independent, we have $(\alpha_i - \beta_i) = 0$ for all i or $\alpha_i = \beta_i$ for all i. Therefore $\{e_1, \dots, e_n\}$ is a basis of V.

□

Theorem 2.3.6 *Let V be a finite dimensional vector space. Then any two bases of V have the same number of elements.*

Proof 1. Let $\{e_1, \ldots, e_n\}$ and $\{f_1, \ldots, f_m\}$ be two bases of V. If $m = n$ we have nothing to prove. Suppose $m > n$. Then $m \geq n + 1$. Since $\{e_1, \ldots, e_n\}$ is a basis of V, by Theorem 2.3.3, $\{f_1, \ldots, f_m\}$ is linearly dependent. Therefore there exists $\alpha_i \in \mathbb{R}$, $1 \leq i \leq m$, not all zero, such that $\alpha_1 f_1 + \cdots + \alpha_m f_m = 0$. But we also have $0f_1 + \cdots + 0f_m = 0$. Since $\{f_1, \ldots, f_m\}$ is a basis, 0 is written uniquely as a linear combination of f_i's. Therefore $\alpha_i = 0$ for all $1 \leq i \leq m$, a contradiction. Hence we conclude that $m \leq n$. Similarly, we prove $n \leq m$. So $m = n$.

□

Proof 2. Let $\{v_i\}_{i=1}^m$ and $\{w_j\}_{j=1}^n$ be bases of V. We must show that $m = n$.

Since $\{v_i\}$ is a basis and $w_j \in V$ there is a unique expression (that is, a linear combination)

$$w_j = \sum_{i=1}^{m} \alpha_{ji} v_i. \tag{2.3.3}$$

Arguing similarly, we have

$$v_i = \sum_{r=1}^{n} \beta_{ir} w_r. \tag{2.3.4}$$

Using Equation (2.3.4) in Equation (2.3.3), we get

$$w_j = \sum_{i=1}^{m} \alpha_{ji} v_i = \sum_{i=1}^{m} \alpha_{ji} \left(\sum_{r=1}^{n} \beta_{ir} w_r \right) = \sum_{r=1}^{n} \left(\sum_{i=1}^{m} \alpha_{ji} \beta_{ir} \right) w_r. \tag{2.3.5}$$

Similarly using Equation (2.3.3) in Equation (2.3.4), we get

$$v_i = \sum_{j=1}^{m} \left(\sum_{r=1}^{n} \beta_{ir} \alpha_{rj} \right) v_j. \tag{2.3.6}$$

We already have $w_j = 0w_1 + \cdots + 0w_{j-1} + 1 \cdot w_j + 0w_{j+1} + \cdots + 0w_n$. Since $\{w_j\}$ is a basis, by uniqueness of the expression, we infer from Equation (2.3.5) that

$$\sum_i \alpha_{ji} \beta_{ir} = \begin{cases} 0, & \text{if } r \neq j \\ 1, & \text{if } r = j. \end{cases} \tag{2.3.7}$$

By similar reasoning, we infer from Equation (2.3.6) that

$$\sum_r \beta_{ir} \alpha_{rj} = \begin{cases} 0, & \text{if } j \neq i \\ 1, & \text{if } j = i. \end{cases} \tag{2.3.8}$$

In particular, from Equation (2.3.7), it follows that $\sum_{i=1}^{m} \alpha_{ji}\beta_{ij} = 1$ for all $1 \leq j \leq n$. Summing it over j, we get

$$\sum_{j=1}^{n}\left(\sum_{i=1}^{m} \alpha_{ji}\beta_{ij}\right) = n. \tag{2.3.9}$$

Similarly, we get from Equation (2.3.8)

$$\sum_{i=1}^{m}\left(\sum_{r=1}^{n} \beta_{ir}\alpha_{ri}\right) = m. \tag{2.3.10}$$

But, obviously, the left sides of Equation (2.3.9) and Equation (2.3.10) are the same. Hence $m = n$.

□

Definition 2.3.5 We say a vector space V is of *dimension* n if it has a basis consisting of n elements. That is, if there exists n vectors $e_1, \ldots, e_n$ in V such that any $v \in V$ can be written uniquely as $v = \sum_{i=1}^{n} \alpha_i v_i$, $\alpha_i \in \mathbb{R}$. Note that this is well-defined in view of Theorem 2.3.6. We then write $\dim V = n$.

Exercise 2.3.13 Determine which of the examples (and exercises) in Section 2.1 are finite dimensional and find their dimensions.

Exercise 2.3.14 Let W be a vector subspace V with $\dim W = \dim V$. Then $W = V$.

Definition 2.3.6 Let V be a finite dimensional vector space over $\mathbb{R}$ and $\{e_1, \ldots, e_n\}$ a basis of V. Given any $v \in V$, there exist unique constants $\alpha_i \in \mathbb{R}$ such that $v = \sum_{i=1}^{n} \alpha_i e_i$. We denote $x_i(v) := \alpha_i$. The functions $x_i \colon V \to \mathbb{R}$ are called the *coordinate functions* with respect to the basis $\{e_i\}_{i=1}^{n}$. α_k or $x_k(v)$ is called the kth coordinate of v with respect to the basis $\{e_i\}_{i=1}^{n}$.

This is the significance of the basis. Given a basis we get a coordinate system on V. We thus require n coordinates (no more, no less) to determine an arbitrary vector $v \in V$ uniquely.

In other words, a vector space V is n-dimensional if "it requires n coordinates" to locate the points uniquely. In physics, this is said as "n degrees of freedom for a particle $v \in V$ to move". This intuitive way of thinking of dimension is quite useful geometrically. We shall put this idea into use in the next couple of examples and then give you problems for practice.

Example 2.3.5 Let $W = \{(x, y, z) \in \mathbb{R}^3 \mid y = z\}$. We know that W is the vector subspace of $\mathbb{R}^3$. We show that its dimension is two by exhibiting a basis. If $w \in W$, to "locate" w, we need only 2 "coordinates". For example, if we know its x and y coordinates, then we know its third coordinate z. Thus dimension of W must be 2. To see this, if $(x, y, z) \in W$ then $z = y$ so that $(x, y, y) = x(1,0,0)+y(0,1,0)+y(0,0,1) = x(1,0,0)+y(0,1,1)$. Thus, we see that $L(\{(1,0,0),(0,1,1)\}) = W$. We claim that $(1,0,0)$ and $(0,1,1)$ are linearly independent. This is clear, since neither is a scalar multiple of the other. See Exercise 2.3.3. Thus, $\{(1,0,0),(0,1,1)\}$ is a basis of W so that $\dim(W) = 2$.

Example 2.3.6 Let $\mathbf{S}_n$ denote the set of symmetric matrices in $M(n, \mathbb{R})$. Then $\mathbf{S}_n$ is a vector subspace of $M(n, \mathbb{R})$. What is its dimension? We shall work out the case $n = 2$ and leave the general case as an exercise.

Let

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathbf{S}_2.$$

In general, to "locate" any matrix

$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \in M_{2\times 2},$$

we need four coordinates $(x_{11}, x_{12}, x_{21}, x_{22})$. But, however, if A is symmetric, if we know a_{12} then we know a_{21}. Thus to "locate" A, we need only three coordinates a_{11}, a_{12}, a_{22}. Therefore we expect that the dimension of $\mathbf{S}_2$ is three. To see this, recall the basis $\{E_{ij}\}$ of $M(n, \mathbb{R})$ where E_{ij} is a matrix whose ijth entry is 1 and the all other entries are zero. If $A \in \mathbf{S}_n$, then

$$\begin{aligned} A &= a_{11}E_{11} + a_{12}E_{12} + a_{21}E_{21} + a_{22}E_{22} \\ &= a_{11}E_{11} + a_{12}(E_{12} + E_{21}) + a_{22}E_{22}. \end{aligned}$$

Thus $\{E_{11}, E_{12} + E_{21}, E_{22}\}$ generates $\mathbf{S}_2$. It is easily seen that it is linearly independent so that it is a basis.

Exercise 2.3.15 Let $W = \{(x_1, \ldots, x_n) \mid x_1 + x_2 + \cdots + x_n = 0\} \subset \mathbb{R}^n$. Find a basis and dimension of W.

Exercise 2.3.16 If $V = \mathbb{R}^n$ and $W = \{(x_1, \ldots, x_n) \mid x_1 = x_n\}$, find a basis and dimension of W.

Exercise 2.3.17 If $V = \mathbb{R}^n$ and $W = \{(x_1, \ldots, x_n) \mid x_k = 0 \text{ if k is even}\}$, find a basis and dimension of W.

Exercise 2.3.18 Let $W = \{(x_1, \ldots, x_n) \mid x_k\text{'s are all equal for } k \text{ even}\}$. Find a basis and dimension of W.

Exercise 2.3.19 Find a basis for $\mathbf{S}_n$ of symmetric matrices.

Exercise 2.3.20 Let $V = M(n,\mathbb{R})$ and let $\mathbf{A}_n$ be the set of all skew-symmetric matrices. Find a basis and dimension of $\mathbf{A}_n$.

Exercise 2.3.21 Let $V = M(n,\mathbb{R})$ and $W = \{X \in V \mid \text{trace } X = 0\}$. Find a basis and dimension of W. (Recall $\operatorname{tr}(X) = \sum_i x_{ii}$ if $X = (x_{ij})$).

Exercise 2.3.22 Find a basis of $\mathbb{C}$, considered as a vector space over $\mathbb{R}$ (see Exercise 2.1.21). *Hint:* Any $z \in \mathbb{C}$ can be written as $z = 1 \cdot x + y \cdot i$, $x, y \in \mathbb{R}$.

Exercise 2.3.23 Can you exhibit a basis of $\mathcal{P}_n$ consisting of elements all of degree n? All of degree $\leq n-1$?

Exercise 2.3.24 The notation is as in Exercise 2.3.20. Given a basis $\{v_i\}_{i=1}^m$ of V and $\{w_j\}_{j=1}^n$ of W, find a basis of $V \oplus W$.

Exercise 2.3.25 The notation is as in Exercise 2.1.19. Given a basis $\{v_i\}_{i=1}^m$ of V, find a basis of $\{\star\} \times V$.

Remark 2.3.4 You should realize that "more the merrier" is not true here! That is, if we adjoin more elements to a basis we may not be able to "locate" any vector v "more precisely" than what we could do with the given basis.

For example, if $V = \mathbb{R}^2$ and $e_1 = (1,0)$ and $e_2 = (0,1)$, then $\{e_1, e_2\}$ is a basis of V. Let us take $S = \{e_1, e_2, e_3 = (1,1)\}$. With respect to $\{e_1, e_2, e_3\}$, we can write

$$\begin{aligned} v = (1,1) &= 1e_1 + 1e_2 + 0e_3 \\ &= \frac{1}{2}e_1 + \frac{1}{2}e_2 + \frac{1}{2}e_3 \\ &= 0e_1 + 0e_2 + 1e_3. \end{aligned}$$

That is, there are lots of ways of giving coordinates to v as $(1,1,0)$ or $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ or $(0,0,1)$ and a host of others!

The next definition is the same as Definition 2.2.3 and nearby items.

Definition 2.3.7 Let S be a non-empty subset of a vector space V. An element $v \in V$ of the form $v = \sum_{i=1}^n \alpha_i v_i$, $v_i \in S$, $\alpha_i \in \mathbb{R}$ is called a *finite linear combination* of elements of S. The set of all such finite linear combinations is called the *linear span* of S denoted by Span $\{S\}$, $L(S)$ or $< S >$ and S is called a *set of generators* of $L(S)$.

Example 2.3.7 Every vector space has a set of generators. Since for a vector space V, $V = L(V)$, V itself is a set of generators of V. Any basis is a generating set.

In fact, a basis is a *minimal* set of generators. Can you make this statement precise and prove it? (*Hint:* See Theorem 2.3.7 below.)

Remark 2.3.5 Let $V = \mathbb{R}^2$. Then,

$$\begin{aligned} S_1 &= \{(1,0),(0,1)\}, \\ S_2 &= \{(1,0),(0,1),(1,1)\}, \text{ and} \\ S_3 &= \{(0,0),(1,0),(-1,0),(0,-1),(1,1)\} \end{aligned}$$

are all generating sets of $\mathbb{R}^2$. That is $\mathbb{R}^2$ is spanned by S_1, S_2 as well as S_3. A number of questions arise here: Which spaces can be spanned by a finite set of elements? Also if a space can be spanned by a finite set of elements what is the smallest number of elements required?

Theorem 2.3.7 *Let V be a finite dimensional vector space. Then the following are equivalent:*

(1) $\{e_1, \dots, e_n\}$ *is a basis of* V.

(2) $\{e_1, \dots, e_n\}$ *is a maximal linearly independent set.*

(3) $\{e_1, \dots, e_n\}$ *is a minimal generating set.*

Proof (1) $\Rightarrow$ (2) Let $\{e_1, \dots, e_n\}$ be a basis. We first show that $\{e_i\}$ is linearly independent. If $\sum \alpha_i e_i = 0$, then the zero vector is expressed as a linear combination of e_i. The zero vector already has the expression $0 = 0e_1 + \cdots + 0e_n$. Hence by the uniqueness of the expression, $\alpha_i = 0$ for all i. Therefore $\{e_1, \dots, e_n\}$ is a linearly independent set. We now claim that it is maximal linearly independent. That is, if we add even one more element x to $\{e_i\}_{i=1}^n$, then the resulting set $\{x, e_1, \dots e_n\}$ is linearly dependent. This follows from Theorem 2.3.3. Or, more directly, since $\{e_i\}$ is a basis, we can write x as a linear combination of e_i's: $x = \sum_{i=1}^n \alpha_i e_i$. But then $\alpha_0 x - \alpha_1 e_1 - \cdots - \alpha_n e_n = 0$ with $\alpha_0 = 1$. Thus the set $\{e_1, \dots, e_n, x\}$ is linearly dependent.

(2) $\Rightarrow$ (3) To this end, we need to show two things:

(a) e_i's generate V, that is, any v is a linear combination of e_i's.

(b) No proper subset of $\{e_i\}$ has the property (a), that is, we cannot generate all of V by any proper subset of $\{e_i\}_{i=1}^n$.

To prove (a), let $v \in V$ be given. Since $\{e_1, \dots, e_n\}$ is a maximal linearly independent set, $\{v, e_1, \dots, e_n\}$ is linearly dependent. Hence there exist scalars α_i, $1 \le i \le n$ with at least one of the α_i's different from zero such that $\alpha_0 v + \alpha_1 e_1 + \cdots + \alpha_n e_n = 0$. We want to write $v = -\alpha_0^{-1} \sum_{i=1}^n \alpha_i e_i$.

We need to show that $\alpha_0 \neq 0$. If $\alpha_0 = 0$, then, $\alpha_1 e_1 + \cdots + \alpha_n e_n =$ and not all $\alpha_i = 0$, $1 \leq i \leq n$. Therefore $\{e_i\}_{i=1}^n$ is linearly dependent a contradiction. Hence $v = \sum \beta_i e_i$, with $\beta_i = -\alpha_0^{-1}\alpha_i$ and we see tha $\{e_1, \ldots, e_n\}$ is a generating set for V. This proves (a).

It is a minimal generating set. For, if

$$\{e_1, \ldots, \hat{e}_i, \ldots, e_n\} := \{e_1, \ldots, e_n\} \setminus \{e_i\}$$

generates V, then e_i is a linear combination of e_j's, $1 \leq j \neq i \leq n$. Bu then $\{e_i\}_{i=1}^n$ is linearly dependent, a contradiction.

(3) $\Rightarrow$ (1) Since $\{e_i\}_{i=1}^n$ is a generating set, any $v \in V$ is written as linear combination of e_i's: $v = \sum \alpha_i e_i$. We need only to show that thi expression is unique. If not, let $v = \sum \beta_i e_i$, be a different expression. Ther there exists i such that $\alpha_i \neq \beta_i$. Hence

$$v - v = \sum_{j \neq i} (\alpha_j - \beta_j) e_j + (\alpha_i - \beta_i) e_i = 0.$$

'e conclude that $e_i = -(\alpha_i - \beta_i)^{-1} \sum_{j \neq i} (\alpha_j - \beta_j) e_j$. We now make the llowing:

Claim: $\{e_1, \ldots, \hat{e}_i, \ldots, e_n\}$ is a generating set of V.

This claim contradicts the minimality of $\{e_1, \ldots, e_n\}$.

Let $w \in V$ be arbitrary. Since $\{e_i\}_{i=1}^n$ is a generating set, we can write

$$\begin{aligned}
w &= \sum \gamma_k e_k \\
&= \sum_{k \neq i} \gamma_k e_k + \gamma_i e_i \\
&= \sum_{k \neq i} \gamma_k e_k - \gamma_i (\alpha_i - \beta_i)^{-1} \left(\sum_{j \neq i} (\alpha_j - \beta_j) e_j \right) \\
&= \sum_{k \neq i} [\gamma_k + \gamma_i (\beta_i - \alpha_i)^{-1} (\alpha_k - \beta_k)] e_k \\
&= \sum_{k \neq i} \xi_k e_k.
\end{aligned}$$

That is, w is in the span of $\{e_1, \ldots, \hat{e}_i, \ldots, e_n\}$. In other words,

$$\{e_1, \ldots, \hat{e}_i, \ldots, e_n\}$$

is a generating set.

$\square$

Exercise 2.3.26 Find the dimension of the set of solutions of

(1) $x + 4z + t = 0$, $x + y + 2z - 4t = 0$.

(2) $x + 2y = 0$, $y - z = 0$, $x + y + z = 0$.

Exercise 2.3.27 In $\mathcal{P}_n$, exhibit a basis consisting of elements each of which has degree n.

Exercise 2.3.28 Find a basis for the vector space $M_{m\times n}(\mathbb{R})$ consisting of the set of $m \times n$ matrices with real entries.

Exercise 2.3.29 Show that the following *elementary operations* on a subset $\{v_1, \dots, v_k\}$ of a vector space V "preserve" linear independence or dependence of the family:

(1) Interchanging two of the vectors.

(2) Multiplying a vector by a non-zero scalar.

(3) Replacing any v_i by $v_i + \alpha v_j$ for any scalar α and any $j \neq i$.

What is expected of you is this: If we have $\{v_1, \dots, v_i, \dots, v_j, \dots, v_n\}$ and use the first elementary operation to get $\{v_1, \dots, v_j, \dots, v_i, \dots, v_n\}$ then the first set is linearly dependent (respectively independent) if and only if the second is so. Similar remarks apply to others.

Theorem 2.3.8 *Let V be a vector space of dimension n and let W be a subspace of V. Any basis $\{w_1, \dots, w_k\}$ of W can be extended to a basis $\{v_1, \dots, v_n\}$ of V such that $v_i = w_i$ for $1 \leq i \leq k$.*

Can you think of a special case of W in Theorem 2.3.8? What does the theorem say when $W = \{0\}$, the trivial subspace?

Proof W is the subspace spanned by $\{w_1, \dots, w_k\}$. Hence $W \subseteq V$. If $W = V$, then we are through. If $W \subset V$, then there exists $v_{k+1} \in V$ such that $v_{k+1} \notin W$, that is, $v_{k+1} \in V \setminus W$.

Claim: $\{w_1, \dots, w_k, v_{k+1}\}$ is linearly independent.

For, if not, let $\sum_{i=1}^{k} \alpha_i w_i + \alpha_{k+1} v_{k+1} = 0$. Then $\alpha_{k+1} = 0$. For, if $\alpha_{k+1} \neq 0$, $v_{k+1} = -\alpha_{k+1}^{-1} \sum_{i=1}^{k} \alpha_i w_i$, and hence it is in W, a contradiction, since by our choice $v_{k+1} \notin W$. Hence the claim.

Now if $W_1 =$ Span $\{w_1, \dots, w_k, v_{k+1}\} = V$, then we are through. If not, there exists $v_{k+2} \in V \setminus W_1$ and $\{w_1, \dots, w_k, v_{k+1}, v_{k+2}\}$ is linearly independent as above. If $W_2 =$ Span $\{w_1, \dots, w_k, v_{k+1}, v_{k+2}\} = V$, then we are through. Otherwise we continue the above process. Since V is finite dimensional this process must end after a finite number of steps. In fact, the

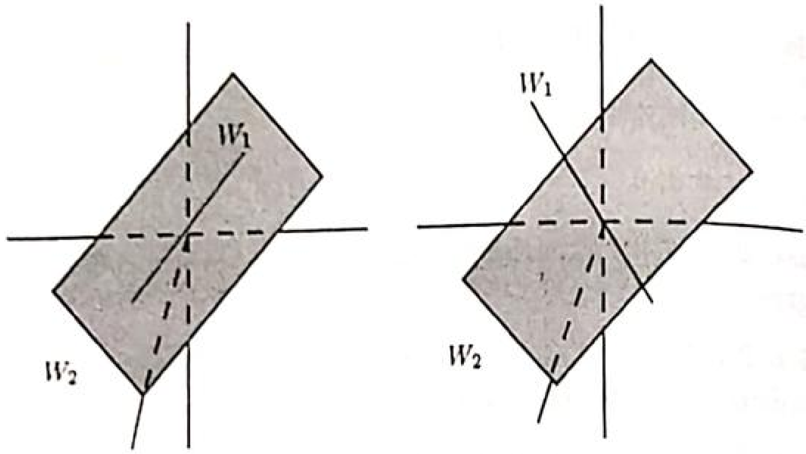


Figure 2.3.1 $\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2)$.

process ends exactly after $r := (\dim V - \dim W)$-number of steps. Do you see why? At the end of r steps, we shall have $\{w_1, \dots, w_k, v_{k+1}, \dots, v_{k+r}\}$, a linearly independent set of $k + r = n$ elements. By Theorem 2.3.5, this set is a basis.

□

Corollary 2.3.9 *Let V be a finite dimensional vector space. Then V has a basis.*

Theorem 2.3.10 *Let V be a vector space and W_1 and W_2 be subspaces of V. Then*

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2). \tag{2.3.11}$$

The geometric meaning underlying this theorem:

If $V = \mathbb{R}^3$, and W_1 and W_2 are, say, one-dimensional vector subspaces of V, then they are lines through the origin.

If $W_1 \neq W_2$, then $W_1 + W_2$ will be the plane containing both the lines and $W_1 \cap W_2 = (0)$ so that the equality,

$$2 = \dim(W_1 + W_2) = (\dim W_1 = 1) + (\dim W_2 = 1) - (\dim(W_1 \cap W_2) = 0),$$

holds.

If $W_1 = W_2$, then by Exercise 2.2.21, $W_1 + W_2 = W_1$, and $W_1 \cap W_2 = W_1$, so that

$$1 = \dim(W_1 + W_2) = (\dim W_1 = 1) + (\dim W_2 = 1) - (\dim W_1 \cap W_2 = 1).$$

We can do a similar analysis, when W_1 is a line and W_2 is a plane. Here there are two cases: $W_1 \subseteq W_2$ or $W_1 \cap W_2 = \{0\}$ (see Figure 2.3.1).

In the first case, $W_1 + W_2 = W_2$ so that

$$2 = \dim(W_1 + W_2) = (\dim W_1 = 1) + (\dim W_2 = 2) - (\dim(W_1 \cap W_2) = 1).$$

In the second case, $W_1 + W_2 = \mathbb{R}^3$ so that

$$3 = \dim(W_1 + W_2) = (\dim W_1 = 1) + (\dim W_2 = 2) - (\dim(W_1 \cap W_2) = 0).$$

We leave it to the reader the case when W_1 and W_2 are both planes through the origin. Here again there are two cases: W_1 and W_2 intersect in a line or coincide. Is it possible that $W_1 \cap W_2 = \{0\}$?

Proof $W_1 \cap W_2 \subseteq W_i$ for $i = 1, 2$. Let $\{u_i\}_{i=1}^k$, be a basis of $W_1 \cap W_2$. Extend this to a basis

$$\{u_1, \dots, u_k, v_1, \dots, v_m\} \text{ of } W_1$$

and a basis

$$\{u_1, \dots, u_k, w_1, \dots, w_n\} \text{ of } W_2.$$

Then

$$\dim W_1 + \dim W_2 - \dim(W_1 \cap W_2) = (m + k) + (n + k) - k = m + n + k.$$

Claim: $\{u_1, \dots, u_k, v_1, \dots, v_m, w_1, \dots, w_n\}$ is a basis of $W_1 + W_2$.

Let $x \in W_1 + W_2$. Then $x = w_1 + w_2$, where $w_1 \in W_1$ and $w_2 \in W_2$. Now $w_1 = \sum_{i=1}^k \alpha_i u_i + \sum_{j=1}^m \beta_j v_j$. Similarly, $w_2 = \sum_{i=1}^k \gamma_i u_i + \sum_{j=1}^n \delta_j w_j$. Hence $x = \sum_{i=1}^k (\alpha_i + \gamma_i) u_i + \sum_{j=1}^m \beta_j v_j + \sum_{j=1}^n \delta_j w_j$. This shows that the set $\{u_1, \dots, u_k, v_1, \dots, v_m, w_1, \dots, w_n\}$ spans $W_1 + W_2$. We now show that it is linearly independent.

Assume that $\sum \alpha_i u_i + \sum \beta_j v_j + \sum \gamma_r w_r = 0$. We need to show that $\alpha_i, \beta_j, \gamma_r$'s are all zero. Now

$$\sum_{i=1}^k \alpha_i u_i + \sum_{j=1}^m \beta_j v_j = -\sum_{=1}^n \gamma_r w_r.$$

The expression on the right side is an element of W_2 and that on the left lies in W_1. Thus $-\sum_{r=1}^n \gamma_r w_r \in W_1 \cap W_2$ and hence we can write

$$-\sum_{r=1}^n \gamma_r w_r = \sum_{j=1}^k \alpha_j u_j$$

so that

$$\sum_{j=1}^k \alpha_j u_j + \sum_{r=1}^n \gamma_r w_r = 0.$$

$\dots, u_k, w_1, \dots, w_r\}$ is a basis of W_2 and hence is a linearly inde-
set. We therefore conclude that $\alpha_j = 0$ and $\gamma_r = 0$ for all j and

articular,

$$\sum_{i=1}^{k} \alpha_i u_i + \sum_{j=1}^{m} \beta_j v_j = -\sum_{r=1}^{n} \gamma_r w_r = 0.$$

e $\{u_1, \dots, u_k, v_1, \dots, v_m\}$ is a basis of W_1, it is linearly independent.
therefore conclude that $\alpha_i = 0$ and $\beta_j = 0$. Thus α_i, β_j and $\gamma_r = 0$ for
i, j, r. This completes the proof.

□

The rest of the section may be omitted in the first reading.

Definition 2.3.8 Let $W_1, \dots, W_k$ be subspaces of a vector space V. Let $W = W_1 + \dots + W_k$. Then we say that W is the *direct sum* of the W_i if for each j, $1 \leq j \leq k$, $W_j \cap (W_1 + \dots + W_{j-1} + W_{j+1} + \dots + W_n) = \{0\}$. We write $W = W_1 \oplus \dots \oplus W_k$.

Remark 2.3.6 As we saw above when $k = 2$, direct sum implies that $W_1 \cap W_2 = \{0\}$. However, when $k > 2$, to say $W = W_1 \oplus \dots \oplus W_k$ implies much more than $W_1 \cap \dots \cap W_k = \{0\}$. The intersection of each W_j with the *sum* of the other W_i's has to be the zero vector.

Exercise 2.3.30 Show that if W_i, $i = 1, 2$ are subspaces of V with

$$\dim W_1 + \dim W_2 > \dim V,$$

then $W_1 \cap W_2 \neq \{0\}$. What can you say if we always have

$$\dim W_1 + \dim W_2 = \dim V?$$

Exercise 2.3.31 Show that $M(n, \mathbb{R}) = \mathbf{S}_n \oplus \mathbf{A}_n$. (Notation as in Exercise 2.1.13 and Exercise 2.1.14). *Hint:* Given $X \in M(n, \mathbb{R})$, what can you say about $(X + X^t)/2$ and $(X - X^t)/2$? X^t stands for the transpose of X whose (i, j)th element is the (j, i)th element of X.

Exercise 2.3.32 Let S be the vector space of sequences. Let W be the vector subspace of constant sequences and $\mathcal{N}$ be the vector subspace of null sequences. Show that $S = W \oplus \mathcal{N}$.

Exercise 2.3.33 Let $V := \mathcal{F}(X, \mathbb{R})$ be the vector space of real valued functions on X. Fix $x_0 \in X$. Let W be the vector subspace of functions vanishing at x_0: $W := \{f \in \mathcal{F}(X, \mathbb{R}) \mid f(x_0) = 0\}$. Is there a vector subspace W' such that $V = W \oplus W'$?

3. Lines and Quotient Spaces

This chapter has three sections. The first section is very important. The second and third sections may be learnt during a second reading of the book.

3.1 Definition of a Line

We have already seen that the one-dimensional vector subspaces in $\mathbb{R}^2$ and $\mathbb{R}^3$ are lines passing through the origin. Recall that any one-dimensional vector subspace is of the form $\mathbb{R}v$, for some *nonzero* v and conversely. Therefore, if a line ℓ in $\mathbb{R}^2$ is given, it must be parallel to one of the lines passing through the origin, that is, to $\mathbb{R}d$ for some nonzero $d \in \mathbb{R}^2$. From our study in Chapter 1, to get a line $\ell \subset \mathbb{R}^2$ (or $\mathbb{R}^3$), we know that we must consider a set of the form $p + \mathbb{R}d$ where p is any point on the given line ℓ. This motivates our definition of a line in any arbitrary vector space V.

Definition 3.1.1 Let V be a vector space and $0 \neq d \in V$. A line passing through $p \in V$ and having *direction* d, is denoted by $\ell(p; d)$ and defined as

$$\ell(p;d) = \{v \in V \mid \text{ there exists } t \in \mathbb{R}, v = p + td\} = p + \mathbb{R}d.$$

In $\mathbb{R}^n$, if $x = (x_1, \ldots, x_n) \in \ell(p; d)$, then, there exists $t \in \mathbb{R}$ such that

$$\begin{aligned}(x_1, \ldots, x_n) &= (p_1, \ldots, p_n) + t(d_1, \ldots, d_n) \text{ for } t \in \mathbb{R} \\ &= (p_1 + td_1, \ldots, p_n + td_n).\end{aligned}$$

Hence $x_i = p_i + td_i$ so that $t = \frac{x_i - p_i}{d_i}$ for $1 \leq i \leq n$. Eliminating t, we get

$$\frac{x_1 - p_1}{d_1} = \frac{x_2 - p_2}{d_2} = \cdots = \frac{x_n - p_n}{d_n}.$$

(This procedure is called the elimination of t.) When $n = 3$, d_1, d_2, d_3 are known as *direction cosines* of the line. This is why d is called the *direction* *vector* for the line $\ell(p;d)$. When $n=2$, the above equations are given by

$$\frac{x-p_1}{d_1}=\frac{y-p_2}{d_2}.$$

Hence $d_2(x-p_1)=d_1(y-p_2)$ or

$$y=\frac{d_2}{d_1}x-d_2p_1+p_2=mx+c,$$

where $m=d_2/d_1$ is the slope.

We encourage the reader to contemplate on the geometric meaning of the propositions in this section before starting on their proofs.

Proposition 3.1.1 *$\ell(p;d)=\ell(q;d)$ if and only if $(q-p)$ is a multiple of d.*

Proof Suppose $\ell(p;d)=\ell(q;d)$. Then $x\in\ell(p;d)$ implies $x\in\ell(q;d)$. Then there exist $s,t\in\mathbb{R}$ such that $x=p+sd=q+td$. Hence $q-p=(s-t)d$, that is, $q-p$ is a multiple of d.

Conversely, suppose $q-p$ is a multiple of d, that is, $q-p=\alpha d$ for some $\alpha\in\mathbb{R}$. Let $v\in\ell(p;d)$. Then, $v=p+td=(q-\alpha d)+td=q+(t-\alpha)d$ for some $t\in\mathbb{R}$. Therefore, $v\in\ell(q;d)$ and so $\ell(p;d)\subseteq\ell(q;d)$. Similarly, $\ell(q;d)\subseteq\ell(p;d)$ and hence $\ell(p;d)=\ell(q;d)$.

□

The proofs of the next two propositions are left as exercises.

Proposition 3.1.2 *$\ell(p;d)=\ell(p;\alpha d)$ for any $\alpha\in\mathbb{R}\setminus\{0\}$.*

□

Proposition 3.1.3 *$\ell(p;d)=\ell(q;d)$ for any $q\in\ell(p;d)$.*

□

Proposition 3.1.4 *Any two distinct points determine a unique line.*

Proof Let ℓ be any line such that $p,q\in\ell$. If d is the direction of ℓ, then $\ell=\ell(p;d)=\ell(q;d)$ by Proposition 3.1.3 above. By Proposition 3.1.1, $p-q=td$ for some $t\in\mathbb{R}$, $t\neq 0$ since $p\neq q$. Therefore,

$$\ell=\ell(p;d)=\ell(q;d)=\ell(p;td)=\ell(p;p-q)=\ell(q;q-p).$$

Thus the only line having p and q on it is $\ell(p;p-q)$ (which is the same as $\ell(q;p-q)$).

□

Notation. For $x, y \in V$, $x \neq y$, we let $\ell(x, y)$ denote the unique line joining x and y. Note that from the proof of Proposition 3.1.4 it follows that $\ell(x, y) = \ell(x; y - x)$. Thus $\ell(x, y) = \{x + t(y - x) \mid t \in \mathbb{R}\}$. Let $V = \mathbb{R}^3$ and $p = (x_1, y_1, z_1)$ and $q = (x_2, y_2, z_2) \in \mathbb{R}^3$. If $r = (x, y, z) \in \ell(p, q)$, then $r = p + tq$ for some $t \in \mathbb{R}$. Equating the components and eliminating t as earlier, we get

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1},$$

the standard equation of a line in $\mathbb{R}^3$, joining the points (x_i, y_i, z_i) for $i = 1, 2$.

Definition 3.1.2 We say that the two lines $\ell(p; d_1)$ and $\ell(q; d_2)$ are *parallel* if $d_1 = \alpha d_2$ for some $\alpha \in \mathbb{R}$. ($\alpha \neq 0$, as $d_i \neq 0$).

Exercise 3.1.1 In $\mathbb{R}^2$, two distinct lines ℓ and ℓ' are parallel if and only if $\ell \cap \ell' = \emptyset$. Is an analogous statement true in any $\mathbb{R}^n$?

Proposition 3.1.5 (Euclid's parallel postulate) *Given ℓ and $q \notin \ell$ there exists a unique line $\ell(q; d)$ such that $\ell(q; d)$ is parallel to ℓ.*

Proof Let d be a direction of ℓ. Then $\ell(q; d)$ is a line passing through q and parallel to ℓ. Suppose there exists another line $\ell(q; d_1)$ such that $\ell(q; d_1)$ is parallel to ℓ. Then $d_1 = \alpha d$ for some $\alpha \in \mathbb{R}$. Hence by Proposition 3.1.2, we find that $\ell(q; d) = \ell(q; d_1)$.

□

Thus our definition of a line in $\mathbb{R}^2$ has the following properties:

(1) Any two distinct points determine a unique line.

(2) Given a line ℓ and a point p not on ℓ, there exists a unique line $\ell(p; d)$ parallel to ℓ.

Therefore $\mathbb{R}^2$ is a model of plane geometry in which the Euclidean parallel postulate is a theorem and not a postulate!

Exercise 3.1.2 Can you define parallel planes in $\mathbb{R}^3$ using the concepts defined so far? Can you generalize your definitions?

3.2 Affine Spaces

Let V be an arbitrary vector space.

Definition 3.2.1 An *affine subspace* is a non-empty set such that for all $x, y \in S$, the line joining x and y also lies in S. This means that if $x, y \in S$ then $tx + (1-t)y \in S$ for all $t \in \mathbb{R}$. Note that S need not be a vector subspace. Note also that $t \in \mathbb{R}$ is arbitrary.

Example 3.2.1 Let W be a vector subspace of V. Let $v \in V$ be fixed. Then the set $S := v + W$ is an affine space. For, let $x, y \in S$ be arbitrary. Then $x = v + w_1$ and $y = v + w_2$ for some $w_i \in W$. The line joining x and y is the set $\ell(x, y) := \{tx + (1-t)y \mid t \in \mathbb{R}\}$. Let $z \in \ell(x, y)$. We need to show that $z \in S$. For this, we must show that z is of the form $v + w$ for some $w \in W$. Since $z \in \ell(x, y)$, $z = tx + (1-t)y$ for some $t \in \mathbb{R}$. We have

$$z = t(v + w_1) + (1-t)(v + w_2) = (t + 1 - t)v + tw_1 + (1-t)w_2 = v + w$$

where $w := tw_1 + (1-t)w_2$. Since W is a vector subspace we know that $w \in W$. Thus z is of the form $v + w$ for some $w \in W$ and hence $z \in S$.

The converse is also true and that is the content of the next theorem.

Theorem 3.2.1 *A non-empty subset S of V is an affine space if and only if it is of the form $v + W$ for some $v \in V$ and a vector subspace W of V.*

Proof One way is Example 3.2.1. To prove the other way, we work backwards. If the result is true and $S = v + W$, note that $v \in S$ as we can write $v = v + 0 \in v + W$. Then, $S = v + W$ implies that $W = S - v$. This suggests the following approach.

Fix $v \in S$. Consider $W := S - v$. We wish to show that W is a vector subspace. Clearly, $0 \in W$, as $0 = v - v \in S - v$. If $w_i \in W$, we have to show that $w_1 + w_2 \in W$. We can write $w_1 = x - v$ and $w_2 = y - v$ for some $x, y \in S$. Now, $w_1 + w_2 = (x - v) + (y - v) = x + y + (-1)v - v$. This will be in S, if we can show that $x + y - v \in S$ for $x, y, v \in S$. (Note that the coefficients add up to 1. This follows from the lemma below). Assuming this for a moment, we have shown that $w_1 + w_2 = z - v$ for some $z \in S$, that is, it lies in W.

To show that if $w \in W$ and $\alpha \in \mathbb{R}$, then $\alpha w \in W$. Let $w = x - v$ for $x \in S$. Then

$$\alpha w = \alpha x - \alpha v = \alpha x - (\alpha - 1)v - v = \alpha x + (1 - \alpha)v - v.$$

Since S is affine and $x, v \in S$ we see that $\alpha x + (1 - \alpha)v \in S$. Thus the displayed equation shows that $\alpha w = z - v$ where

$$z = \alpha x + (1 - \alpha)v \in S.$$

Hence $\alpha w \in W$. We have therefore shown that W is a vector subspace. $\square$

Lemma 3.2.2 *S is an affine subspace if and only if $\sum_{i=1}^n \alpha_i v_i \in S$ for all $v_i \in S$ and $\alpha_i \in \mathbb{R}$ with $\sum_i \alpha_i = 1$.*

Proof Note that for $n = 2$, this is just definition. We proceed by induction. Assume the result for n. Let $v_i \in S$, $\alpha_i \in \mathbb{R}$ for $1 \leq i \leq n+1$ be arbitrary with $\sum_i \alpha_i = 1$. There exists at least one i such that $\alpha_i \neq 1$. Without loss of generality, we assume that $\alpha_{n+1} \neq 1$. Let us denote α_{n+1} by α. Now, $\sum_{i=1}^n \alpha_i = 1 - \alpha \neq 0$. Hence if we let $\beta_i := \frac{\alpha_i}{1-\alpha}$, then $\sum_{i=1}^n \beta_i = 1$. By induction hypothesis, $\sum_{i=1}^n \beta_i v_i \in S$. Call it w. Hence $(1-\beta)w + \beta v_{n+1} \in S$. But this is nothing other than $\sum_{i=1}^{n+1} \alpha_i v_i$.

□

Now if you go back to Chapter 1 on systems of linear equations, you will realize the following:

Let S be the set of solutions of a (possibly) non-homogeneous system (1.2.1). Assume $S \neq \emptyset$. Then S is an affine space and $S = a + S_h$ for some (and hence any) a in S. Here S_h stands for the set of solutions of the associated homogeneous system.

Definition 3.2.2 The dimension of an affine space S is $\dim W$ if $S = v + W$ for some $v \in S$.

An affine space of dimension $\dim V - 1$ is called a *hyperplane*.

Exercise 3.2.1 A hyperplane in $\mathbb{R}^2$ is a line. A hyperplane in $\mathbb{R}^3$ is a plane.

3.3 Quotient Space

Definition 3.3.1 Let W be a vector subspace of a vector space V over $\mathbb{R}$. By a *coset* of W in V we mean a set of the form $v + W := \{v + w \mid w \in W\}$ for some $v \in V$.

We look at a couple of examples to get a feeling of this concept.

Example 3.3.1 Let $V = \mathbb{R}^2$ and $W = \mathbb{R}e_1 = \{(x, 0) \mid x \in \mathbb{R}\}$. Take any $v = (a, b) \in \mathbb{R}^2$. Then we have

$$\begin{aligned} v + W &= \{(a,b) + (x,0) \mid x \in \mathbb{R}\} \\ &= \{(a+x, b) \mid x \in \mathbb{R}\} \\ &= \{(x', b) \mid x' \in \mathbb{R}\}. \end{aligned}$$

That is, $v + W$ is the line ℓ through (a, b) parallel to the x-axis (see Figure 3.3.1). In the notation of the last section $\ell = \ell((a,b); e_1)$. Note that if $(a', b') \in \ell$, then $(a', b') + W = (a, b) + W$. (Check!)

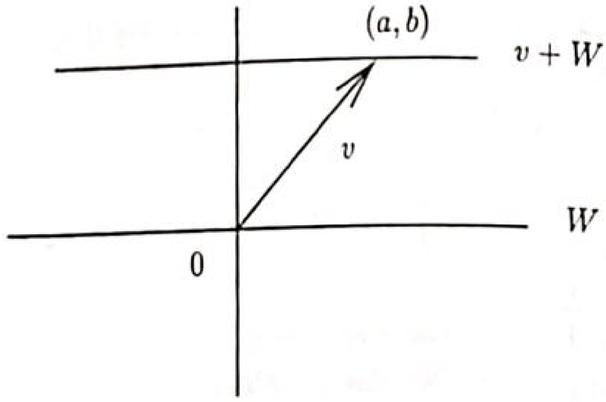


Figure 3.3.1 Coset of W.

Example 3.3.2 We shall be brief here. Let $V = \mathbb{R}^3$ and

$$W = \{(x, y, 0) \mid x, y \in \mathbb{R}\}$$

be the xy-plane. Then the coset $v + W$ for any $v \in \mathbb{R}^3$ is the plane parallel to the xy-plane through the point $v = (a, b, c)$ at "height" c.

Just to make sure that you understand these two examples, work out the following exercise:

Exercise 3.3.1 Let $W := \{(x, y) \in \mathbb{R}^2 \mid ax + by = 0\}$ for a fixed $(a, b) \in \mathbb{R}^2 \setminus (0, 0)$. Show that W is a one-dimensional subspace of $V = \mathbb{R}^2$ and that the cosets of W in V are the lines parallel to the line $ax + by = 0$ and hence are given by $ax + by = c$ for $c \in \mathbb{R}$.

We now want to talk of the set of cosets of W in V. In Example 3.3.1, the set of cosets are the lines parallel to the x-axis. We denote the set of cosets of W in V by V/W (read as V mod W).

We want to define "addition" of two elements of V/W. Let us look xample 3.3.1 to see how to go about doing this. Any "point" ξ ounced 'xi') of V/W is a line given as, say, $y = a$. Thus we may e the "addition" of two such "points" given by $y = a$ and $y = b$ as the oint" given by $y = a + b$ (see Figure 3.3.2).

Note that the point ξ_1 (respectively ξ_2) of V/W given by $y = a$ (respectively by $y = b$) is the coset $(0, a) + W$ (respectively by $(0, b) + W$). So what we have done is to define $\xi_1 + \xi_2 = \xi$ where ξ is the coset $(0, a + b) + W$. That is,

$$((0, a) + W) + ((0, b) + W) := (0, a + b) + W.$$

We can also multiply the points of V/W by scalars (that is, real numbers) as follows:

$$\alpha((0, a) + W) := (0, \alpha a) + W.$$

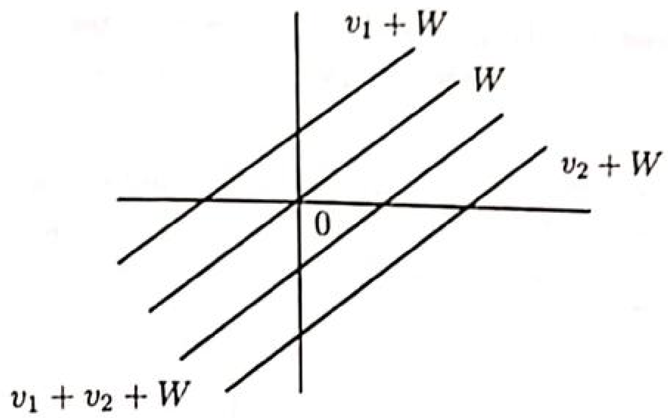


Figure 3.3.2 Addition of cosets.

We invite the reader to show that V/W with respect to these two operations becomes a vector space over $\mathbb{R}$. What is its dimension?

After this special but illuminating case we wish to do a similar thing for V/W for any vector space V and any vector subspace W of V. Before doing this we observe a crucial fact about the cosets.

Lemma 3.3.1 *Let W be a vector subspace of a vector space V over $\mathbb{R}$. Let $v_i + W$ be cosets for $i = 1, 2$. Then exactly one of the following is true:*

(1) $(v_1 + W) \cap (v_2 + W) = \emptyset$.

(2) $v_1 + W = v_2 + W$.

Moreover, $v_1 + W = v_2 + W$ if and only if $v_1 - v_2 \in W$.

Proof *This is geometrically clear (see Figure 3.3.3). For $v_1 + W$ and $v_2 + W$ are "parallel subspaces" and hence either they coincide or they are disjoint.*

We prove this algebraically. Let $v_1 + W$ and $v_2 + W$ have non-empty intersection. Then there exists $v \in V$ such that $v \in v_i + W$. Since $v \in v_i + W$ there exists $w_i \in W$ such that $v = v_i + w_i$. But then $v_1 - v_2 = w_2 - w_1$. Since W is a vector space, $w_1 - w_2 \in W$. Therefore $v_1 - v_2 \in W$. (This proves the last statement of the lemma.) Let $u = v_1 - v_2 \in W$.

Let $x \in v_1 + W$. Then $x = v_1 + w$ for some $w \in W$. Since

$$v_1 = (v_1 - v_2) + v_2$$

and $u = v_1 - v_2 \in W$ we see that $x = u + v_2 + w$ or $x = v_2 + w'$ where $w' = u + w \in W$. Thus $x \in v_2 + W$. Since x was an arbitrary element of $v_1 + W$, we have thus proved that $v_1 + W \subset v_2 + W$. Interchanging v_1 and

v_2 we see that $v_2 + W \subset v_1 + W$. Thus we have proved that if the cosets $v_1 + W$ and $v_2 + W$ have non-empty intersection, then they are the same.

Do you see that the last statement of the lemma is the algebraic version of Proposition 3.1.1? Think about this before going further. You may also want to go back and compare the proof of this assertion with that of Proposition 3.1.1.

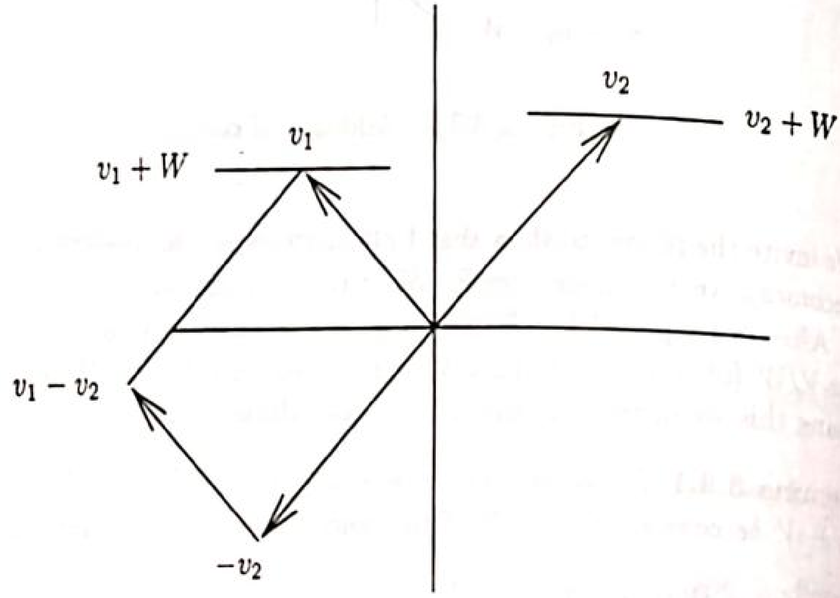


Figure 3.3.3 Illustration for Lemma 3.3.1.

Definition 3.3.2 Let $\xi \in V/W$ be a coset. If $\xi = v + W$, then v is called a *representative* of ξ.

The representative is by no means unique. For instance, in Example 3.3.1, for the coset $y = a$, any point on the line (that is, any point of the form (x, a)) is a representative. What we saw in the proof of Lemma 3.3.1 is that if $\xi = v_1 + W$ as well as $\xi = v + W$, then $v - v_1$ is an element of W. Thus, any two representatives of a coset of W differ by an element of W. In particular, $x + W = W$ if and only if $x \in W$. (But, this is a solution of Exercise 2.2.21!)

We now show how to define "addition" of two cosets $\xi_i := v_i + W$ in V/W. Let v_i be a representative of ξ_i for $i = 1, 2$. Then we define $\xi_1 + \xi_2$ to be the coset whose representative is $v_1 + v_2$. That is, $\xi_1 + \xi_2 = (v_1 + v_2) + W$.

We need to show that this coset $\xi_1 + \xi_2$ is defined without any ambiguity. This is usually called "the well-definedness" of the concept. Can you identify the possible source of confusion in the definition of $\xi_1 + \xi_2$? For, ८०८

I may choose a representative v_i for ξ_i, for $i = 1, 2$. You may choose u_i to be a representative of ξ_i. According to me the sum $\xi_1 + \xi_2$ is the coset $(v_1 + v_2) + W$ whereas according to you the sum is $(u_1 + u_2) + W$.

Which of us is right? Well, we both are! That is, I claim that

$$(v_1 + v_2) + W = (u_1 + u_2) + W.$$

The equality holds if and only if $(v_1 + v_2) - (u_1 + u_2) \in W$. That is, if and only if $(v_1 - u_1) + (v_2 - u_2) \in W$. But this is true, since v_i and u_i are representatives of the same coset and hence $v_i - u_i \in W$. Since W is a vector subspace, $(v_1 - u_1) + (v_2 - u_2) \in W$. Hence the claim is proved. Thus, to define the sum of two cosets we may use any two representatives.

Now I am sure you know how to define scalar multiplication on V/W. Did you get the following definition? If $\alpha \in \mathbb{R}$ and $\xi = v + W \in V/W$, then $\alpha(\xi) := \alpha v + W$. As earlier, we may show that this is well-defined. If $\xi = u + W$, then $\alpha v + W = \alpha u + W$ as $\alpha v - \alpha u = \alpha(v - u) \in W$ since $v - u \in W$. (Why is $v - u \in W$?)

Thus we have defined addition and scalar multiplication on the set V/W. We claim that V/W with these two operations is a vector space over $\mathbb{R}$. We state this as a theorem.

Theorem 3.3.2 *Let W be a vector subspace of V. Let V/W denote the set of cosets of V with respect to W. The following operations are well-defined:*

(1) $\xi_1 + \xi_2 = (v_1 + v_2) + W$ *where* $\xi_i := v_i + W \in V/W$.

(2) $\alpha\xi = (\alpha v) + W$ *where* $\xi = v + W \in V/W$.

V/W with these operations becomes a vector space.

Proof That the operations are well-defined is proved above. Suppose we want to prove the associativity of the addition on V/W: Let

$$\xi_i = v_i + W, \ 1 \le i \le 3$$

be arbitrary. Then we need to show that

$$(\xi_1 + \xi_2) + \xi_3 = \xi_1 + (\xi_2 + \xi_3).$$

That is, to show that

$$((v_1 + v_2) + W) + (v_3 + W) = (v_1 + W) + (v_2 + v_3 + W).$$

The left side is $[(v_1+v_2)+v_3]+W$, while the right side is $[v_1+(v_2+v_3)]+W$. But by the associativity of addition in V, we know that

$$(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$$

so that both the cosets $[(v_1 + v_2) + v_3] + W$ and $[v_1 + (v_2 + v_3)] + W$ are equal to $(v_1 + v_2 + v_3) + W$.

What is the zero element of V/W?

The rest of the proof goes along the same lines. The burden of showing something in V/W relies on th. observation that its analogue is true in V.

□

V/W is called the *quotient space* of V with respect to W. What is its dimension?

Theorem 3.3.3 *Let V be a finite dimensional vector space and W a vector subspace of V. Then* $\dim V/W = \dim V - \dim W$.

Proof Let $\{w_1, \ldots, w_m\}$ be a basis of W. Extend this to a basis

$$\{w_1, \ldots, w_m, v_1, \ldots, v_n\}$$

of V. We show that $\{v_i + W\}_{i=1}^n$ is a basis of V/W. Let $x + W \in V/W$. Since $x \in V$, $x = \sum_{i=1}^m \alpha_i w_i + \sum_{j=1}^n \beta_j v_j$ for some α_i's and β_j's. Hence

$$\begin{aligned} x + W &= \sum_i \alpha_i w_i + \sum_j \beta_j v_j + W \\ &= \sum_{j=1}^n \beta_j v_j + W \qquad \text{by Exercise 2.2.21} \\ &= \sum_{j=1}^n \beta_j (v_j + W). \end{aligned}$$

Therefore every element of V/W can be expressed as a linear combination of $\{v_1 + W, \ldots, v_n + W\}$. We prove that $\{v_1 + W, \ldots, v_n + W\}$ is linearly independent. Consider $\alpha_1(v_1 + W) + \cdots + \alpha_n(v_n + W)$ as the zero element V/W for $\alpha_i \in \mathbb{R}$. Then

$$\sum_i \alpha_i(v_i + W) = \Big(\sum_i \alpha_i v_i\Big) + W = W.$$

It follows that $\sum_{i=1}^n \alpha_i v_i \in W$. Since $\{w_j\}_{j=1}^m$ is a basis of W, we can write $\sum_{i=1}^n \alpha_i v_i = \sum_{j=1}^m \beta_j w_j$ and thus $\sum_{i=1}^n \alpha_i v_i - \sum_{j=1}^m \beta_j w_j = 0$. Since $\{w_1, \ldots, w_m, v_1, \ldots, v_n\}$ is a basis of V, we deduce that $\alpha_i = 0$ for all i and $\beta_j = 0$ for all j. Thus $\sum_i \alpha_i(v_i + W) = 0$ in V/W implies that $\alpha_i = 0$ for all i. Hence $\{v_i + W\}_{i=1}^n$ is linearly independent and we have $\dim V/W = n$ and $\dim V - \dim W = m + n - m = n$.

□

4. Linear Transformations

If X and Y are any two arbitrary sets, there is no obvious restriction on the kind of maps between X and Y, except that it is one-one or onto. However if X and Y have some additional structure, we wish to consider those maps which in some sense "preserve" the extra structure on the sets X and Y.

4.1 Linear Transformation

Informally, a "linear transformation" preserves algebraic operations. The sum of two vectors is mapped to the sum of their images and the scalar multiple of a vector is mapped to the same scalar multiple of its image. More precisely, we have the following definition:

Definition 4.1.1 A *linear transformation (or a linear map)* T from a vector space V to a vector space W is a map which satisfies:

(i) $T(0) = 0$.

(ii) $T(v_1 + v_2) = T(v_1) + T(v_2)$, for $v_1, v_2 \in V$.

(iii) $T(\alpha v) = \alpha T(v)$, $\alpha \in \mathbb{R}$ and $v \in V$.

Remark 4.1.1 Condition (i) is not needed (see Proposition 4.1.1).

Remark 4.1.2 This remark may be omitted in the first reading.

Geometrically, a linear map sends lines passing through the origin to lines passing through the origin or onto the origin.

Let $z \in \ell(x, y)$, $x, y \in V$. Then we can write

$$z = x + t(y - x) = (1 - t)x + ty$$

for a unique $t \in \mathbb{R}$. We say that z divides the line segment

$$[x, y] := \{x + s(y - x) \mid 0 \leq s \leq 1\}$$

in the ratio $(1-t):t$. Note that the sense of direction is important here. For, if we write $z = y + s(x-y)$ then $s = (1-t)$ so that z divides the line segment $[y, x]$ in the ratio $(1-s):s$, that is, in the ratio $t:(1-t)$. Now returning to our earlier notation, if v lies on the line joining v_1 and v_2 and divides it in the ratio $t:1-t$, then $T(v)$ also lies on the line joining $T(v_1)$ and $T(v_2)$ and divides it in the same ratio. This statement is not quite true, since $T(v_1)$ and $T(v_2)$ could be equal so that there is no unique line joining them. However, with a proper convention — in such a case the line is to be taken as the point $T(v_i)$ — the observation remains correct.

Let us look at some examples of linear transformations. The verifications are left to the reader.

Example 4.1.1 The map $T: V \to W$ defined by $Tv = 0$ for all $v \in V$ is a linear transformation.

Example 4.1.2 The simplest kind of linear transformation $T: V \to V$ is the identity map I where $I(v) = v$ for $v \in V$. More generally, if $\alpha \in \mathbb{R}$, then $T = \alpha I$ defined by $\alpha I(v) = \alpha v$ is linear.

Example 4.1.3 Let $V = \mathbb{R}^m$ and $W = \mathbb{R}^n$ with $m \le n$. Consider the map $T: V \to W$ given by

$$T(x_1, \dots, x_m) = (x_1, \dots, x_m, 0, \dots, 0), \ (n-m \text{ zeroes}).$$

Then T is a one-one linear map called the *natural inclusion* of $\mathbb{R}^m$ into $\mathbb{R}^n$.

Example 4.1.4 Let $V = \mathbb{R}^m$ and $W = \mathbb{R}^n$ but now assume that $m \ge n$. Consider $T: V \to W$ defined by $T(x_1, \dots, x_m) = (x_1, \dots, x_n)$. That is, we drop the last $m-n$ coordinates of the vector from $\mathbb{R}^m$. (T is called the *natural projection* of $\mathbb{R}^m$ onto $\mathbb{R}^n$.

Example 4.1.5 Let $V = \mathbb{R}^n = W$. Fix scalars $\alpha_i \in \mathbb{R}$, $1 \le i \le n$. Let $Tx := (\alpha_1 x_1, \dots, \alpha_n x_n)$. Then T is a linear transformation.

Example 4.1.6 Let $V = \mathbb{R}^2 = W$. Consider $R_\theta: V \to W$ given by

$$R_\theta\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) := \begin{pmatrix} \cos\theta\, x + \sin\theta\, y \\ -\sin\theta\, x + \cos\theta\, y \end{pmatrix}.$$

Then R_θ is a linear map. It is the rotation by an angle θ from the positive x-axis. See Example 6 in Section 4.6.

Example 4.1.7 Consider $\mathbb{C}$ as a vector space over $\mathbb{R}$ (as in Exercise 2.1.21). Consider the map $T: \mathbb{C} \to \mathbb{C}$ given by $z \mapsto iz$, the multiplication by i. T is linear map.

Exercise 4.1.1 The conjugation map from $\mathbb{C}$ to itself given by $z \mapsto \bar{z}$ is linear.

Exercise 4.1.2 Fix $v \in V$. Consider the map $T_v : V \to V$ given by

$$Tx = x + v.$$

Then T_v is a linear map if and only if $v = 0$. (T_v is called the *translation* of V by v.)

Exercise 4.1.3 Let $T : \mathbb{R} \to \mathbb{R}$ be given by $T(x) = x^2$. Then T is not linear.

Exercise 4.1.4 Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $T(x, y) = (x, y, xy)$. Then T is not linear.

Exercise 4.1.5 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(x, y) = (x, y + 3)$. Is T linear?

Exercise 4.1.6 Fix $A \in M(n, \mathbb{R})$. Consider the map

$$L_A : M(n, \mathbb{R}) \to M(n, \mathbb{R})$$

defined by $L_A(X) = AX$, the product of two matrices. Then L_A is a linear map. Is the square map $X \mapsto X^2$ linear?

Exercise 4.1.7 The map T of Example 2.1.3 is a bijective linear map.

Proposition 4.1.1 *Let $T : V \to W$ be a linear map. Then the following are true:*

(1) $T(0) = 0$. *That is, T maps the zero element of V to that of W.*

(2) $T(-v) = -T(v)$ *for $v \in V$, that is, T maps the negative of an element of V to the negative of the image.*

(3) $T(v_1 - v_2) = T(v_1) - T(v_2)$.

Proof Just for the purpose of this proof, we shall let 0_V (respectively 0_W) denote the zero element of V (respectively W). Since $0_V = 0_V + 0_V$, we have

$$T(0_V) = T(0_V + 0_V).$$

But,

$$0_V + 0_V = 1 \cdot 0_V + 1 \cdot 0_V = 2 \cdot 0_V$$

so that

$$T(0_V) = T(2 \cdot 0_V) = 2 \cdot T(0_V).$$

Hence $2 \cdot T(0_V) - T(0_V) = 0_W$ or $T(0_V) = 0_W$. This proves (1).

To prove (2), $T(-x) = T((-1)\cdot x) = (-1)\cdot T(x) = -T(x)$ where w have used (4) of Theorem 2.1.1 twice.

You should now prove (3) just to show that you have understood th above arguments.

□

Definition 4.1.2 If V and W are two vector spaces, we denote the set o all linear maps from V to W by $L(V,W)$.

$L(V,\mathbb{R})$ is usually denoted by V^* and called the *dual* of V.

Any linear transformation $T : V \to V$ is called an *endomorphism* of V. We let End $(V) = L(V,V)$.

The following proposition states that $L(V,W)$ is a vector space.

Proposition 4.1.2 *Let $S,T \in L(V,W)$ and $\alpha \in \mathbb{R}$. Then $S+T$ and αS defined by $(S+T)(x) = Sx+Tx$ and $(\alpha S)x = \alpha Sx$ are again linear maps. With these operations $L(V,W)$ is a vector space.*

Proof A routine verification and hence left to the reader.

□

Definition 4.1.3 Any linear map $T : V \to \mathbb{R}$ is called a *linear functional* or *linear form*. It is customary to denote the linear functionals by letters f, g, etc. The set of linear forms on a given vector space is denoted by V^* and is called the *dual* of V.

Example 4.1.8 Show that the f defined at the end of Section 1.1 is a linear functional. (What is the vector space? The underlying field?)

Example 4.1.9 Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x_1, \dots, x_n) = x_i$, for a fixed i. Then f_i is a linear form. f_i's are also called *coordinate functions*. Can you generalize this?

Exercise 4.1.8 Let $\{v_i\}_{i=1}^n$ be a basis of V. Define $f_i : V \to \mathbb{R}$ by $f_i(v) = \alpha_i$ if $v = \sum_{k=1}^n \alpha_k v_k$. Then f_i is a linear form. That is, the coordinate functions with respect to a basis are linear forms.

Example 4.1.10 Let us find all linear maps $f : \mathbb{R} \to \mathbb{R}$. If we fix any nonzero vector, say, $x_0 \in \mathbb{R}$, then any $x \in \mathbb{R}$ can be written as $x = \alpha x_0$. If $f(x_0) = y_0$, then $f(x) = f(\alpha x_0) = \alpha f(x_0) = \alpha y_0$. In particular, if $x_0 = 1$, then $f(x) = f(1)x$. Thus all linear maps from $\mathbb{R}$ to $\mathbb{R}$ are of the form $f(x) = \alpha x$ for a fixed $\alpha \in \mathbb{R}$.

Exercise 4.1.9 Let V be a one-dimensional vector space. Find all linear forms on V.

Example 4.1.11 We extend the reasoning above to find all linear forms $f: \mathbb{R}^n \to \mathbb{R}$. Fix the standard basis $\{e_1, \ldots, e_n\}$ of $\mathbb{R}^n$. Let f be a linear form and let $\alpha_i := f(e_i)$. Then

$$f(x) = f\left(\sum_{i=1}^{n} x_i e_i\right) = \sum_{i=1}^{n} f(x_i e_i) = \sum_{i=1}^{n} x_i f(e_i) = \sum_{i=1}^{n} x_i \alpha_i.$$

Conversely, if $(\alpha_1, \ldots, \alpha_n)$ is any n-tuple and if we define $f(x) = \sum_{i=1}^{n} x_i \alpha_i$, then f is a linear form. Is there any special reason for us to employ the standard basis in the above argument?

Exercise 4.1.10 Let $f_i : \mathbb{R}^m \to \mathbb{R}$ be arbitrary functions. Let $T: \mathbb{R}^m \to \mathbb{R}^n$ be defined by $T(x_1, \ldots, x_m) = (f_1(x), \ldots, f_n(x))$. When is T linear? *Hint:* Review Exercise 4.1.3 through Exercise 4.1.5.

Exercise 4.1.11 Let V be an n-dimensional vector space. Find all linear forms on it.

If the reader goes through the last two examples and exercises, he will understand the most important fact:

> **Any linear transformation $T : V \to W$ is completely determined by its action on a basis of V.**

This is the content of the next theorem.

Theorem 4.1.3 *Let V and W be vector spaces. Let $\{v_1, \ldots, v_n\}$ be a basis of V. Let w_i, $1 \le i \le n$ be any set of (not necessarily distinct) vectors in W. Then there is a unique linear map $T : V \to W$ such that $T(v_i) = w_i$.*

Proof Let $v \in V$. Then, $v = \sum_{i=1}^{m} \alpha_i v_i$. Define $T : V \to W$ by

$$T(v) = \sum_i \alpha_i w_i.$$

We have $v_i = 0v_1 + 0v_2 + \cdots + 1v_i + 0v_{i+1} + \cdots + 0v_n$. Hence

$$T(v_i) = 0w_1 + 0w_2 + \cdots + 1w_i + 0w_{i+1} + \cdots + 0w_n.$$

Therefore $T(v_i) = w_i$.

Claim: T is a linear map.

Let $u, v \in V$. Then $v = \sum_{i=1}^{n} \alpha_i v_i$ and $u = \sum_{i=1}^{n} \beta_i v_i$. Therefore,

$$T(v) = \sum_{i=1}^{n} \alpha_i w_i,\ T(u) = \sum_{i=1}^{n} \beta_i w_i$$

By Exercise 2.3.11, $v + u = \sum_i (\alpha_i + \beta_i) v_i$. Therefore

$$\begin{aligned} T(v+u) &= T\left(\sum_{i=1}^{n} (\alpha_i + \beta_i) v_i\right) \\ &= \sum_{i=1}^{n} (\alpha_i + \beta_i) w_i \\ &= \sum_{i=1}^{n} \alpha_i w_i + \sum_{i=1}^{n} \beta_i w_i \\ &= T(v) + T(u). \end{aligned}$$

Let $\alpha \in \mathbb{R}$, then $\alpha v = \alpha \left(\sum_{i=1}^{n} \alpha_i v_i\right) = \sum_{i=1}^{n} (\alpha \alpha_i) v_i$. Therefore

$$T(\alpha v) = \sum_{i=1}^{n} (\alpha \alpha_i) w_i = \alpha \sum_{i=1}^{n} \alpha_i w_i = \alpha T(v).$$

Hence T is a linear map.

To prove uniqueness part of the theorem, let T' be a linear map from to W such that $T'(v_i) = w_i$. Then we claim $T = T'$.

Let $v \in V$. Then $v = \sum_{i=1}^{n} \alpha_i v_i$. Therefore

$$T'(v) = T'\left(\sum_{i=1}^{n} \alpha_i v_i\right) = \sum_{i=1}^{n} \alpha_i T'(v_i) = \sum_{i=1}^{n} \alpha_i w_i = T(v).$$

Since $v \in V$ was arbitrary, $T' \equiv T$. Thus T is unique.

From Proposition 4.1.2, we know that the set $L(V, W)$ of linear ma from V to W is a vector space. What is its dimension?

Theorem 4.1.4 *Let V and W be vector spaces. Then the dimension* $L(V, W)$ *is* $\dim V \times \dim W$.

Proof This is an application of Theorem 4.1.3. We shall sketch the pro leaving the details to the reader. Another proof can be obtained using t results of Section 4.4 (see Exercise 4.4.8).

Fix a basis $\{v_i\}_{i=1}^m$ of V and $\{w_j\}_{j=1}^n$ of W. Define $T_{ij} \in L(V,W)$ by setting

$$T_{ij}(v_r) = \begin{cases} 0 & \text{if } v_r \neq v_i \\ w_j & \text{if } v_r = v_i \end{cases}$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$. (Such a T_{ij} exists and is unique by Theorem 4.1.3.) The theorem follows from the claims:

(1) $\{T_{ij}\}$ generate $L(V,W)$.

(2) $\{T_{ij}\}$ is linearly independent.

To prove claim (1), Theorem 4.1.3 is again invoked. If $T \in L(V,W)$ and $T(v_i) = y_i \in W$, express y_i in terms of w_j's. Now you are on your own.

□

Exercise 4.1.12 Let $A\colon V \to W$ and $B\colon W \to U$ be a linear map. Then $B \circ A\colon V \to U$ is a linear map.

Exercise 4.1.13 (Dilations) Fix $\alpha \in \mathbb{R}$ and let $T_\alpha : V \to V$ be given by $T_\alpha(v) = \alpha v$. If $\alpha = 0$, then $T_\alpha(v) = 0$ for all $v \in V$. If $\alpha = 1$, then $T_\alpha(v) = v$, that is, T_α is the identity map.

Exercise 4.1.14 Let $\dfrac{d}{dX} : \mathcal{P}_n \to \mathcal{P}_n$ be defined by

$$\frac{d}{dX}\left(\sum_{k=0}^{n} a_k X^k\right) := \sum_{k=1}^{n} k a_k X^{k-1}.$$

Then $\dfrac{d}{dX}$ is a linear map of $\mathcal{P}_n$ into the vector subspace $\mathcal{P}_{n-1}$.

To solve the next three exercises, you need results from analysis.

Exercise 4.1.15 Let $S, T\colon C[0,1] \to \mathbb{R}$ be defined by $S(f) := f(t_0)$, and $T(f) := \int_0^1 f(t)\,dt$ where $t_0 \in [0,1]$. Then S and T are linear.

Exercise 4.1.16 Let C be the set of all convergent real sequences and let $T : C \to \mathbb{R}$ be defined by $(x_n) \mapsto \lim x_n$. Then T is a linear transformation.

Exercise 4.1.17 Let $\mathcal{D}[0,1]$ be the set of all continuously differentiable functions on $[0,1]$ and let $T : \mathcal{D}[0,1] \to C[0,1]$ by $f \mapsto f'$. Then T is linear.

Exercise 4.1.18 Let $V = \mathbb{R}^n$ and W be the subspace given by

$$W := \{x \in \mathbb{R}^n \mid x_n = 0\}.$$

Consider $P : V \to W$ given by $(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_{n-1}, 0)$. Then P is a linear map called the natural projection. What is P^2?

Exercise 4.1.19 Let $V := \mathbb{R}^n$ and $W := \mathbb{R}^{n+1}$. Define $\varphi: V \to W$ by $\varphi(x_1, \dots, x_n) = (x_1, \dots, x_n, 0)$. Then φ is a linear map called the *natural inclusion.*

Exercise 4.1.20 Can you think of generalizations of Exercises 4.1.18 and 4.1.19?

Exercise 4.1.21 Let $V = \mathcal{P}_n$ and $W = \mathbb{R}^m$. For $P(X) \in V$, and $\alpha \in \mathbb{R}$, we let $P(\alpha)$ be the "value" of P at $X = \alpha$, obtained by substituting α for X. Let $\alpha_1, \dots, \alpha_m$ be any real scalars. The map $T: V \to W$ given by $TP = (P(\alpha_1), \dots, P(\alpha_m))$ is a linear map.

Exercise 4.1.22 Let V be a vector space. Fix a basis $\{v_i\}_{i=1}^n$ of V. To define linear forms on V, by Theorem 4.1.3 it is enough to define $f(v_i)$. Define $f_i \in V^*$ by setting

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

Show that $\{f_i \mid 1 \leq i \leq n\}$ is a basis of V^*, called the basis of V^* dual to the given basis $\{v_i\}$ of V.

Exercise 4.1.23 Let $V = \mathbb{R}^n$. What is the basis of V^* dual to the standard basis of $\mathbb{R}^n$?

Exercise 4.1.24 Let V be the vector space of solutions f of

$$y^{(n)} + P_1(x)y^{(n-1)} + \dots + P_n(x)y = 0.$$

(Refer to Exercise 2.1.22.) Let $W = \mathbb{R}^k$, for a fixed k with $1 \leq k \leq n$. Consider $T : V \to W$ given by $Tf = (f(0), f'(0), \dots, f^{(k-1)}(0))$. Then T is linear.

The following exercise is an important one. It introduces a family of linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$.

Exercise 4.1.25 Let $A := (a_{ij})$ be an $n \times m$ matrix with real entries. We let $V := \mathbb{R}^m$ and $W := \mathbb{R}^n$. We write the elements of these vector spaces as column vectors (rather than row vectors as we have been doing so far):

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \in \mathbb{R}^m, \text{ and } y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n.$$

We define $Tx := Ax$, the matrix multiplication of $n \times m$ matrix A by the $m \times 1$ matrix x thereby getting an $n \times 1$ matrix. The latter is considered as a vector in $\mathbb{R}^n$. Show that T is a linear map.

Convention. Whenever we write Ax where A is an $n \times m$ matrix and $x \in \mathbb{R}^m$, we assume x is written as a column vector or an $m \times 1$ matrix so that the matrix multiplication is defined.

Exercise 4.1.26 Go through Example 4.1.11. Can you reformulate the result there in light of Exercise 4.1.25?

Exercise 4.1.27 Let $M_{n\times m}(\mathbb{R})$ denote the set of all $n \times m$ real matrices. From Exercise 4.1.25, we have a map $A \mapsto T$ (the linear map associated to A) from $M_{n\times m}(\mathbb{R})$ to $L(\mathbb{R}^m, \mathbb{R}^n)$. Investigate whether this map is one-one, onto.

Exercise 4.1.28 Let A be as in Exercise 4.1.25. Let V and W be vector spaces with $\dim V = m$ and $\dim W = n$. Fix a basis $\{v_i\}$ of V and $\{w_j\}$ of W. Can you associate a linear transformation T from V to W?

Exercise 4.1.29 Let $T\colon V \to W$ be linear. Let V^* and W^* be their duals (see Definition 4.1.3). We define a map $T^*\colon W^* \to V^*$ as follows. Given $g \in W^*$, $Tg \in V^*$ is given by $Tg(v) := g(Tv)$ for all $v \in V$. Show that T^* is a linear map. It is called the *adjoint* of T.

4.2 Representation of Linear Maps by Matrices

Let V be an m-dimensional vector space and W be an n-dimensional vector space. Let $\{v_1, \dots, v_m\}$ and $\{w_1, \dots, w_n\}$ be bases of V and W respectively. The aim of this section is to show that there exists a bijective linear map from $L(V, W)$ to $M_{n\times m}(\mathbb{R})$. This fact was hinted at in the last few exercises of the previous section.

Let $T\colon V \to W$ be a linear map. Then $T(v_i) \in W$. Therefore,

$$T(v_i) = \sum_{j=1}^{n} a_{ij} w_j, \quad 1 \le i \le m.$$

We write this in an expanded form

$$\begin{aligned}
Tv_1 &= a_{11}w_1 + \cdots + a_{1n}w_n \\
Tv_2 &= a_{21}w_1 + \cdots + a_{2n}w_n \\
\vdots \quad & \quad \vdots \qquad\qquad \vdots \\
Tv_m &= a_{m1}w_1 + \cdots + a_{mn}w_n.
\end{aligned} \tag{4.2.1}$$

We define the matrix $M_w^v(T)$ of T with the choice of bases $\{v_i\}$ of V and $\{w_j\}$ of W to be the transpose of the matrix of the coefficients in Equation (4.2.1). That is,

$$M_w^v(T) := \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1i} & a_{2i} & \cdots & a_{mi} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}.$$

The matrix $M_w^v(T)$ is called the matrix *associated* with T with respect to the bases $\{v_1, \ldots, v_m\}$ and $\{w_1, \ldots, w_n\}$. Also, $M_w^v(T)$ is called the matrix representation of T with respect to these bases. Note that the matrix $M_w^v(T)$ is an $n \times m$ matrix whose first column is the coefficients of Tv_1 when expressed as a linear combination of w_j and so on.

> **The matrix $M_w^v(T)$ is the $n \times m$ matrix whose ith column is the coefficients of Tv_i when expressed as a linear combination of w_j, $1 \le i \le m$.**

This is the secret recipe which allows you to write the matrix of a linear transformation with respect to the given bases. Let us put it into use.

Example 4.2.1 Let $V = \mathbb{R}^2 = W$. Consider the linear map

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

We use the standard basis $\{e_1 = (1,0), e_2 = (0,1)\}$ on both V and W. We want the matrix associated to T with respect to these bases. We prepare the matrix using the recipe

$$\begin{aligned} Te_1 = (1,1) &= (1,0) + (0,1) = 1 \cdot e_1 + 1 \cdot e_2 \\ Te_2 = (1,-1) &= (1,0) + (0,-1) = 1 \cdot e_1 - 1 \cdot e_2. \end{aligned}$$

The first column of $M_e^e(T)$ is therefore $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ while the second is $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Hence

$$M_e^e(T) = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

To consolidate this idea, do the next couple of exercise before you proceed.

Exercise 4.2.1 Let the notation be as above. Let

$$T(x, y) = (ax + by, cx + dy),$$

Find the matrix representation of T with respect to the standard basis.

Exercise 4.2.2 Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^4$. Let T be defined by

$$T(x, y) = (x, y, x + y, x - y).$$

Find the matrix of T with respect to the standard bases.

Exercise 4.2.3 Let $\{v_1, \dots, v_n\}$ be a basis of $\mathbb{R}^n$. Find the matrix $A \in M(n, \mathbb{R})$ when considered as a linear map from $\mathbb{R}^n$ to itself which takes the standard basis vector e_i to v_i for $1 \le i \le n$.

Let us do one more example which will be an eye opener.

Example 4.2.2 Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(x, y) = (x + y, x - y)$. We choose $\{v_1 = (0, 1), v_2 = (1, 0)\}$ as a basis for both the domain and range of T. What is the corresponding matrix of T? Let us do it in a systematic way as earlier:

$$\begin{aligned} Tv_1 = (1, -1) &= (0, -1) + (1, 0) = (-1) \cdot v_1 + 1 \cdot v_2 \\ Tv_2 = (1, 1) &= (1, 0) + (0, 1) = 1 \cdot v_1 + 1 \cdot v_2. \end{aligned}$$

Thus the required matrix is

$$\begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Compare this with that of Example 4.2.1. Do you see the difference? Even though the basis in Example 4.2.1 is same as that in Example 4.2.2 **as sets**, the order in which they are listed seems to matter. The lesson we learn from it is

> **In writing the matrix associated to a linear map with respect to the given bases, the order in which the elements appear in the lists matter.**

Just to make sure that you appreciate this, redo Exercise 4.2.1 with the basis $\{e_2, e_1\}$ and compare the new matrix with the earlier one (see also Exercise 4.2.12).

Example 4.2.3 Let p_1 be the map $p_1\colon \mathbb{R}^2 \to \mathbb{R}$ by $p_1(x,y) = x$. Then $p_1(e_1) = p_1(1,0) = 1$ and $p_1(e_2) = p_1(0,1) = 0$. Therefore the matrix associated with p_1 is the 1×2 matrix $(1,0)$.

Example 4.2.4 Recall the linear map T, the multiplication by i in $\mathbb{C}$ (see Example 4.1.7). As a basis of $\mathbb{C}$ we take $v_1 = 1$ and $v_2 = i$ (see Exercise 2.3.22). What is the matrix representation of T with respect to this basis?

$$\begin{aligned} Tv_1 &= i\cdot 1 = i = 0\cdot v_1 + 1\cdot v_2 \\ Tv_2 &= i\cdot i = -1 = -v_1 = (-1)v_1 + 0\cdot v_2. \end{aligned}$$

Thus

$$M_v^v(T) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Exercise 4.2.4 More generally, let $\lambda = a + ib \in \mathbb{C}$ be a complex number. Consider the map $T\colon \mathbb{C} \to \mathbb{C}$ given by $Tz = \lambda z$, the complex multiplication of z by λ. Show that T is a linear map. Compute the matrix of T with respect to the "usual" basis of $\mathbb{C}$.

Exercise 4.2.5 What is the matrix representation of the conjugation map $z \mapsto \bar{z}$ of $\mathbb{C}$ with respect to the "usual" basis of $\mathbb{C}$?

Exercise 4.2.6 If $A\colon \mathbb{R}^2 \to \mathbb{R}^3$ is given by

$$A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ x+y \\ y \end{pmatrix},$$

write down the matrix of A with respect to the standard basis of $\mathbb{R}^2$ and $\mathbb{R}^3$.

Exercise 4.2.7 Let $p_1 : \mathbb{R}^2 \to \mathbb{R}^2$ be the *projection* of $\mathbb{R}^2$ onto its subspace $\mathbb{R}e_1$ defined by $p_1(x,y) = (x,0)$. The matrix associated with p_1 is

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Generalize this to the projection of $\mathbb{R}^n$ onto its subspace

$$W := \{x \in \mathbb{R}^n \mid x_i = 0,\ i \geq m\} \text{ for } 1 \leq m \leq n.$$

Exercise 4.2.8 Let $p_i\colon \mathbb{R}^2 \to \mathbb{R}^2$, $i = 1,2$, be defined by $p_1(x,y) = (x,0)$ and $p_2(x,y) = (0,x)$. Show that $p_1 \circ p_2 = 0$ while $p_2 \circ p_1$ is not. In particular, composition of linear maps is not commutative: $S \circ T$ need not be $T \circ S$, $S,T \in \mathrm{End}(V)$.

What are the matrices of p_i with respect to the standard basis? Do they commute?

Exercise 4.2.9 Let V be a vector space and $\{v_1, \ldots, v_n\}$ a basis of V. Let $\phi: V \to V$ be a linear transformation defined by

$$\phi(v_1) = v_2, \phi(v_2) = v_3, \ldots, \phi(v_{n-1}) = v_n, \phi(v_n) = 0.$$

Let A be the matrix associated with ϕ. We have

$$\phi(v_1) = v_2 = 0v_1 + 1v_2 + 0v_3 + \cdots + 0v_n$$
$$\vdots$$

and

$$\phi(v_i) = v_{i+1} = 0v_1 + 0v_2 + \cdots + 1v_{i+1} + 0v_{i+2} + \cdots + 0v_n.$$

Then we have

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

We have $\phi^n \equiv 0$ and $A^n = 0$.

Exercise 4.2.10 Let $\tau: \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ be a permutation and let $V = \mathbb{R}^n$. Let $\phi_\tau: \mathbb{R}^n \to \mathbb{R}^n$ be defined by $e_i \mapsto e_{\tau(i)}$. Extend ϕ_τ linearly. As a specific example consider the case of $\mathbb{R}^3$, and let

$$\tau = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}.$$

Then, $\phi_\tau(e_1) = e_{\tau(1)} = e_3$, $\phi_\tau(e_2) = e_{\tau(2)} = e_1$ and $\phi_\tau(e_3) = e_{\tau(3)} = e_2$. Hence the matrix of ϕ_τ is given by

$$\phi_\tau = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Matrices corresponding to permutations are called *permutation matrices.*

We now proceed to investigate the relationship between linear maps and the associated matrices.

Let the notation be as in the beginning of the section. Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}$$

be an $n \times m$ matrix. Is there a linear map $T\colon V \to W$ whose matrix wi respect to the given bases is A? If such a T exists, we then must ha $Tv_i = \sum_{j=1}^{n} a_{ji} w_j$. Once T is defined on the basis of V, we can extend "linearly" to all of V (see Theorem 4.1.3).

More specifically, if $v := \sum_{r=1}^{m} \lambda_r v_r$, then

$$\begin{aligned} Tv &= \sum_{r=1}^{m} \lambda_r T v_r \\ &= \sum_{r=1}^{m} \lambda_r \left(\sum_{j=1}^{n} a_{jr} w_j \right) \\ &= \sum_{j=1}^{n} \left(\sum_{r=1}^{m} a_{jr} \lambda_r \right) w_j \\ &= \sum_{j=1}^{n} b_j w_j \end{aligned}$$

where b_j is the $(j1)$th entry of the matrix product

$$A \cdot \lambda = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix}.$$

By the very definition, T is a linear map whose matrix is A.

We have thus proved the following result.

Proposition 4.2.1 *Let A be an $n \times m$ matrix. Fix bases $\{v_i\}_{i=1}^{m}$ of V and $\{w_j\}_{i=1}^{n}$ of W. Let $v \in V$, $v = \sum_{i=1}^{m} \lambda_i v_i$. Define a map T such tha $T(v) = \sum_{j=1}^{n} b_j w_j$, where the b_j's are defined as $b_j = \sum_{r=1}^{m} a_{jr} \lambda_r$. That is,*

$$\begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

Then $T\colon V \to W$ is a linear map such that the matrix of T with respect to the bases $\{v_i\}$ and $\{w_j\}$ is A.

Exercise 4.2.11 Let V and W be vector spaces. Fix a basis $\{v_i\}_{i=1}^{m}$ of V and a basis $\{w_j\}_{j=1}^{n}$ of W. Then the map $\varphi\colon L(V, W) \to M_{n \times m}(\mathbb{R})$ given by $T \mapsto M_w^v(T)$ is a bijective (one-to-one and onto) linear map.

Theorem 4.2.2 *Let V, W, U be vector spaces. Let $\{v_1, \ldots, v_m\}$ be a basis for V, $\{w_1, \ldots, w_n\}$ be a basis for W and let $\{u_1, \ldots, u_k\}$ be a basis for U. Let $T\colon V \to W$, $S\colon W \to U$ be linear maps. Let A and B be the matrices associated with T and S respectively. Then the matrix associated with $S \circ T$ is AB.*

Proof We have $Tv_i = \sum_{j=1}^n a_{ij} w_j$ and $Sw_j = \sum_{r=1}^k b_{jr} u_r$. Hence

$$\begin{aligned} S \circ T(v_i) &= S\left(\sum_{j=1}^n a_{ij} w_j\right) \\ &= \sum_{j=1}^n a_{ij} S w_j \\ &= \sum_{j=1}^n a_{ij} \left(\sum_{r=1}^k b_{jr} u_r\right) \\ &= \sum_{r=1}^k \left(\sum_{j=1}^n a_{ij} b_{jr}\right) u_r \\ &= \sum_{r=1}^k c_{ir} u_r \end{aligned}$$

where $c_{ir} = \sum_{j=1}^n a_{ij} b_{jr}$, the irth entry of the product matrix AB.

□

Exercise 4.2.12 Let $A\colon V \to W$ be a linear map. Fix bases $\{v_1, \ldots, v_m\}$ of V and $\{w_1, \ldots, w_n\}$ of W. We denote by $M(A)$ the matrix of the linear map with respect to these bases. Find the matrix of A with respect to the following bases:

(1) The basis of V is $\{v_1', v_2', v_3', \ldots, v_n'\}$ where $v_1' = v_2$, $v_2' = v_1$ and $v_i' = v_i$ for $i \geq 3$ and the basis of W is as given.

(2) The basis of V remains as given but the basis of W is $\{w_j'\}$ where $w_i' = w_j$ and $w_j' = w_i$ and $w_k' = w_k$ if $k \neq i$ and $k \neq j$.

(3) The basis of V is $\{v_i'\}$ where $v_1' = v_1 + v_2$ and $v_i' = v_i$ for $i \geq 2$ and the basis of W is as given.

(4) The basis of V is $\{\alpha v_1, v_2, \ldots, v_m\}$, $\alpha \neq 0$ and the basis of W remains unchanged.

Exercise 4.2.13 Let the notation be as in Exercise 4.1.29. Fix bases $\{v_1, \ldots, v_n\}$ of V and $\{w_1, \ldots, w_n\}$ of W. Let f_i and g_j be the dual bases of V^* and W^* respectively (see Exercise 4.1.22). Is there any relation between the matrices $M_w^v(T)$ and $M_f^g(T^*)$?

4.3 Kernel and Image of a Linear Transformation

Definition 4.3.1 Let $T : V \to W$ be a linear map. Then the *kernel* of T, denoted by $\ker T$, is defined by

$$\ker T = \{v \in V \mid T(v) = 0\}$$

and the image of T, denoted by Im T, is defined by

$$\text{Im } T := T(V) = \{w \in W \mid \text{ there exists } v \in V \text{ such that } T(v) = w\}.$$

Exercise 4.3.1 $\ker T$ is a subspace of V and Im T is a subspace of W.

Definition 4.3.2 Let $T\colon V \to W$ be a linear map. The dimension of the vector subspace $\ker T$ (respectively the dimension of $\operatorname{Im} T$) is called the *nullity* (respectively the *rank*) of T. That is,

$$\begin{aligned} \text{nullity of } T &= \dim \ker T \\ \text{rank of } T &= \dim \operatorname{Im} T. \end{aligned}$$

Example 4.3.1 We wish to find the range and kernel of $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ x-y \end{pmatrix}.$$

We have $\ker(T) = \{(x, y) \in \mathbb{R}^2 \mid T(x, y) = (0, 0)\}$. Now $T(x, y) = (0, 0)$ implies that $x + y = 0$ and $x - y = 0$ which implies that $x = 0$, $y = 0$. Hence $\ker(T) = \{(0, 0)\}$.

Let $(x, y) \in \mathbb{R}^2$. Does there exist $(x_1, y_1) \in \mathbb{R}^2$ such that $T(x_1, y_1) = (x, y)$? That is, can we find $(x_1, y_1) \in \mathbb{R}^2$ such that $(x_1 + y_1, x_1 - y_1) = (x, y)$. To find such x_1, y_1, we need to solve the system of linear equations

$$\begin{aligned} x_1 + y_1 &= x \\ x_1 - y_1 &= y. \end{aligned}$$

On solving, we find that $x_1 = \dfrac{x+y}{2}$, and $y_1 = \dfrac{x-y}{2}$. Therefore given any $(x, y) \in \mathbb{R}^2$, we can find $(x_1, y_1) \in \mathbb{R}^2$ such that $T(x_1, y_1) = (x, y)$. Hence the range is $\mathbb{R}^2$.

✓ **Example 4.3.2** Find the range and kernel of $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+z \\ x+y+2z \\ 2x+y+3z \end{pmatrix}.$$

$$\ker(T) = \{ (x,y,z) \in \mathbb{R}^3 \mid T(x,y,z) = (0,0,0) \}.$$

Now $T(x,y,z) = (0,0,0)$ if and only if

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

Hence $z = -x$, $y = x$. Therefore,

$$\ker(T) = \{ (\alpha, \alpha, -\alpha) \mid \alpha \in \mathbb{R} \} = \{ \alpha(1,1,-1) \mid \alpha \in \mathbb{R} \}.$$

Let $(x,y,z) \in \mathbb{R}^3$. If $(x,y,z) \in \text{Im}\,(T)$, then there exists $(x_1,y_1,z_1) \in \mathbb{R}^3$ such that $T(x_1,y_1,z_1) = (x,y,z)$, or

$$(x_1+z_1, x_1+y_1+2z_1, 2x_1+y_1+3z_1) = (x,y,z).$$

Therefore $x_1+z_1 = x$ so that $z_1 = x - x_1$. Now, $x_1+y_1+2z_1 = y$ and hence $x_1+y_1+2(x-x_1) = y$. That is, $-x_1+y_1 = y-2x$. Finally, $2x_1+y_1+3z_1 = z$ and so $2x_1+y_1+3(x-x_1) = z$ or, $-x_1+y_1 = z-3x$. Therefore we get, $y-2x = z-3x$ and hence $y+x = z$. Thus if $(x,y,z) \in \text{Im}\,(T)$, $x+y = z$ and hence

$$\text{Im}\,(T) = \{(x,y,x+y)\} = \{ x(1,0,1) + y(0,1,1) \mid x,y \in \mathbb{R} \}.$$

Remark 4.3.1 Do you see that one of the recurring themes, the solution of a system of linear equations, is at work in these examples?

Exercise 4.3.2 Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ x+y \\ y \end{pmatrix}.$$

Find the range and kernel of T.

Exercise 4.3.3 Find the kernel and image of each of the linear maps of the exercises in Section 5.1.

Exercise 4.3.4 True or false? If V, W are vector spaces and $T : V \to W$ is a linear map and $\{v_1, \dots , v_n\}$ is a linearly independent set of vectors in V, then $\{T(v_i)\}_{i=1}^n$ is linearly independent.

Exercise 4.3.5 If $\{e_1, \dots , e_n\}$ is a basis of a vector space V, and $y_1, \dots , y_n$ are arbitrary elements of a vector space W, then we have already seen that there exists a unique linear map T such that $Te_i = y_i$ (see Theorem 4.1.3). If $\{v_1, \dots , v_n\}$ is an arbitrary set of vectors in V, and $\{w_1, \dots , w_n\}$ is an arbitrary set of vectors in W, does there exist a linear map $T : V \to W$ such that $Tv_i = w_i$?

Exercise 4.3.6 If W is a subspace of a vector space V, and $T : W \to X$ a linear transformation, does there exist a linear map $\tilde{T} : V \to X$ such that $\tilde{T}(w) = T(w)$ for all $w \in W$. If so, how many? *Hint:* Look first at $V = \mathbb{R}^2$, $W = \mathbb{R}e_1 = \{(x, 0) \mid x \in \mathbb{R}\}$ and $X = \mathbb{R}$ (recall Theorem 4.1.3).

Exercise 4.3.7 If V and W are vector spaces and $S, T : V \to W$ are linear transformations such that $\ker(T) = \ker(S)$ and $\mathrm{Im}(T) = \mathrm{Im}(S)$, is $S = T$?

Exercise 4.3.8 Let V and X be vector spaces. If W is a vector subspace of V, does there exist a transformation $T : V \to X$ such that $\ker(T) = W$? *Hint:* Theorem 4.1.3 and Theorem 2.3.8.

In the examples considered so far, we see that

$$\dim V = \dim \ker(T) + \dim\ \mathrm{Im}(T).$$

In fact, we have the following theorem:

Theorem 4.3.1 (Rank-Nullity Theorem) *Let V and W be finite dimensional vector spaces and $T : V \to W$ be a linear transformation. Then*

$$\dim V = \dim\ \mathrm{Im}(T) + \dim \ker(T) = \mathrm{rank}\, T + \mathrm{nullity}\, T. \tag{4.3.1}$$

Proof Let $\{u_1, \dots , u_k\}$ be a basis for $\ker T$. This can be extended to a basis of V, say, $\{u_1, \dots , u_k, v_1, \dots , v_n\}$. We now prove that $\{Tv_1, \dots , Tv_n\}$ is a basis of $\mathrm{Im}(T)$. Let $w \in \mathrm{Im}(T)$. Then there exists $v \in V$ such that $T(v) = w$. Now $v = \sum_{i=1}^k \alpha_i u_i + \sum_{i=1}^n \beta_i v_i$. Therefore,

$$\begin{aligned} w = T(v) &= T\left(\sum \alpha_i u_i + \sum \beta_j v_j\right) \\ &= \sum \alpha_i T(u_i) + \sum \beta_j T(v_j) \\ &= \sum \beta_j T(v_j), \qquad \text{since } u_i \in \ker(T). \end{aligned}$$

This implies that $\{Tv_1, \dots , Tv_n\}$ spans $\mathrm{Im}(T)$.

We claim that $\{Tv_i\}$ is linearly independent. Let $\sum_{i=1}^{n} \alpha_i T v_i = 0$ for some scalars α_i. Then $T(\sum_{i=1}^{n} \alpha_i v_i) = 0$. This means that

$$\sum_{i=1}^{n} \alpha_i v_i \in \ker(T).$$

But then $\sum_{i=1}^{n} \alpha_i v_i = \sum_{i=1}^{k} \beta_j u_j$ since $\{u_1, \ldots, u_k\}$ is a basis of $\ker(T)$. Therefore,

$$\sum_{i=1}^{n} \alpha_i v_i - \sum_{i=1}^{k} \beta_i u_i = 0.$$

Since $\{u_1, \ldots, u_k, v_1, \ldots, v_n\}$ is a basis of V, we see that $\alpha_i = 0$ and $\beta_j = 0$ for all i and j. Hence the claim follows.

Thus $\{Tv_i\}_{i=1}^{n}$ is linearly independent and spans $\mathrm{Im}\,(T)$ so that dimension of $\mathrm{Im}(T)$ is n. Now, $\dim \mathrm{Im}(T) = n = n + k - k = \dim V - \dim \ker(T)$. Hence, we have $\dim V = \dim \ker(T) + \dim \mathrm{Im}\,(T)$.

□

Example 4.3.3 Let us do Example 4.3.2 again. We have found that $\dim \ker T = 1$. Proceeding as in Example 4.3.2, we find that if (x, y, z) is in $\mathrm{Im}\,(T)$ then we have to solve for the system

$$\begin{aligned} x_1 + z_1 &= x \\ x_1 + y_1 + 2z_1 &= y \\ 2x_1 + y_1 + 3z_1 &= z. \end{aligned}$$

Subtracting the second equation from the third, we get $x_1 + z_1 = z - y$. This along with the first equation implies that $x = z - y$ or $z = x + y$. Thus, the $\mathrm{Im}\,(T) \subset W := \{(x, y, z) \in \mathbb{R}^3 \mid z = x + y\}$. This is a vector subspace of dimension 2. Now Equation (4.3.1) implies that $\dim \mathrm{Im}\,(T) = 2$: $\dim \mathbb{R}^3 = \dim \ker(T) + \dim \mathrm{Im}\,(T) = 1 + \dim \mathrm{Im}\,(T)$. Thus $\mathrm{Im}\,(T)$ is a two-dimensional vector subspace of the two-dimensional vector space W and hence $\mathrm{Im}\,(T) = W$.

This is an example of how one uses theory to cut down excessive computations, shorten the arguments and gain insight.

Exercise 4.3.9 Let $V = \mathbb{R}^n$ and A be a $n \times n$ matrix. If $Ax = 0$ has a unique solution then $Ax = b$ has a unique solution for every $b \in \mathbb{R}^n$.

Exercise 4.3.10 Can you construct a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^4$ such that $\mathrm{Im}(T) = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1 + x_2 + x_3 + x_4 = 0\}$?

Exercise 4.3.11 Can you construct a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ such that $\mathrm{Im}(T) = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\}$?

Exercise 4.3.12 Let $T : V \to V$ be a linear map such that $\operatorname{Im} T = \ker T$. What can you say about T^2? (By the way, can you construct such a map from $\mathbb{R}^2$ to $\mathbb{R}^2$?)

4.4 Linear Isomorphism

Definition 4.4.1 Let V and W be vector spaces over $\mathbb{R}$. A linear map $T: V \to W$ is said to be a (linear) *isomorphism* if T is one-one and onto. We then say V is *isomorphic* to W.

Exercise 4.4.1 If $T: V \to W$ is an isomorphism, then the set theoretic inverse $T^{-1}: W \to V$ is linear and an isomorphism.

Exercise 4.4.2 Isomorphism is an equivalence relation. This means the following:

(i) V is isomorphic to itself.

(ii) If V is isomorphic to W, then W is isomorphic to V.

(iii) If V is isomorphic to W, W is isomorphic to U, then V is isomorphic to U.

Before we say why this concept is important, let us look at some examples.

Example 4.4.1 Let V be the space of polynomials (with real coefficients) of degree less than or equal to n and W be $\mathbb{R}^{n+1}$. Then the map

$$T: P := \sum_{i=1}^{n} \alpha_i X^i \mapsto (\alpha_0, \dots, \alpha_n) \in \mathbb{R}^{n+1}$$

is an isomorphism.

Example 4.4.2 Let V be any vector space over $\mathbb{R}$. Let $T: V \to V$ be defined by $T(x) = \alpha x$, $\alpha \neq 0$. Then T is a linear isomorphism.

Example 4.4.3 Let $V = M_{n\times m}(\mathbb{R})$ be the set of $n \times m$ matrices with real entries and let $W = \mathbb{R}^{mn}$. Define $f: V \to W$ by

$$f(A) = (\alpha_{11}, \dots, \alpha_{1m}, \alpha_{21}, \dots, \alpha_{2m}, \dots, \alpha_{n1}, \dots, \alpha_{nm}).$$

Here $A = (\alpha_{ij}) \in V$. It is easily seen that f is an isomorphism.

Example 4.4.4 This is a slightly more abstract isomorphism, worth learning thoroughly. Let V and W be n-dimensional vector spaces over $\mathbb{R}$. Let $\{v_i\}_{i=1}^n$, (respectively $\{w_i\}_{i=1}^n$) be a basis of V (respectively W) over $\mathbb{R}$. Given $x \in V$, we can then write $x = \sum \alpha_i v_i$. We let $f(x) = \sum \alpha_i w_i$. That is, $f(v_i) = w_i$ and extend this linearly over $\mathbb{R}$. Then f is an isomorphism of V onto W. We leave the proof of this assertion to the reader.

We invite the reader to check that all the isomorphisms in the previous examples were obtained this way. In Example 4.4.1, $\{P_i = X^i\}_{i=0}^n$ is a basis of V. T maps X^i to $e_i \in \mathbb{R}^{n+1}$ and extended linearly to all of V. If $P \in V$, $P = \sum \alpha_i X^i$, then

$$T(P) = \sum \alpha_i f(X^i) = \sum \alpha_i e_i = (\alpha_0, \alpha_1, \dots, \alpha_n) \in \mathbb{R}^{n+1}.$$

We leave the other cases as exercises to the reader. The theorem below says that any isomorphism between V and W "arises" this way.

Theorem 4.4.1 *Let $T: V \to W$ be a linear map. Then T is an isomorphism if and only if $\{T(v_i)\}_{i=1}^n$ is a basis of W for any basis $\{v_i\}_{i=1}^n$ of V.*

Proof Let $\{v_i\}_{i=1}^n$ be a basis of V and let T be a (linear) isomorphism. We need to prove that $\{T(v_i)\}_{i=1}^n$ is a basis of W.

We first of all show that $\{T(v_i)\}_{i=1}^n$ is linearly independent. If

$$\sum \alpha_i T(v_i) = 0$$

for some scalars $\alpha_i \in \mathbb{R}$, then by linearity of T, we have

$$\sum \alpha_i T(v_i) = T\Big(\sum \alpha_i v_i\Big) = 0.$$

Since T is one-one and $T(0) = 0$, we see that $\sum \alpha_i v_i = 0$. But $\{v_i\}_{i=1}^n$ is a basis of V and hence $\sum \alpha_i v_i = 0$ if and only if $\alpha_i = 0$ for all i. Thus $\{T(v_i)\}_{i=1}^n$ is a linearly independent set.

To prove that $\{T(v_i)\}_{i=1}^n$ is a basis, it is now enough to show that $\{T(v_i)\}_{i=1}^n$ spans W. Since $T: V \to W$ is onto, given $w \in W$, there exists $v \in V$ such that $T(v) = w$. Write $v = \sum \alpha_i v_i$. Then by linearity,

$$w = T(v) = T\Big(\sum \alpha_i v_i\Big) = \sum \alpha_i T(v_i).$$

This implies w is a linear combination of $T(v_i)$'s. Thus $\{T(v_i)\}_{i=1}^n$ is a basis of W.

The converse is essentially Example 4.4.4 and hence left as an exercise.

□

Exercise 4.4.3 Isomorphic vector spaces have the same dimension.

Exercise 4.4.4 Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(x, y) = (ax + by, cx + dy)$. Then T is an isomorphism if and only if $ad - bc \neq 0$. Write explicitly T^{-1} when it exists. What are the matrices of T and T^{-1} (whenever the latter exists) with respect to standard basis of $\mathbb{R}^2$?

The importance of the concept of linear isomorphism is as follows. Let $f\colon V \to W$ be an isomorphism. Let us assume that we want to solve the vector equation: Find $x_j \in V$ such that the equation $\sum_j \alpha_{ij} x_j = \beta_i$, where $\alpha_{ij} \in \mathbb{R}$, $\beta_i \in V$, is satisfied. We can solve this in V, that is, find x_j, if and only if we can solve the equations $\sum \alpha_{ij} y_j = f(\beta_i)$ in W. It may happen that this second system is readily solvable. Then we take $x_j = f^{-1}(y_j)$ and these x_j's solve the original system.

Exercise 4.4.5 Let $T\colon V \to W$ be linear. Then T is one-one if and only if $\ker(T) = 0$. *Hint:* If $Tx = Ty$, then $T(x - y) = 0$.

We include a proof of the following theorem for completeness sake. However, we urge the reader to write out a proof on his/her own.

Theorem 4.4.2 *Let V be a finite dimensional vector space. Let $T : V \to V$ be a linear map. Then the following are equivalent:*

(1) *T is an isomorphism.*

(2) $\ker T = \{0\}$.

(3) $\operatorname{Im}(T) = V$.

Proof Assume that (1) holds. Since $T(0) = 0$ for any linear map, if $v \in V$ is such that $T(v) = 0$, then $v = 0$ since T is one-one.

Assume that (2) holds. To show that $\operatorname{Im} T = V$. By the rank-nullity theorem (Equation (4.3.1)), we have

$$\dim \operatorname{Im} T = \dim V - \dim \ker T = \dim V - 0 = \dim V.$$

Since $\operatorname{Im} T$ is a vector subspace of V, by Exercise 2.3.14, $\operatorname{Im} T = V$. Thus (3) follows.

Assume that (3) holds, that is, T is onto. Hence $\dim \operatorname{Im} T = \dim V$. To prove that T is an isomorphism, it is therefore enough to show that it is one-one. If T is not one-one, then there exist $v_i \in V$ such that $v_1 \neq v_2$ and $T(v_1) = T(v_2)$. This implies that $T(v_1 - v_2) = 0$, by linearity of T. Thus a nonzero element $v_1 - v_2 \in \ker T$ so that $\dim \ker T \geq 1$. Using the rank-nullity theorem Equation (4.3.1), we get

$$\dim \operatorname{Im} T = \dim V - \dim \ker T < \dim V.$$

This contradicts our assumption that $\dim \operatorname{Im} T = \dim V$. So, we conclude that T is one-one.

□

Remark 4.4.1 The reader should compare this result with the following: Let X be a finite set. Then a map $f\colon X \to X$ is a bijection if and only if it is one-one if and only if it is onto. In fact, the reader can supply a proof of Theorem 4.4.2 using this fact and Theorem 4.4.1. In light of the fact that such a result is false in the case of infinite sets, he may want to investigate the validity of the theorem for infinite dimensional vector spaces.

Exercise 4.4.6 Any n-dimensional vector space is isomorphic to $\mathbb{R}^n$ and then any two n-dimensional vector spaces are isomorphic.

Exercise 4.4.7 Let T be the map as in Example 2.1.3. Then T is an isomorphism.

Exercise 4.4.8 The map in Exercise 4.2.11 is an isomorphism. Hence conclude that $\dim L(V, W) = \dim(V) \times \dim(W)$.

Exercise 4.4.9 Let V denote the space in Exercise 2.1.22. Let $T\colon V \to \mathbb{R}^n$ be defined by

$$Tf = (f(a), f'(a), \dots , f^{n-1}(a)).$$

Then T is a linear isomorphism. *Hint:* You need results from the theory of linear systems of ordinary differential equations.

Exercise 4.4.10 Let V be a vector space and W a subspace of V. Then there exists an onto linear map $T\colon V \to V/W$ such that $W = \ker T$.

Exercise 4.4.11 Let V and U be vector spaces and $T : V \to U$ be a linear map with kernel K. Let $W = \operatorname{Im} T$. Then $W \simeq V/K$. (This is called the *Fundamental Theorem of Homomorphisms.*)

Exercise 4.4.12 If W is a subspace of V, every subspace of V/W is of the form T/W where T is a subspace of V containing W.

Exercise 4.4.13 This is an extension of Exercise 4.2.12. Let $\{v_1, \dots , v_m\}$ and $\{v'_1, \dots , v'_m\}$ be two bases of V and $\{w_1, \dots , w_n\}$ and $\{w'_1, \dots , w'_n\}$ be two bases of W. Let $M^v_w(A)$ (respectively, $M^{v'}_{w'}(A)$) denote matrix of a linear transformation $A\colon V \to W$ with respect to the bases $\{v_1, \dots , v_m\}$ and $\{w_1, \dots , w_m\}$ (respectively $\{v'_1, \dots , v'_m\}$ and $\{w'_1, \dots , w'_n\}$). Find the relation between A^w_v and $A^{w'}_{v'}$.

4.5 Geometric Ideas and Some Loose Ends

Given an $n \times m$ matrix $A = (a_{ij})$, one may think of it as a listing of (column) vectors from $\mathbb{R}^m$. For, we may write $A = (C_1, \ldots, C_m)$ where

$$C_j = \begin{pmatrix} a_{ij} \\ \vdots \\ a_{nj} \end{pmatrix}$$

is the jth column. In particular, we think of A as the linear map which takes the jth vector in the standard basis of $\mathbb{R}^m$ to $C_j \in \mathbb{R}^n$: $Ae_j = C_j$, $1 \leq i \leq m$. We find this way of looking at matrices quite geometric and useful on many occasions.

The choice of a basis for a vector space V allows us to set up an isomorphism from V to $\mathbb{R}^n$, where $n = \dim V$. Let $\{v_1, \ldots, v_n\}$ be a basis of V. Write $v = \sum_{i=1}^n a_i v_i$. Then the map $T \colon V \to \mathbb{R}^n$ defined by

$$Tv = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

is a linear isomorphism. (Isn't this Exercise 4.4.6?)

Let V and W be vector spaces, $T \colon V \to W$ a linear map. Let $\{v_i\}_{i=1}^m$ be a basis of V and $\{w_j\}_{j=1}^n$ a basis of W. We want to look at the matrix of T with respect to these bases in a geometric way.

If $A = (a_{ij})$ is the matrix representation of T with respect to these bases, then the column vector

$$\xi_j = \begin{pmatrix} a_{j1} \\ \vdots \\ a_{jn} \end{pmatrix}$$

stands for the vector $y_j = \sum a_{ji} w_i \in W$.

We shall put these ideas into use to prove Theorem 4.5.1.

Let $A = (a_{ij})$ be an $m \times n$ matrix. We write

$$A = \begin{pmatrix} R_1 \\ \vdots \\ R_m \end{pmatrix}$$

where $R_i := (a_{i1}, \ldots, a_{in})$ is the ith row. Similarly, we write

$$A = (C_1, \ldots, C_n) \quad \text{where} \quad C_j = \begin{pmatrix} a_{ij} \\ \vdots \\ a_{mj} \end{pmatrix}$$

is the jth column.

We may consider R_i (respectively, C_j) as a row vector in $\mathbb{R}^n$ (respectively as column vector in $\mathbb{R}^m$). The vector subspace spanned by R_i's (respectively C_j's) is called the *row space* (respectively *column space*) of A.

The *row rank* of A is defined to be the dimension of the vector subspace of $\mathbb{R}^n$ spanned by R_i's. The *column rank* of A is defined to be the dimension of the vector subspace of $\mathbb{R}^m$ spanned by C_j's.

If we think of A as the linear map which takes the jth element of the standard basis of $\mathbb{R}^n$ to the jth column C_j, then the column space is nothing other than Im (A). Hence the column rank of A is the dimension of Im (A).

Exercise 4.5.1 Compute the row and column rank S of

$$1.\ \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix} \quad 2.\ \begin{pmatrix} 1 & 0 & 2 \\ 0 & 2 & 2 \\ 0 & 0 & 0 \end{pmatrix} \quad 3.\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The main result of this section is the following theorem:

Theorem 4.5.1 *The row rank and the column rank of an $m \times n$ matrix $A = (\alpha_{ij})$ are equal.*

Proof Let k be the row rank of A. Let $\{v_1, \dots, v_k\}$ be a basis of the row space of A. We then can write $R_i = \sum_{j=1}^{k} \alpha_{ij} v_j$. Let $v_r = (b_{r1}, \dots, b_{rn})$ for $1 \le r \le k$. Since

$$R_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in}) = \sum_{r=1}^{k} \alpha_{ir} v_r = \sum_{r=1}^{k} \alpha_{ir}(b_{r1}, \dots, b_{rn}),$$

we get an equation among the coordinates

$$\alpha_{ij} = \sum_{r=1}^{k} \alpha_{ir} b_{rj} \text{ for } 1 \le i \le m \text{ and } 1 \le j \le n.$$

That is,

$$\begin{aligned} \alpha_{1j} &= \alpha_{11}b_{1j} + \alpha_{12}b_{2j} + \cdots + \alpha_{1k}b_{kj} \\ \alpha_{2j} &= \alpha_{21}b_{1j} + \alpha_{22}b_{2j} + \cdots + \alpha_{2k}b_{kj} \\ \vdots \quad & \qquad \vdots \\ \alpha_{mj} &= \alpha_{m1}b_{1j} + \alpha_{m2}b_{2j} + \cdots + \alpha_{mk}b_{kj}. \end{aligned}$$

Hence, we get, for $1 \le j \le n$,

$$\begin{pmatrix} \alpha_{1j} \\ \alpha_{2j} \\ \vdots \\ \alpha_{mj} \end{pmatrix} = b_{1j} \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{m1} \end{pmatrix} + \cdots + b_{kj} \begin{pmatrix} \alpha_{1k} \\ \alpha_{2k} \\ \vdots \\ \alpha_{mk} \end{pmatrix}$$

But the left side of the above equation is C_j. Thus any column is a linear combination of k vectors. Hence the dimension of the column space is at most k. Thus the column rank is less than or equal to the row rank.

A similar argument yields the reverse inequality. Hence the result.

□

Exercise 4.5.2 Complete the proof of Theorem 4.5.1.

Definition 4.5.1 The common value of the row and column ranks of A is called the *rank* of A and denoted by rank A.

Exercise 4.5.3 For an $m \times n$ matrix, what is the largest possible value of rank A?

Exercise 4.5.4 If A is an 11×7 matrix, show that the rows of A are linearly dependent.

Exercise 4.5.5 If A is a 3×5 matrix, show that the columns of A are linearly dependent.

Definition 4.5.2 We say an $n \times n$ matrix A is *non-singular* if rank $A = n$.

Exercise 4.5.6 Show that A is non-singular if and only if A is invertible. That is, there exists a matrix B such that $AB = BA = I$. *Hint*: There is a linear map corresponding to A. Rank $A = n$ says something about the linear map.

Exercise 4.5.7 Let A be an $m \times n$ matrix and B an $n \times k$ matrix. Show that the rank of AB is at most the minimum of the ranks of A and B. Can it be strictly less than the minimum? *Hint:* Look at the linear map $T: \mathbb{R}^2 \to \mathbb{R}^2$ which maps e_1 to e_2 and e_2 to zero. Let A be the matrix of this linear map.

Exercise 4.5.8 Let the notation be as in Exercise 4.5.7 True or false? If A is of rank r and B is of maximal rank, that is, the rank of A is $\min\{n, k\}$, then the rank of AB is the rank of A.

4.6 Some Special Linear Transformations

In this section, we show how a geometric object transforms under some special linear transformations. The idea behind this exercise is that the reader will learn how to look at linear maps as geometric maps.

In the following series of examples, we show how the geometric picture "E" in $\mathbb{R}^2$ transforms under certain linear maps of $\mathbb{R}^2$. The letter E is considered as the subset, some of whose special points are the origin, e_1, $(0, 1/2)$, $(1, 1/2)$, e_2 and $e_1 + e_2$ (see Example 1 on page 80).

We explain by means of an example how the picture of $T(E)$ is drawn where

$$T = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}$$

is as in Example 5 below. First note that we can write

$$T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = AB \qquad \text{say.}$$

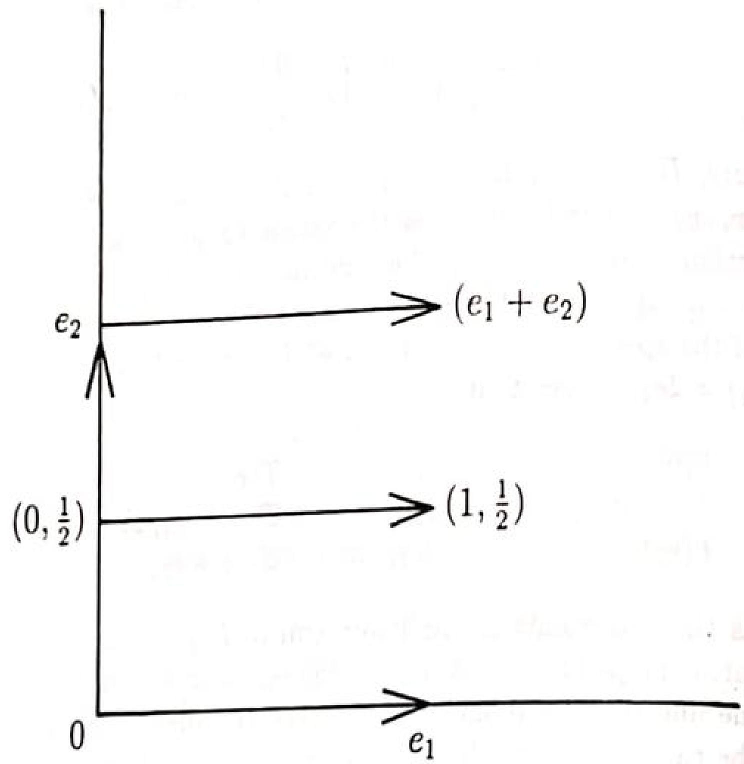
Clearly, B sends e_1 to e_1 and e_2 to $2e_2$, or more generally, sends (x, y) to $(x, 2y)$. Thus it stretches the vector (x, y) by a factor of 2 in the y-direction. Obviously, A is the reflection with respect to the y-axis. Thus T is a composition of these two maps. Now what are the images (under T) of the special points listed above? Using the fact that $T(e_1) = -e_1$ and $T(e_2) = 2e_2$, we see that

$$\begin{array}{lcllcl} T(0) & = & 0, & T(e_1) & = & -e_1, \\ T(0e_1 + e_2/2) & = & e_2, & T(e_1 + e_2/2) & = & -e_1 + e_2, \\ T(e_2) & = & 2e_2, \text{ and} & T(e_1 + e_2) & = & -e_1 + 2e_2. \end{array}$$

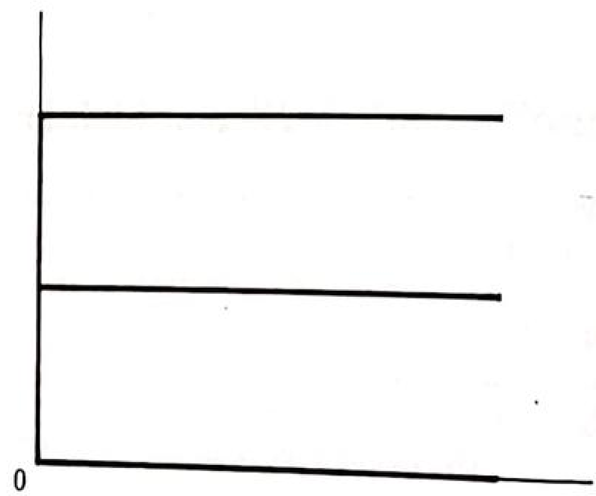
Thus the end points of the lower arm of E go to 0 and $-e_1$. Since T is linear, it maps the points on the line segment joining 0 and e_1 into points of the line segment 0 and $-e_1$. (In fact, if p divides the line segment $[x, y]$ in the ratio $t : 1 - t$, (that is, $p = tx + (1 - t)y$, then Tp divides the "line segment" $[Tx, Ty]$ in the ratio $t : 1 - t$. It is possible that Tx and Ty are scalar multiples of each other and hence the reason for the quotation marks). Thus under T the lower arm goes to the lower arm of TE as shown in the figure. Proceeding in a similar way, the reader can show that the image TE is as shown in the figure in Example 5.

The reader must convince himself of the validity of the figures of the other examples, given on pages 80–85, in an analogous manner.

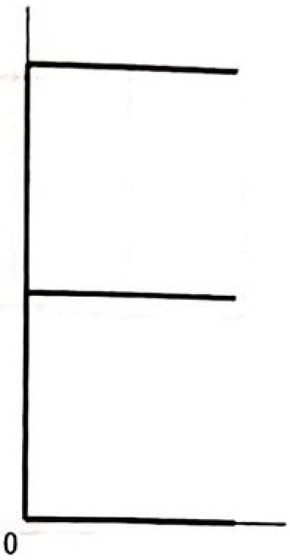
1. *Identity:*
$$A_1 = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



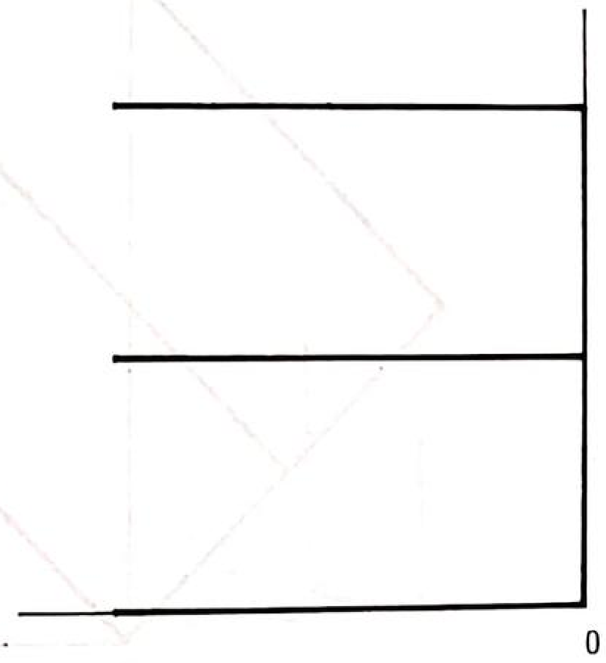
2. *Stretching along e_1 (x-direction):*
$$A_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$



3. *Stretching along e_2 (y-direction):* $A_3 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$

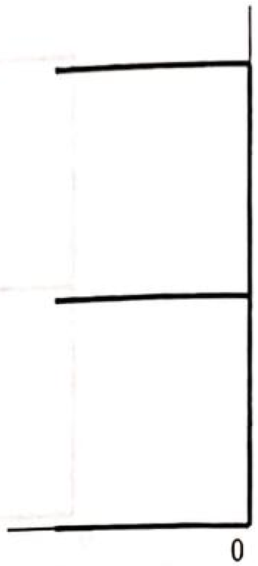


4. *Reflection with respect to y-axis:*
$A_4 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$



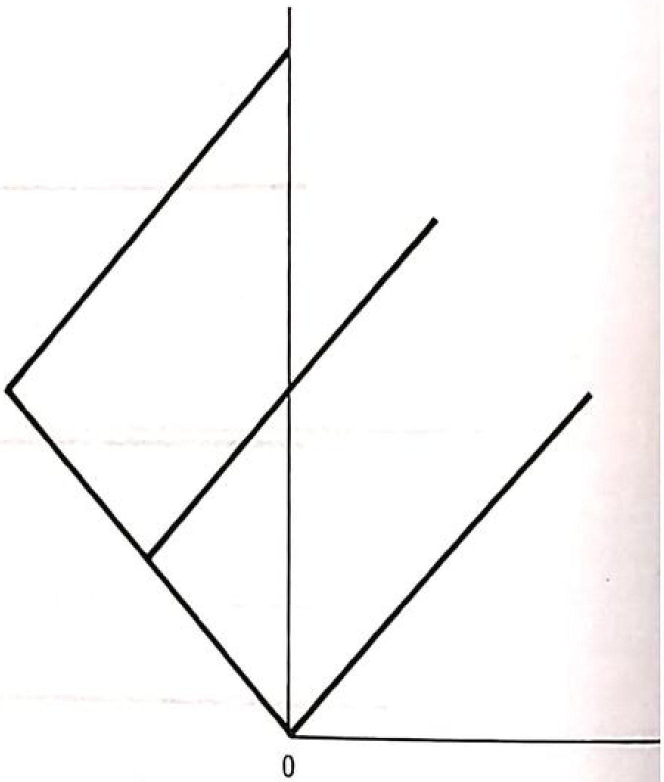
5. *Stretching along y-direction and reflection with respect to y-a*

$$A_5 = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}$$



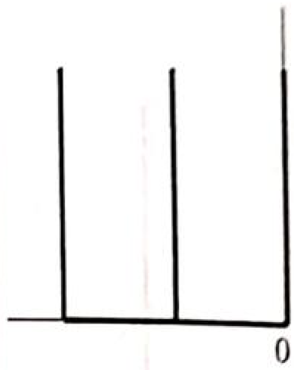
6. *Rotation by an angle θ:*

$$A_6 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$



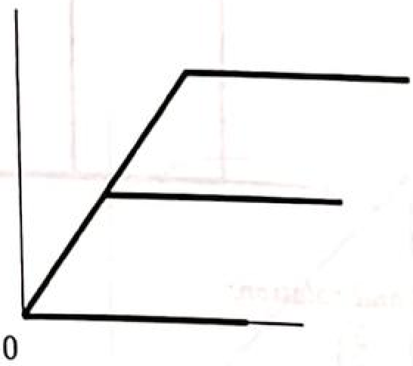
7. *Rotation by $\theta = \pi/2$:*

$$A_7 = \begin{pmatrix} \cos\frac{\pi}{2} & -\sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{pmatrix}$$



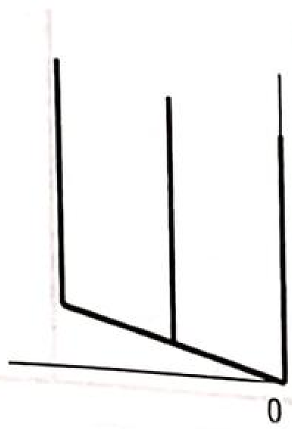
8. *Shear:*

$$A_8 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$



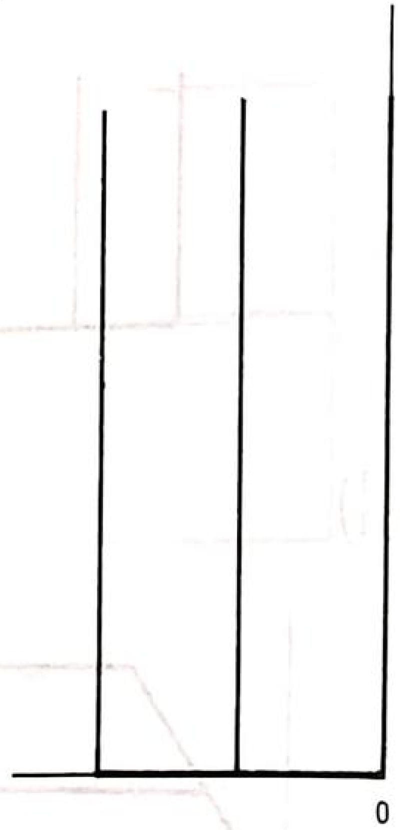
9. *Shear and rotation:*

$$A_9 = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$$



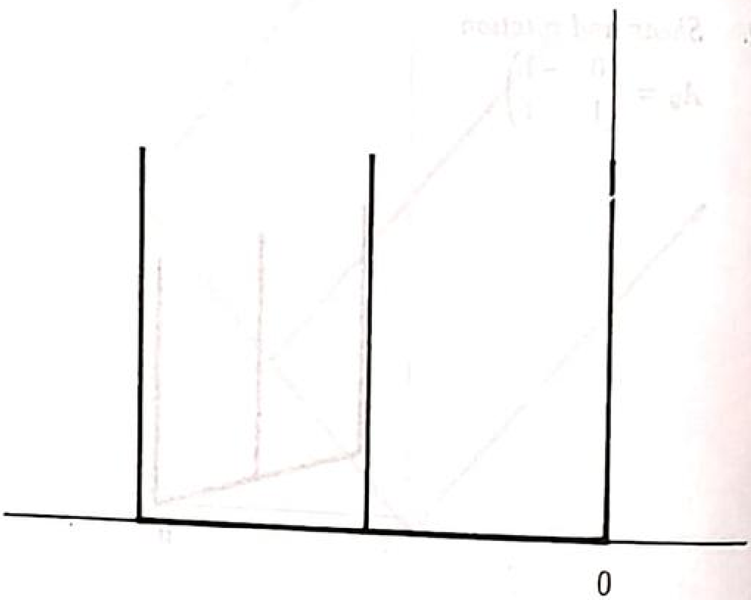
10. *Rotation and stretching:*

$$A_{10} = \begin{pmatrix} 0 & -1 \\ 2 & 0 \end{pmatrix}$$



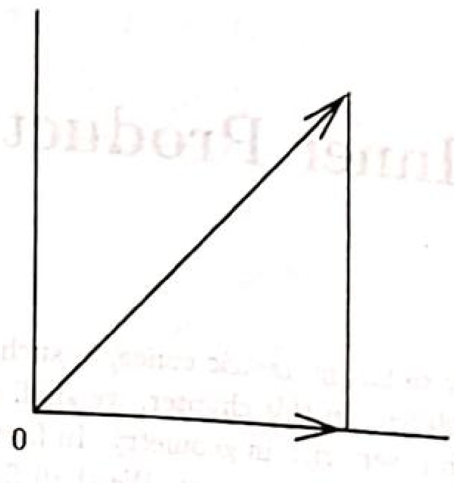
11. *Stretching and rotation:*

$$A_{11} = \begin{pmatrix} 0 & -2 \\ 1 & 0 \end{pmatrix}$$



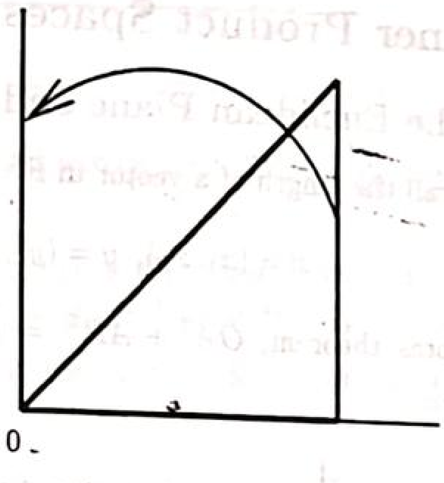
12. *Projection:*

$$A_{12} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$



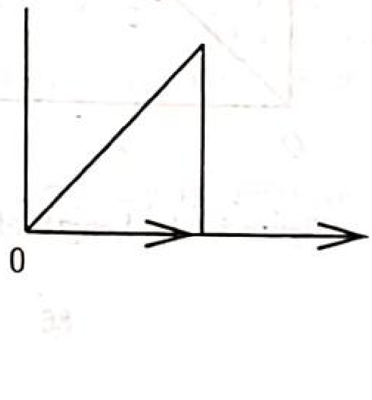
13. *Projection and rotation:*

$$A_{13} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$



14. *Projection and stretching:*

$$A_{14} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$



Scanned by CamScanner

5. Inner Product Spaces

In our study so far, no metric concepts such as length, angle and dista
were encountered. In this chapter, we shall study a special class of vec
spaces which is very rich in geometry. In fact, a model of Euclidean Geo
etry is provided by these spaces. We shall first look into the most famil
of these, namely the Euclidean plane. This will allow us to fine-tune
geometric insight.

5.1 Inner Product Spaces

5.1.1 The Euclidean Plane and the Dot Product

We shall recall the length of a vector in $\mathbb{R}^2$. Let

$$x = (x_1, x_2),\ y = (y_1, y_2) \in \mathbb{R}^2.$$

By Pythagoras theorem, $OA^2 + AP^2 = OP^2$ or $x_1^2 + x_2^2 = OP^2$ (
Figure 5.1.1).

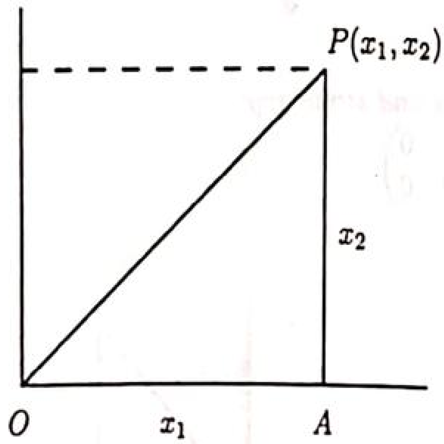


Figure 5.1.1 Length of a vector in $\mathbb{R}^2$.

We define the *length* or *norm* of x as the positive $\sqrt{x_1^2 + x_2^2}$ and denote it by $||x||$.

Again, by Pythagoras theorem (see Figure 5.1.2), we have

$$||x - y||^2 = (x_1 - y_1)^2 + (x_2 - y_2)^2.$$

Hence $||x - y|| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$. Let $d(x, y) = ||x - y||$. This gives the *distance* between the points x and y.

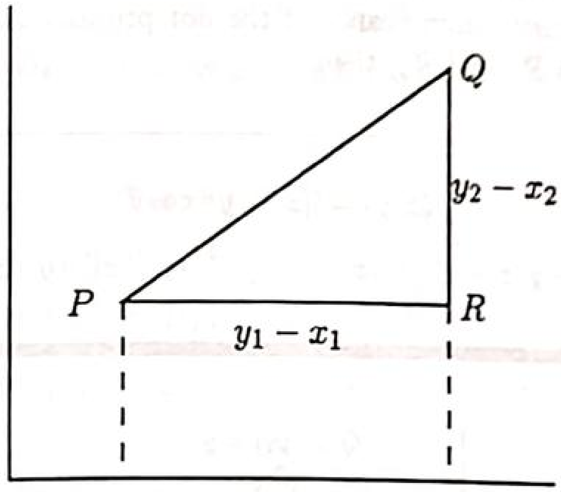


Figure 5.1.2 Distance between x and y.

All these concepts can be captured by an additional structure, called the *dot product* on $\mathbb{R}^2$.

Definition 5.1.1 Let $x = (x_1, x_2)$, $y = (y_1, y_2)$. Then the dot product of x and y, denoted by $\langle x, y\rangle$ (or, $x \cdot y$) is defined as $\langle x, y\rangle = x_1y_1 + x_2y_2$.

Note that we get the earlier notions such as length and distance from the dot product:

(a) If $x = y$, then $\langle x, x\rangle = x_1^2 + x_2^2 = ||x||^2$. Thus $||x|| = \langle x, x\rangle^{\frac{1}{2}}$.

(b) $d(x, y) = ||x - y|| = \langle x - y, x - y\rangle^{\frac{1}{2}} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$.

Now that we know that this new notion captures the old metric concepts, we may wonder what its geometric meaning is.

Let $P = (x_1, x_2) = x$ and $Q = (y_1, y_2) = y$. From plane trigonometry we know that given two sides of a triangle and the included angle between them, the remaining side can be calculated using the *law of cosines*. From Figure 5.1.3, we have

$$PQ^2 = OP^2 + OQ^2 - 2OP \cdot OQ \cos\theta.$$

By the definition of the length of a vector and the dot product, we have

$$(x_1 - y_1)^2 + (x_2 - y_2)^2 = ||x||^2 + ||y||^2 - 2\,||x||\,||y||\cos\theta.$$

Simplifying, we get

$$(x_1^2+x_2^2)+(y_1^2+y_2^2)-2(x_1y_1+x_2y_2) = (x_1^2+x_2^2)+(y_1^2+y_2^2)-2\,||x||\,||y||\cos\theta$$

or $||x||\,||y||\cos\theta = x_1y_1 + x_2y_2 = \langle x, y\rangle$ and so $\langle x, y\rangle = ||x||\,||y||\cos\theta$.

Thus the geometric significance of the dot product is: If θ is the angle between the lines $\mathbb{R}x$ and $\mathbb{R}y$, then

$$\langle x, y\rangle = ||x||\,||y||\cos\theta$$
$$\langle x - y, x - y\rangle = ||x||^2 + ||y||^2 - 2\,||x||\,||y||\cos\theta.$$

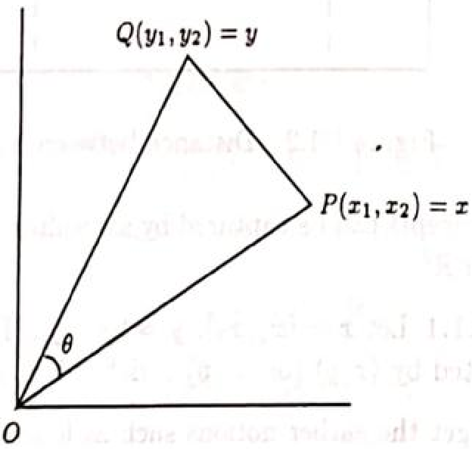


Figure 5.1.3 Angle between vectors.

We now introduce a dot product on $\mathbb{R}^n$ in a way similar to that on $\mathbb{R}^2$.

Definition 5.1.2 If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are in $\mathbb{R}^n$, then their dot product $\langle x, y\rangle$ (or, $x \cdot y$) is defined as $\langle x, y\rangle = \sum_{i=1}^n x_iy_i$.

We ask the reader to verify that the dot product has the following properties:

(1) $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$, if and only if $x = 0$.

(2) $\langle x, y\rangle = \langle y, x\rangle$, for all $x, y \in \mathbb{R}^n$.

(3) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$, for all $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.

(4) $\langle x + z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$, for all $x, y, z \in \mathbb{R}^n$.

(2) and (3) imply that $\langle x, \beta y\rangle = \beta\langle x, y\rangle$ for $x, y \in \mathbb{R}^n$, $\beta \in \mathbb{R}$. (2) and (4) imply that $\langle x, y+z\rangle = \langle x, y\rangle + \langle x, z\rangle$ for all $x, y, z \in \mathbb{R}^n$.

Definition 5.1.3 If $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ then the *length* or *norm* of the vector x is denoted by $||x||$ and given by

$$||x|| = \langle x, x\rangle = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}.$$

For $x, y \in \mathbb{R}^n$, the *distance* between x and y is defined as

$$d(x, y) = ||x - y|| = \sqrt{\langle x - y, x - y\rangle} = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{\frac{1}{2}}.$$

5.1.2 General Inner Product Spaces

We abstract the properties of the dot product on $\mathbb{R}^n$ in the following definition.

Definition 5.1.4 An *inner product* or a *dot product* on a vector space V is a map $\langle,\rangle : V \times V \to \mathbb{R}$ satisfying the following properties: For $x, y, z \in V$ and $\alpha \in \mathbb{R}$,

(i) $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ if and only if $x = 0$,

(ii) $\langle x, y\rangle = \langle y, x\rangle$,

(iii) $\langle x + z, y\rangle = \langle x, y\rangle + \langle z, y\rangle$ and $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$,

(iv) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$.

$(V, \langle,\rangle)$ is called an *inner product space*. For brevity sake, we may say V is an inner product space without explicitly mentioning the inner product $\langle\, ,\, \rangle$.

Example 5.1.1 The dot product defined above on $\mathbb{R}^n$ is an inner product.

Convention. Unless specified otherwise the inner product on $\mathbb{R}^n$ will be assumed to be the dot product.

Example 5.1.2 For $x, y \in \mathbb{R}^2$, where $x = (x_1, x_2)$, $y = (y_1, y_2)$,

$$\langle x, y\rangle = y_1(x_1 + 2x_2) + y_2(2x_1 + 5x_2)$$

defines an inner product. We shall prove that if $\langle x, x \rangle = 0$ then $x = 0$. We have

$$\begin{aligned} \langle x, x \rangle &= x_1^2 + 4x_1x_2 + 5x_2^2 \\ &= x_1^2 + 4x_1x_2 + 4x_2^2 + x_2^2 \\ &= (x_1 + 2x_2)^2 + x_2^2. \end{aligned}$$

Thus $\langle x, x \rangle = 0$ if and only if $(x_1 + 2x_2)^2 = 0$ and $x_2^2 = 0$. This is true if and only if $x_1 = 0 = x_2$, that is, if and only if $x = 0$. The rest of the verifications are left as an easy exercise to the reader.

Exercise 5.1.1 Let $V = \mathbb{C}$. Show that $\langle z, w \rangle := \operatorname{Re}(z\overline{w})$ defines an inner product on $\mathbb{C}$.

Exercise 5.1.2 Let $V = \mathbb{R}^2$ and define $\langle x, y \rangle := y_1(2x_1 + x_2) + y_2(x_1 + x_2)$. Show that this defines an inner product on $\mathbb{R}^2$.

Exercise 5.1.3 If $f, g \in C[0,1]$ define $\langle f, g \rangle = \int_0^1 f(t)g(t)\,dt$. Note first of all that the integral exists (thanks to analysis!). Here also, the crucial thing to show is that $\langle f, f \rangle = 0$ if and only if $f = 0$. This follows from Exercise 5.1.4 from Analysis. The rest of the properties follow from well-known properties of the integral. Thus $(C[0,1], \langle , \rangle)$ is an inner product space.

Exercise 5.1.4 Let $f\colon [0,1] \to \mathbb{R}$ be continuous with $f(t) \geq 0$ for $t \in [0,1]$. Then $\int_0^1 f(t)\,dt = 0$ if and only if $f(t) = 0$ for all $t \in [0,1]$. *Hint:* To prove the nontrivial part, assume that $\int_0^1 f = 0$. If f is not identically 0, let t_0 be such that $f(t_0) > 0$. Let $\alpha := f(t_0)$ and $\varepsilon := \alpha/2$. For $\varepsilon/2$, by continuity of f at t_0, there is a δ such that $f(t) \in (\varepsilon/2, \frac{3\varepsilon}{2})$ for $t \in (t_0 - \delta, t_0 + \delta)$. Using various properties of the integral, we see that

$$\int_0^1 f(t)\,dt \geq \int_{t_0-\delta}^{t_0+\delta} f(t)\,dt \geq \int_{t_0-\delta}^{t_0+\delta} \frac{\varepsilon}{2}\,dt = \varepsilon\delta > 0.$$

This contradicts our assumption that $\int_0^1 f(t)\,dt = 0$.

Exercise 5.1.5 Let $V = M(n, \mathbb{R})$. Define $\langle A, B \rangle := \operatorname{tr} AB^t$. Show that this defines an inner product on $M(n, \mathbb{R})$ ($tr(X) := \sum_i x_{ii}$ if $X = (x_{ij})$).

To gain facility with computations involving inner products, do the next exercise. Think of distribution of multiplication over addition.

Exercise 5.1.6 Let V be an inner product space. Then

$$\langle ax + by, cv + dw \rangle = ac\,\langle x, v \rangle + ad\,\langle x, w \rangle + bc\,\langle y, v \rangle + bd\,\langle y, w \rangle .$$

What are $\langle x + y, x + y \rangle$, $\langle x + y, x - y \rangle$?

Exercise 5.1.7 Fix $a \in V$, an inner product space. Show that the maps from V to $\mathbb{R}$ given by $x \mapsto \langle x, a \rangle$ and $y \mapsto \langle a, y \rangle$ are linear.

Exercise 5.1.8 Let $\{v_1, \ldots, v_n\}$ be a (not necessarily the standard) basis of $\mathbb{R}^n$. Let $\alpha_i \in \mathbb{R}$ be given for $1 \leq i \leq n$. Show that there exists a unique vector $x \in \mathbb{R}^n$ such that $x \cdot v_i = \alpha_i$ for all i. *Hint:* Think of a linear map!

Definition 5.1.5 Let V be an inner product space. We can imitate the definition of length or norm defined as earlier. The *length* or *norm* of a vector $v \in V$ is $||v|| := \langle v, v \rangle^{\frac{1}{2}}$, the positive square root of the non-negative number $\langle v, v \rangle$.

If the reader feels uncomfortable with abstract inner product space, he may assume that the inner product space is $\mathbb{R}^n$ with the dot product introduced above. However, he should notice that nowhere (except in some examples) we shall have to use the way the inner product is defined. We shall use only the defining properties of the inner product.

Lemma 5.1.1 *Let V be an inner product space. The norm function*

$$|| \; || : V \to \mathbb{R}$$

has the following properties:

(1) $||x|| \geq 0$ *and* $||x|| = 0$ *if and only if* $x = 0$, *for* $x \in V$.

(2) $||\alpha x|| = |\alpha| \, ||x||$, $x \in V$ *and* $\alpha \in \mathbb{R}$.

Furthermore, given a nonzero vector $v \in V$, there is a vector $u \in V$ such that $||u|| = 1$ and $v = ||v|| u$. This u is called the unit vector *along v.*

Proof (1) is easy. We shall prove (2).
It suffices to show that $||\alpha x||^2 = |\alpha|^2 \, ||x||^2$.

$$||\alpha x||^2 = \langle \alpha x, \alpha x \rangle = \alpha \langle x, \alpha x \rangle = \alpha \langle \alpha x, x \rangle = \alpha^2 \langle x, x \rangle .$$

(Can you justify the steps above?) To prove the last assertion, observe that the equation $v = ||v|| \, u$ suggests that we take $u = \frac{v}{||v||}$. u is of unit length because of (2): $||u|| = \left|\left| \frac{v}{||v||} \right|\right| = \frac{1}{||v||} ||v|| = 1$.

□

A vector u in an inner product space V is said to be of *unit norm* or *unit length* if $||u|| = 1$, that is, if and only if $\langle u, u \rangle = 1$. The above construction of a unit vector along a given nonzero vector is used quite often in the sequel.

Theorem 5.1.2 (Cauchy-Schwarz Inequality) *Let V be an inner product space. If $x, y \in V$, then $|\langle x, y\rangle| \leq ||x||\,||y||$. Further, equality holds if and only if one is a multiple of the other (that is, x and y are linearly dependent).*

Remark 5.1.1 In the case of $\mathbb{R}^2$ with the dot product, Cauchy-Schwarz inequality is obvious. For, since $\langle x, y\rangle = ||x||\,||y|| \cos\theta$, we have

$$\begin{aligned} |\langle x, y\rangle| &= |\,||x||\;||y||\,\cos\theta\,| \\ &\leq ||x||\,||y||\;|\cos\theta| \\ &\leq ||x||\;||y||. \end{aligned}$$

Proof We shall give three proofs — one is elementary, geometric. The second uses calculus and the third uses results on quadratic equations. Also, we indicate a proof of Cauchy-Schwarz inequality in the special case of $\mathbb{R}^n$ with the dot product. The last one is to convince you how the level of abstractions help us understand the underlying principles.

Proof 1. If $x = 0$ or $y = 0$, then $\langle x, y\rangle = 0$ and either $\langle x, x\rangle = 0$ or $\langle y, y\rangle = 0$. Hence the result. Now consider the case when $||x|| = ||y|| = 1$. Consider $\langle x - y, x - y\rangle$. Then

$$\begin{aligned} 0 \leq \langle x - y, x - y\rangle &= \langle x, x\rangle + \langle y, y\rangle - 2\langle x, y\rangle \\ &= 2 - 2\langle x, y\rangle \quad \text{as } ||x|| = ||y|| = 1. \end{aligned}$$

This gives $\langle x, y\rangle \leq 1$.

Similarly $\langle x + y, x + y\rangle \geq 0$ yields $-\langle x, y\rangle \leq 1$. Hence

$$|\langle x, y\rangle| \leq 1 = ||x||\,||y||. \tag{5.1.1}$$

We now prove the statement concerning the equality. Let $|\langle x, y\rangle| = 1$. Then either $\langle x, y\rangle = 1$ or -1. If $\langle x, y\rangle = 1$, from the above chain of inequalities we deduce that $\langle x - y, x - y\rangle = 0$ or $x = y$. If $\langle x, y\rangle = -1$, we see that $x = -y$. Thus equality holds if and only if either $x + y = 0$ or $x - y = 0$, that is, if and only if $x = \pm y$.

Now suppose x and y are nonzero (not necessarily of unit length). Then $u = \frac{x}{||x||}$ and $v = \frac{y}{||y||}$ are of unit length (by Lemma 5.1.1). By the previous case $|\langle u, v\rangle| \leq 1$. Therefore,

$$\left|\left\langle \frac{x}{||x||}, \frac{y}{||y||}\right\rangle\right| = \left|\frac{1}{||x||}\frac{1}{||y||}\langle x, y\rangle\right| \leq 1.$$

From this we get $|\langle x, y\rangle| \leq ||x||\,||y||$.

If x and y are nonzero, then the equality means $\langle x, y\rangle = ||x||\,||y||$ or $-\langle x, y\rangle = ||x||\,||y||$. Assume the first happens.

Then

$$\begin{aligned}
\langle x, y\rangle &= \|x\|\,\|y\| \\
&\iff \left\langle \frac{x}{\|x\|}, \frac{y}{\|y\|} \right\rangle = 1 \\
&\iff \frac{x}{\|x\|} = \frac{y}{\|y\|} \\
&\iff x = \frac{\|x\|}{\|y\|} y.
\end{aligned}$$

The other case is similar.

□

Proof 2. Fix x and y in V. If $y = 0$, then the result is obviously true. So, we assume that $y \neq 0$. Consider the real valued function of the real variable $f(t) := \langle x + ty, x + ty\rangle$. We want to investigate the extremum points of f:

$$\begin{aligned}
f(t) &= \langle x + ty, x + ty\rangle \\
&= \langle x, x\rangle + 2t\langle x, y\rangle + t^2 \langle y, y\rangle .
\end{aligned}$$

Thus $f(t)$ is a polynomial in t with real coefficients.

$$f'(t) = 2\langle x, y\rangle + 2t\langle y, y\rangle .$$

Then if t_0 is an extremum point for f only if $f'(t_0) = 0$, that is, only if $\langle x, y\rangle + t_0 \langle y, y\rangle = 0$. This suggests that we choose $t_0 = -\frac{\langle x,y\rangle}{\langle y,y\rangle}$. Is this point an extremum point? Now $f''(t) = 2\langle y, y\rangle > 0$, since $y \neq 0$, for all t, in particular for $t = t_0$. Hence $f(t_0)$ is a minimum. That is, $0 \leq f(t_0) \leq f(t)$ for all t. But $f(t_0) \geq 0$ since $f(t) \geq 0$ for all t. That is,

$$\langle x, x\rangle - 2\frac{(\langle x, y\rangle)^2}{\langle y, y\rangle} + \frac{(\langle x, y\rangle)^2}{\langle y, y\rangle} \geq 0.$$

It follows that

$$\langle x, x\rangle \geq \frac{\langle x, y\rangle^2}{\langle y, y\rangle} \text{ or } |\langle x, y\rangle| \leq \langle x, x\rangle^{\frac{1}{2}} \langle y, y\rangle^{\frac{1}{2}}.$$

The equality case is again dealt with by carefully retracing the above chain of inequalities.

□

Proof 3. Let $p(t) := at^2 + bt + c$ be a quadratic polynomial in t with real coefficients. Recall that $p(t)$ is always nonnegative (or always nonpositive) if and only if it has no real roots or a double real root. This happens if and only if the discriminant $b^2 - 4ac < 0$ or equal to 0. Now $f(t)$ as in the second proof is a quadratic polynomial in t with real coefficients $a = \langle y, y \rangle$, $b = 2\langle x, y \rangle$ and $c = \langle x, x \rangle$. Also, $f(t)$ is always nonnegative. So we conclude that $b^2 - 4ac \leq 0$. From this the required result follows.

□

Remark 5.1.2 Let us consider $\mathbb{R}^n$ with the dot product. The Cauchy-Schwarz inequality in this case reads as follows:

$$\left|\sum_{i=1}^{n} x_i y_i\right| \leq \left(\sum_i x_i^2\right)^{\frac{1}{2}} \left(\sum_i y_i^2\right)^{\frac{1}{2}}, \quad \text{for all } x_i, y_j \in \mathbb{R}. \tag{5.1.2}$$

This concrete inequality is quite useful in analysis. Note that the inequality is a special case of what we proved above. Another more classical proof of this follows from the Lagrange's identity

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 = \left(\sum x_i^2\right)\left(\sum y_i^2\right) - \sum_{1 \leq i < j \leq n} (x_i y_j - x_j y_i)^2.$$

Exercise 5.1.9 Let a_i, $1 \leq i \leq n$ be positive real. Let $\alpha \in \mathbb{R}$. Show that

$$\left(\sum_{i=1}^{n} \sqrt{a_i}\right)^2 \leq \left(\sum_{i=1}^{n} a_i^{\alpha}\right)\left(\sum_{i=1}^{n} a_i^{1-\alpha}\right),$$

with equality if and only if either $\alpha = 1/2$ or $\alpha \neq 1/2$ but all the a_i's are equal. *Hint:* Let $x_i := a_i^{\alpha/2}$ and $y_i := a_i^{(1-\alpha)/2}$ in Equation (5.1.2).

Exercise 5.1.10 What does the Cauchy-Schwarz inequality mean for V in Exercise 5.1.5?

Corollary 5.1.3 *Let V be an inner product space. The norm function $|| \; || : V \to \mathbb{R}$ has the following properties:*

(1) $||x|| = 0$ *if and only if* $x = 0$, *for* $x \in V$.

(2) $||\alpha x|| = |\alpha| \, ||x||$ *for* $x \in V$ *and* $\alpha \in \mathbb{R}$.

(3) $||x + y|| \leq ||x|| + ||y||$, *for* $x, y \in V$. *This is known as the* Triangle Inequality.

(4) $|\, ||x|| - ||y|| \,| \leq ||x - y||$ *for all* $x, y \in V$.

Proof (1) and (2) were proved ea·lier. To prove the triangle inequality, we compute

$$\begin{aligned} ||x+y||^2 &:= \langle x+y, x+y\rangle \\ &= \langle x,x\rangle + 2\langle x,y\rangle + \langle y,y\rangle \\ &\leq ||x||^2 + 2\,||x||\,||y|| + ||y||^2 \qquad \text{by Cauchy-Schwarz} \\ &= (||x|| + ||y||)^2. \end{aligned}$$

The triangle inequality follows.

Observe $||x|| = ||(x-y)+y|| \leq ||x-y|| + ||y||$ by triangle inequality. It follows that $||x|| - ||y|| \leq ||x-y||$. Interchanging x and y in this inequality, we get

$$||y|| - ||x|| \leq ||y-x|| = ||(-1)(x-y)|| = |-1|\,||x-y|| = ||x-y||.$$

Thus, $\pm(||x|| - ||y||) \leq ||x-y||$. Hence the last assertion follows.

□

Definition 5.1.6 A *metric* on a set X is a function $d\colon X \times X \to \mathbb{R}$ with the following properties:

(i) $d(x,y) \geq 0$ for $x, y \in X$ and $d(x,y) = 0$ if and only if $x = y$.

(ii) $d(x,y) = d(y,x)$ for all $x, y \in X$.

(iii) $d(x,z) \leq d(x,y) + d(y,z)$ for all $x, y, z \in X$ (Triangle inequality).

Proposition 5.1.4 *Let V be an inner product space. If we define*

$$d(x,y) := ||x-y|| \ \textit{for}\ x, y \in V,$$

then d is a metric on V.

Proof We shall show that d satisfies the triangle inequality. As is to be expected, we use triangle inequality for the norm. Let $x, y, z \in V$. Then

$$\begin{aligned} d(x,z) &= ||x-z|| \\ &= ||(x-y)+(y-z)|| \\ &\leq ||x-y|| + ||y-z|| = d(x,y) + d(y,z). \end{aligned}$$

The rest of the proof is easy and left to the reader.

□

Exercise 5.1.11 Show that "distance" is translation invariant. That is,

$$d(x+z, y+z) = d(x,y) \text{ for all } x, y, z \in V.$$

In the notation of Exercise 4.1.2, this says:

$$d(T_z(x), T_z(y)) = d(x,y) \text{ for all } x, y, z \in V.$$

5.2 Orthogonality

Unless specified otherwise, V will stand for an inner product space.

In Section 5.1 we saw how the existence of an inner product on a vector space induces notions such as the length of a vector and the distance between two vectors. In this section, we shall see how to define the angle between two nonzero vectors of an inner product space.

The remarks made after Definition 5.1.1, motivate the following definition of the angle between two nonzero vectors x and y.

Definition 5.2.1 If x and y are two nonzero vectors in an inner product space V, then by Cauchy-Schwarz inequality, we have

$$-1 \leq \frac{\langle x, y\rangle}{||x||\,||y||} \leq 1.$$

From trigonometry (or more rigorously, from analysis), it follows that there exists a unique $\theta \in [0, \pi]$ such that $\cos\theta = \frac{\langle x, y\rangle}{|x||y|}$. This θ is called the *angle* between the nonzero vectors x and y.

Exercise 5.2.1 Compute the angle between

(1) $v = e_1$ and $w = e_1 + e_2$ in $\mathbb{R}^2$,

(2) $v = (x, y)$ and $(-y, x)$ $x \neq 0 \neq y$ in $\mathbb{R}^2$, and

(3) $(x_1, \ldots, x_{2k-1}, x_{2k})$ and $(-x_2, x_1, -x_4, x_3, \ldots, -x_{2k}, x_{2k-1})$ in $\mathbb{R}^{2k}$.

Exercise 5.2.2 Find $\cos\theta$ where θ is the angle between the vectors $f(t) = t$ and $g(t) = t^2$ in Exercise 5.1.3.

Exercise 5.2.3 Let V be as in Exercise 5.1.3. Let $f(t) = t$. Let $h(t) = t^2$. Compute $g := h - 3\langle h, f\rangle f$. What is the angle between f and g? (If you are intrigued by this, the mystery behind this construction will be solved in a later section.)

Definition 5.2.2 Let x and y be vectors in an inner product space V. We say x and y are *orthogonal* if $\langle x, y\rangle = 0$. This definition is meaningful since $\cos(\pi/2) = 0$. Also, it coincides with what we have in $\mathbb{R}^2$ with the dot product: $\langle x, y\rangle = 0$ implies $\cos\theta = 0$ which implies $\theta = \frac{\pi}{2}$. We write $x \perp y$ to denote $\langle x, y\rangle = 0$.

Exercise 5.2.4 Let $x, y, z \in V$. Let $x \perp y$ and $x \perp z$. Then $x \perp (\alpha y + \beta z)$ for all $\alpha, \beta \in \mathbb{R}$. More generally, the set $\{v \in V \mid \langle x, v\rangle = 0\}$ is a vector subspace of V. It is denoted by $x^\perp$.

Assume that $v = (\alpha, \beta) \neq 0$ in $\mathbb{R}^2$. What is the geometric description of $v^\perp$? $v^\perp$ is given by $\{(x, y) \in \mathbb{R}^2 \mid \alpha x + \beta y = 0\}$. Thus $v^\perp$ is a straight line passing through the origin perpendicular to the vector (α, β). Proceeding similarly, we see that if $v = (\alpha, \beta, \gamma) \in \mathbb{R}^3$ is nonzero, then

$$v^\perp = \{(x, y, z) \in \mathbb{R}^3 \mid \alpha x + \beta y + \gamma z = 0\}.$$

Thus $v^\perp$ is the plane through the origin with normal $v = (\alpha, \beta, \gamma)$.

Example 5.2.1 This generalizes the last observation. We wish to find the dimension of $W = \{v \in V \mid \langle v, x \rangle = 0\}$ for a fixed nonzero $x \in V$.

Define $T_\alpha : V \to \mathbb{R}$ by $T_\alpha(v) = \langle x, \alpha \rangle$. Then T_α is a linear map (Exercise 5.1.7) and $W = \ker(T_\alpha)$. Since $\alpha \neq 0$, $\mathrm{Im}\,(T_\alpha) = \mathbb{R}$. For, $T_\alpha(\alpha) = \langle \alpha, \alpha \rangle \neq 0$ lies in the vector subspace $\mathrm{Im}\,(T_\alpha) \subset \mathbb{R}$. Therefore $\dim \mathrm{Im}\,(T_\alpha) \geq 1$ and hence equals 1. Now, by Equation (4.3.1), we have

$$\dim V = \dim \ker(T_\alpha) + \dim \mathrm{Im}\,(T)$$

which implies that $\dim W = \dim V - 1$. If $\alpha = 0$, then $W = V$.

Exercise 5.2.5 Let v and w be nonzero vectors in V with $v \perp w$. Let $\alpha, \beta \in \mathbb{R}$ be such that $\alpha v + \beta w = 0$. Then $\alpha = 0 = \beta$. That is, two nonzero vectors orthogonal to each other are linearly independent.

Exercise 5.2.6 If a vector $z \in V$ is orthogonal to all the vectors in V, then $z = 0$. Consequently, if $\langle x, v \rangle = \langle y, v \rangle$ for all $v \in V$, then $x = y$.

This simple exercise is quite often used to show the equality of two vectors in an inner product space.

Exercise 5.2.7 Let W_i, $i = 1, 2$, be vector subspaces of V. Assume that each vector in one of them is orthogonal to all of the other. Show that $W_1 \cap W_2 = \{0\}$.

Exercise 5.2.8 Let S be any nonempty subset of V. Let

$$S^\perp := \{v \in V \mid \langle v, s \rangle = 0, \text{ for all } s \in S\}.$$

Then $S^\perp$ is a vector subspace. Can you think of two proofs — one direct and the other using Exercise 2.2.27?

Exercise 5.2.9 Let $v = (\alpha, \beta) \in \mathbb{R}^2$ be nonzero. Describe $v^\perp$ as $\mathbb{R}w$ for a suitable w.

Exercise 5.2.10 Let v and w be two nonzero vectors in $\mathbb{R}^3$. Assume that the set of vectors orthogonal to both of them is a plane (through the origin). Then each is a scalar multiple of the other. (Do you see this geometrically?)

Exercise 5.2.11 Let $v = (\alpha, \beta, \gamma)$ be a nonzero vector in $\mathbb{R}^3$. Find a basis of $W := v^\perp$. *Hint:* W is described by a linear equation and you have learnt Section 1.1!

Exercise 5.2.12 This is a continuation of Exercise 5.2.11. Give a pair of equations whose solution set is the line joining the origin and v.

Lemma 5.2.1 (Pythagoras Theorem) *Let $x, y \in V$ be orthogonal to each other. Then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.*

More generally, let $\{v_i\}_{i=1}^k$ be a set of vectors such that they are pairwise orthogonal, that is, $v_i \perp v_j$ if $i \neq j$. Then

$$\left\|\sum_{i=1}^k v_i\right\|^2 = \sum_{i=1}^k \|v_i\|^2.$$

Proof Do you see why this is called the Pythagoras theorem?

A simple computation yields the result:

$$\begin{aligned}\left\|\sum_{i=1}^k v_i\right\|^2 &= \left\langle \sum_{i=1}^k v_i, \sum_{j=1}^k v_j \right\rangle \\ &= \sum_{1\leq i,j\leq k} \langle v_i, v_j\rangle \\ &= \sum_{i=1}^k \langle v_i, v_i\rangle \qquad \text{as } \langle v_i, v_j\rangle = 0 \text{ for } i \neq j.\end{aligned}$$

Did you notice how the sums were indexed at the top right?

□

Exercise 5.2.13 $\|x+y\| = \|x\| + \|y\|$ if and only if one is a nonnegative scalar multiple of the other.

Exercise 5.2.14 For any $x, y \in V$, we have

$$4\langle x, y\rangle = \|x+y\|^2 - \|x-y\|^2.$$

This is known as *polarization identity.*

Exercise 5.2.15 Prove that for any two vectors $x, y \in V$,

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

Geometrically this means that the sum of the squares of the diagonals equals the sum of the squares of the sides of a parallelogram.

Exercise 5.2.16 Prove that $||x|| = ||y||$ if and only if $x - y \perp x + y$. (The geometric meaning of this is that a parallelogram is a rhombus if and only if the diagonals are perpendicular.)

Exercise 5.2.17 Prove that x and y are orthogonal if and only if

$$||x - y||^2 = ||x||^2 + ||y||^2.$$

(This is Pythagoras theorem and its converse.)

5.3 Some Geometric Applications

We mentioned at the beginning of this chapter that $\mathbb{R}^2$ with dot product is a model for Euclidean geometry. In this section, we give indications for this assertion by proving some results from geometry using the notions developed so far. This also serves to instill a geometric way of looking at linear algebra. To start with, let us examine the last few exercises at the end of Section 5.2.

Example 5.3.1 *The sum of the squares of the diagonals of a parallelogram equals the sum of the squares of its sides.*

To prove this we need to turn this geometric problem into the language of linear algebra. Without loss of generality, let us assume that three of the vertices are the vectors 0, x and y. Draw a picture. Then the vertex of the fourth side is $x + y$. (Recall the geometric interpretation of the vector addition.) The length of the side whose end points are 0 and x (respectively y) is $||x||$ (respectively $||y||$). So we know the lengths of sides. One of the diagonals has its endpoints 0 and $x + y$ and so its length is $||x + y||$. The other diagonal has endpoints x and y and so its length is $||x - y||$. Therefore, what we are asked to show is that

$$2(||x||^2 + ||y||^2) = ||x + y||^2 + ||x - y||^2.$$

This is easy: We write $||x + y||^2 = \langle x + y, x + y\rangle$ and expand the right side. We do similarly for $||x - y||^2$. Add the results to get what we want.

$$\begin{aligned}
\langle x + y, x + y\rangle &= \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle \\
\langle x - y, x - y\rangle &= \langle x, x\rangle - 2\langle x, y\rangle + \langle y, y\rangle .
\end{aligned}$$

The sum of the right sides in these equations is $2(\langle x, x\rangle + \langle y, y\rangle)$.

We shall be brief in the rest of this section.

Example 5.3.2 *A parallelogram is a rhombus if and only if the diagonals are perpendicular to each other.*

We use the notation as in Example 5.3.1. The direction vector of the diagonal joining 0 and $x+y$ may be taken as $x+y$. The direction vector of the diagonal joining the points x and y may be taken as $x-y$. (Recall direction vectors of $\ell(x,y)$ are of the form $t(x-y)$ for $t \in \mathbb{R}$.) The diagonals are perpendicular if and only if their direction vectors are orthogonal, that is, if and only if $\langle x-y, x+y \rangle = 0$. Thus, we are asked to show that

$$||x|| = ||y|| \text{ if and only if } \langle x+y, x-y \rangle = 0.$$

We invite the reader to verify this.

Example 5.3.3 *A parallelogram is a rectangle if and only if the diagonals are of equal length.*

With the notation as above, what we are supposed to show is that $||x+y|| = ||x-y||$ if and only if the angle between the sides $0x$ (the line segment joining 0 and x) and $0y$ (the line segment joining 0 and y) is $\pi/2$, that is, if and only if cosine of this angle is zero. This is translated in our language as $||x+y|| = ||x-y||$ if and only if $\langle x, y \rangle = 0$. In an inner product space, it is easier to work with the inner product than the norm. So what we would like to establish is

$$\langle x+y, x+y \rangle = \langle x-y, x-y \rangle \text{ if and only if } \langle x, y \rangle = 0.$$

Now this is an easy exercise for the reader.

We now turn our attention to the study of triangles. The first example is the Pythagoras theorem.

Example 5.3.4 *A triangle is right angled if and only if there exists one side whose square equals the sum of the squares of the other two sides.*

How do we find a model for this in linear algebra? Recall that a triangle is a triple of three non-collinear points, that is, they do not lie on a line. These three points are considered as the vertices of the triangle. To find a model for this, we may assume that the vertices are at 0, x and y. The condition for their non-collinearity turns out to be their linear independence. Do you see this? The line joining 0 and x (respectively y) is $\mathbb{R}x$ (respectively $\mathbb{R}y$). The point x (respectively y) lies on $\mathbb{R}y$ (respectively $\mathbb{R}x$) if and only if there are $a, b \in \mathbb{R}$ such that $ax = by$, that is, if and only if x and y are linearly independent. Thus a triangle with a vertex at the origin corresponds to a triple $(0, x, y)$ of points with $\{x, y\}$ linearly independent. The lengths of the sides of the triangle are, therefore, $||x||$, $||y||$ and $||x-y||$. (Refer to Example 5.3.1.)

To simplify the matters, we make a further assumption. If the given triangle has the Pythagorean property $c^2 = a^2 + b^2$, we may assume that the vertex opposite to the longest side c is at 0. Thus what we are supposed to prove is $||x - y||^2 = ||x||^2 + ||y||^2$ if and only if $\langle x, y\rangle = 0$. As usual, the reader proves this.

Example 5.3.5 *If a triangle is isosceles, then the medians to the two sides of equal length are of equal length.*

We keep the notation of Example 5.3.4. We assume that the sides $0x$ and $0y$ are of equal length. This means that $||x|| = ||y||$. The midpoints of the line sides are $\frac{1}{2}x$ and $\frac{1}{2}y$. The medians, under consideration, are the line segments joining (i) x and $y/2$ and (ii) y and $x/2$. Their lengths are $||x - y/2||$ and $||y - x/2||$. What we have to prove is: $||x|| = ||y||$ implies $||x - y/2|| = ||y - x/2||$. We ask the reader to check this. (Remember: It is always better to use inner product than the norm.)

Exercise 5.3.1 Prove the converse of Example 5.3.5.

We now look at a result which belongs to Affine Geometry, that part of geometry which deals with points, lines, planes and their incidence (inclusion) relations and which does not deal with metric concepts such as lengths and angles. First a definition: Given a line segment

$$[x, y] := \{tx + (1 - t)y \mid 0 \le t \le 1\}$$

in a vector space, the point $z := (x + y)/2$ is called the midpoint of the line segment. (If V happens to be an inner product space, then this coincides with our metric requirement: z divides the line segment into two parts of equal length. The point here is that we can define the midpoint of a line segment in any vector space.) The result that we want to prove is taken up in Example 5.3.6.

Example 5.3.6 *The medians of a triangle are concurrent.*

As earlier, we take the vertices at 0, x and y with $\{x, y\}$ linearly independent. The midpoints are $(x + y)/2$, $x/2$ and $y/2$. The lines joining the vertices with the midpoints of the opposite sides are: $r \cdot (x + y)/2$, $sx + (1 - s)(y/2)$ and $ty + (1 - t)(x/2)$. Let us find the point of intersection of the last two lines. Finding the point of intersection is equivalent to solving for s and t in the equation: $sx + (1 - s)(y/2) = ty + (1 - t)(x/2)$. This is rewritten as

$$\left(s - \frac{1 - t}{2}\right) x - \left(\frac{1 - s}{2} - t\right) y = 0.$$

Since x and y are linearly independent, we deduce that $s = \frac{1-t}{2}$ and $t = \frac{1-s}{2}$. It follows that $s = t = 1/3$. Thus the point of intersection is $(x+y)/3$. This point certainly lies on the first line. Thus the lines are concurrent at $(x+y)/3$.

Exercise 5.3.2 Redo Example 5.3.6 without assuming that one of the vertices is at the origin 0. This will give you more symmetric expressions.

The next result that we want to prove is about circles:

Example 5.3.7 *The angle inscribed by semicircle is a right angle.*

We assume that the circle has centre at 0 and radius r in $\mathbb{R}^2$. The diametrically opposite points are given by x and $-x$ for some x with $||x|| = r$. Let y be any point on the circle. Then $||y|| = r$. The angle inscribed is the angle between the lines $\ell(-x, y)$ and $\ell(x, y)$ at y. Their direction vectors are $x + y$ and $x - y$. Thus, we are expected to show that $||x|| = ||y||$ if and only if $\langle x + y, x - y\rangle = 0$. You should have no difficulty in proving this result.

In all the above computations, you might have noticed that whenever we needed to deal with two vectors, we assumed that we were in the two-dimensional space spanned by them. This assumption allows us to see what happens in $\mathbb{R}^2$ and get geometric insight. This point is worth remembering.

With this we end our excursion into geometry.

5.4 Orthogonal Projection onto a Line

This forms the heart of the next few sections.

Let $v \in \mathbb{R}^n$ be any vector. Let u be any unit vector, that is, $||u|| = 1$. We want to find whether there is a vector u_0 in the set $\mathbb{R}u$ which is nearest to v. That is, we are looking for a real number t_0 such that $d(v, t_0u) \leq d(v, tu)$ for all $t \in \mathbb{R}$. We can solve this problem in two ways. One is geometric and the other is analytic in the sense that it uses one variable calculus.

Look at Figure 5.4.1. Let θ be the angle between u and v. The vector $P_u(v)$ is called the *orthogonal projection* of v on u. From Figure 5.4.1, we see that

$$P_u(v) = ||v|| \cos\theta u = ||v|| \frac{\langle u, v\rangle}{||u||\,||v||} \cdot u = \langle u, v\rangle\, u$$

since $||u|| = 1$. Hence we have $P_u(v) = \langle u, v\rangle\, u$. This suggests the following definition.

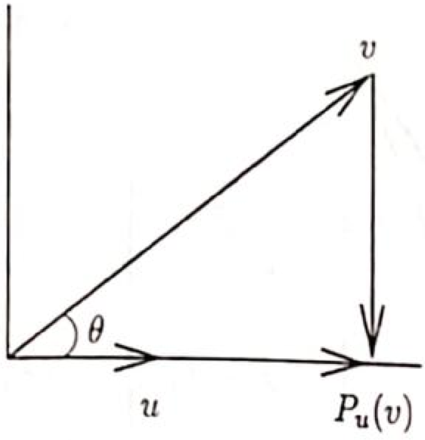


Figure 5.4.1 Orthogonal projection.

Definition 5.4.1 Let V be an inner product space. Let u be a unit vector and $v \in V$ arbitrary. Then the projection $P_u(v)$ of v onto the line (one-dimensional subspace) $\mathbb{R}u$ is defined by $P_u(v) := \langle u, v\rangle\, u$.

From Figure 5.4.1 it is clear that $P_u(v)$ is the point on $\mathbb{R}u$ closest to v. The following proposition asserts this.

Proposition 5.4.1 *For a unit vector u and any $v \in V$, we let*

$$P_u(v) := \langle v, u\rangle\, u.$$

Then $d(P_u(v), v) \leq d(\alpha u, v)$ for any $\alpha \in \mathbb{R}$.

Proof First observe that $(v - P_u(v)) \perp u$:

$$\langle v - P_u(v), u\rangle = \langle v, u\rangle - \langle \langle v, u\rangle\, u, u\rangle = \langle v, u\rangle - \langle u, v\rangle \langle u, u\rangle = \langle u, v\rangle - \langle u, v\rangle,$$

since $\langle u, u\rangle = 1$. Hence $(v - P_u(v)) \perp \alpha u$ for all $\alpha \in \mathbb{R}$ and therefore

$$(v - P_u(v)) \perp (P_u(v) - u) \text{ for all } \alpha \in \mathbb{R}.$$

Use Lemma 5.2.1 to get

$$||v - \alpha u||^2 = ||v - P_u(v) + P_u(v) - \alpha u||^2 = ||v - P_u(v)||^2 + ||P_u(v) - \alpha u||^2.$$

Thus, we get $d(v, \alpha u)^2 \geq d(v, P_u(v))^2$ and equality holds if and only if $\alpha u = P_u(v)$.

□

This is how proofs are written. Do you see why we thought of this proof? Usually, we start with the geometric idea and turn it into a rigorous proof. Here also that is what happened. A close look at Figure 5.4.2 tells us that we have a right angled triangle whose "vertices" are at αu, $P_u(v)$ and v.

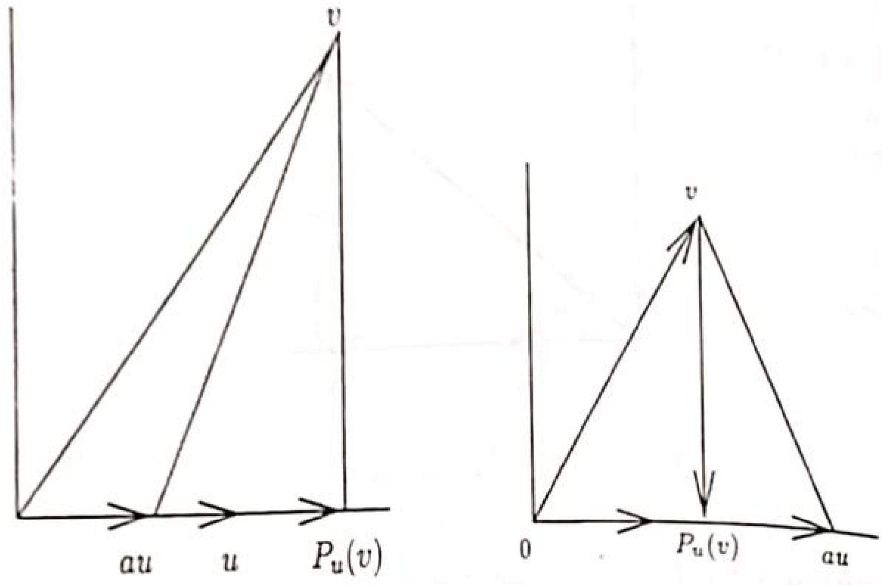


Figure 5.4.2 Projection is the closest approximation.

Naturally, the hypotenuse will be longer. For once, we tried to prove the result in a formal way without this geometric motivation. Hopefully, you understand the proof much better now and appreciate our efforts to put things in a geometric language!

Now we prove this result using calculus. The first impulse would be to consider the function $g(t) := \|v - tu\|$ and try to find its extreme values. A good analyst would not do this. For, the norm function is quite akin to the *modulus function* $|\ |$ on $\mathbb{R}$ which is not differentiable at the origin. Also, inner products are easier to deal with than norms. If you doubt me, go through the proof below.

We consider the function $f(t) = \langle v - tu, v - tu \rangle$. A point t_0 is a minimum of f if and only if it is a minimum of g. But f, as earlier, is a quadratic polynomial in t: $f(t) = \langle u, u \rangle t^2 - 2\langle u, v \rangle t + \langle v, v \rangle$. If t_0 is an extremum point, then $f'(t_0) = 0$. We find that

$$f'(t) = 2t\langle u, u \rangle - 2\langle u, v \rangle .$$

So $t_0 = \frac{\langle u,v \rangle}{\langle u,u \rangle}$. Also, $f''(t) = 2\langle u, u \rangle > 0$ if $u \neq 0$. Thus f attains a minimum at t_0. Therefore, the vector $\frac{\langle u,v \rangle}{\langle u,u \rangle} u$ is closest to v among all the vectors tu of $\mathbb{R}u$.

This motivates the following definition:

Definition 5.4.2 Let u be a unit vector and let v be an arbitrary vector. We define the *orthogonal projection* of v along u by $P_u(v) = \langle u, v \rangle u$. If u

is any nonzero vector then

$$P_u(v) = \left\langle \frac{u}{||u||}, v \right\rangle \frac{u}{||u||} = \frac{\langle u, v\rangle}{\langle u, u\rangle} u.$$

Exercise 5.4.1 Prove that $v - P_u(v) \perp u$, for any $u, v \in V$. (This is already seen and inserted here for future reference.)

5.5 Orthonormal Basis

Given any finite dimensional vector space we know that there exists a basis. But if the finite dimensional vector space has additional structures such as an inner product then we may look for a basis $\{e_1, \dots, e_n\}$ which has some additional properties involving the inner product. For instance, the standard basis $\{e_i\}$ of $\mathbb{R}^n$ has the following properties involving the dot product:

(1) Each of them is of unit norm.

(2) They are all mutually orthogonal to each other.

These properties can be succinctly put as a single condition with the use of the *Kronecker delta* δ_{ij}: δ_{ij} is defined as

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases} \tag{5.5.1}$$

Thus the standard basis $\{e_i\}$ of $\mathbb{R}^n$ has the property that $\langle e_i, e_j\rangle = \delta_{ij}$, for $1 \leq i, j \leq n$. This suggests the following definition:

Definition 5.5.1 A basis $\{v_1, \dots, v_n\}$ of V is said to be *orthonormal* if we have $\langle v_i, v_j\rangle = \delta_{ij}$ for $1 \leq i, j \leq n$.

To emphasize the geometric aspect of this definition, let us reformulate this. A basis $\{v_i\}$ of V is orthonormal if and only if each v_i is of unit length and they are mutually orthogonal to each other.

Example 5.5.1 The standard basis of $\mathbb{R}^n$ is an orthonormal basis.

Example 5.5.2 The basis $\{(e_1 + e_2)/\sqrt{2}, (e_1 - e_2)/\sqrt{2}\}$ is an orthonormal basis of $\mathbb{R}^2$. Can you now construct an orthonormal basis of $\mathbb{R}^n$ which is not the standard basis?

Exercise 5.5.1 Is the set $\{E_{ij}\}$ of $M(n, \mathbb{R})$ an orthonormal basis (see Exercise 5.1.5)?

What is the use of an orthonormal basis? Let $\{v_i\}$ be a basis of V. Let $v \in V$ be given. By the very definition of basis, we know that there exist scalars $\alpha_i \in \mathbb{R}$ such that $v = \sum_i \alpha_i v_i$. In the case of an arbitrary basis, we have no clue to these scalars. But in the case of an orthonormal basis we know what they are! Let us assume that $\{v_i\}$ is an orthonormal basis of V. Write $v = \sum_i \alpha_i v_i$. Let us take the inner product of both sides with the vector v_j. Using the orthonormal properties of the orthonormal basis, we get

$$\langle v, v_j\rangle = \left\langle \sum_i \alpha_i v_i, v_j \right\rangle = \sum_i \langle \alpha_i v_i, v_j\rangle = \sum_i \alpha_i \langle v_i, v_j\rangle = \sum_i \alpha_i \delta_{ij} = \alpha_j.$$

Thus α_j is $\langle v, v_j\rangle$, a quantity which involves the given vector v, the jth basis vector and the inner product. Even though this is simple, to emphasize its practical importance, we elevate it to the status of a theorem.

Theorem 5.5.1 *Let $\{v_i\}_{i=1}^n$ be an orthonormal basis of an inner product space V. If $v = \sum_{i=1}^n \alpha_i v_i$, then the ith coefficient $\alpha_i = \langle v, v_i\rangle$.*

□

Exercise 5.5.2 Let $\{v_i\}$ be an orthonormal basis. Let $x = \sum_i x_i v_i$ and $y = \sum_i y_i v_i$. Show the following:

(1) $\|x\|^2 = \sum_{i=1}^n (\langle v, x_i\rangle)^2$ (2) $\langle x, y\rangle = \sum_i x_i y_i$.

Exercise 5.5.3 This is the converse of Exercise 5.5.2. Let $\{v_i\}$ be a basis of V such that if $v = \sum_i \alpha_i v_i$ then $\|v\|^2 = \sum_i \alpha_i^2$. Prove that $\{v_i\}$ is an orthonormal basis.

Another use of an orthonormal basis is the following: Assume V and W are inner product spaces with orthonormal bases $\{v_i\}_{i=1}^m$ and $\{w_j\}_{j=1}^n$. Let $A\colon V \to W$ be any linear transformation. Then the general theory tells us that $Av_i = \sum_{j=1}^n \alpha_{ij} w_j$ for some scalars $\alpha_{ij} \in \mathbb{R}$. Since $\{w_j\}_{j=1}^n$ is an orthonormal basis of W, we can find α_{ij} explicitly as eariier:

$$\langle Av_i, w_k\rangle = \left\langle \sum_j \alpha_{ij} w_j, w_k \right\rangle = \sum_j \alpha_{ij} \langle w_j, w_k\rangle = \sum_j \alpha_{ij}\delta_{jk} = \alpha_{ik}.$$

Thus with respect to the bases $\{v_i\}_{i=1}^m$ and $\{w_j\}_{j=1}^n$ the matrix for A is

$$M_w^v(A) = \begin{pmatrix} \langle Av_1, w_1\rangle & \dots & \langle Av_m, w_1\rangle \\ \vdots & \ddots & \vdots \\ \langle Av_1, w_n\rangle & \dots & \langle Av_m, w_n\rangle \end{pmatrix} \tag{5.5.2}$$

Perhaps one should record this as another theorem because of its usefulness but we resist the temptation.

Does every inner product space have an orthonormal basis? Before we answer the question, we ask for a basis with a less stringent condition.

Definition 5.5.2 A set $E \subset V$ is said to be *orthogonal* if

(i) $0 \notin E$, and

(ii) $\langle x, y \rangle = 0$ for all $x, y \in E$ $x \neq y$.

An orthogonal basis is a basis which is also an orthogonal set.

Exercise 5.5.4 The following sets are orthogonal:

(1) Any orthonormal basis in V,

(2) $\{x + y, x - y\}$ in $\mathbb{R}^2$ with $(x, y) \neq 0$, and

(3) $\{(a, 0, 0, 0), (0, b, 0, 0), (0, 0, c, 0)\}$ in $\mathbb{R}^4$ with $abc \neq 0$.

Which of these are orthogonal bases?

The next lemma generalizes Exercise 5.2.5.

Lemma 5.5.2 *Any orthogonal set in an inner product space V is linearly independent.*

Proof Let $E = \{v_1, \ldots, v_k\}$ be an orthogonal set (of nonzero vectors). Assume that $\sum_{i=1}^{k} \alpha_i v_i = 0$ for $\alpha_i \in \mathbb{R}$. Then $\left\langle \sum_{i=1}^{k} \alpha_i v_i, v_j \right\rangle = 0$ for $1 \leq j \leq n$ and hence $\sum_{i=1}^{k} \alpha_i \langle v_i, v_j \rangle = 0$. Since $\langle v_i, v_j \rangle = 0$ for $i \neq j$, the only surviving term is $\alpha_j \langle v_j, v_j \rangle$. Since $v_j \in E$, and $0 \notin E$, $\langle v_j, v_j \rangle \neq 0$. This implies $\alpha_j = 0$. As j was arbitrary, we see that $\alpha_j = 0$ for all j.

□

Exercise 5.5.5 Suppose $\{v_1, \ldots, v_n\}$ is an orthogonal basis. Then

$$\left\{ e_i = \frac{v_i}{||v_i||} \;\middle|\; 1 \leq i \leq n \right\}$$

is an orthonormal basis.

Exercise 5.5.6 If $\{v_1, \ldots, v_n\}$ is an orthogonal basis of V and $v = \sum_i \alpha_i v_i$, then $\alpha_i = \dfrac{\langle v, v_i \rangle}{\langle v_i, v_i \rangle}$, $1 \leq i \leq n$.

We now show that $(\mathbb{R}^2, \langle , \rangle)$ has an orthonormal basis. By Exercise 5.5.5 it is enough to produce an orthogonal basis. Let $\{v_1, v_2\}$ be a basis of $\mathbb{R}^2$. Let $u_1 := v_1$. The idea is we want a nonzero vector u_2 which is orthogonal to v_1. Our earlier study of orthogonal projection of a vector onto a nonzero vector suggests a candidate, namely

$$u_2 := v_2 - \frac{\langle v_1, v_2\rangle}{\langle v_1, v_1\rangle} v_1 = v_2 - P_{u_1}(v_2).$$

(See Exercise 5.4.1.) Let $u_2 = v_2 - P_{u_1}(v_2)$. For completeness sake, we shall show that $u_1 \perp u_2$:

$$\begin{aligned}
\langle u_1, u_2\rangle &= \langle u_1, v_2 - P_{u_1}(v_2)\rangle \\
&= \langle u_1, v_2\rangle - \langle u_1, P_{u_1}(v_2)\rangle \\
&= \langle v_1, v_2\rangle - \left\langle v_1, \frac{\langle v_1, v_2\rangle}{\langle v_1, v_1\rangle} v_1 \right\rangle \\
&= \langle v_1, v_2\rangle - \frac{\langle v_1, v_2\rangle}{\langle v_1, v_1\rangle}\langle v_1, v_1\rangle \\
&= 0.
\end{aligned}$$

Further, $u_2 \neq 0$. For, if $u_2 = 0$, then $v_2 = P_{u_1}(v_2)$ is a scalar multiple of v_1. But this contradicts the assumption that $\{v_1, v_2\}$ is a basis. Hence $\{u_1, u_2\}$ is an orthogonal basis.

Imitating this argument, we prove the following theorem:

Theorem 5.5.3 *Let V be any inner product space. Then V has an orthonormal basis.*

Proof By Exercise 5.5.5, it is enough to produce an orthogonal basis of V.

Let $\{v_i\}_{i=1}^n$ be a basis of V. Let $u_1 = v_1$. As in the case of $\mathbb{R}^2$ above, we define

$$u_2 = v_2 - \frac{\langle v_2, v_1\rangle}{\langle v_1, v_1\rangle} v_1.$$

Then $u_2 \perp u_1$ (that is, $v_1 \perp v_2 - P_{u_1}(v_2)$). Therefore $\langle u_1, u_2\rangle = 0$. Also, $u_2 \neq 0$. For, if $u_2 = 0$, then $v_2 = \frac{\langle v_2, v_1\rangle}{\langle v_1, v_1\rangle} v_1$ will imply $\{v_1, v_2\}$ and hence the basis $\{v_i\}_1^n$ is linearly dependent, a contradiction. Let

$$u_3 = v_3 - \frac{\langle v_3, u_1\rangle}{\langle u_1, u_1\rangle} u_1 - \frac{\langle v_3, u_2\rangle}{\langle u_2, u_2\rangle} u_2.$$

Then $\langle u_3, u_1\rangle = 0$ and $\langle u_3, u_2\rangle = 0$ and $u_3 \neq 0$. For otherwise, v_3 is a linear combination of u_1 and u_2 and hence a linear combination of v_1 and v_2. This implies $\{v_1, v_2, v_3\}$ is linearly dependent, a contradiction.

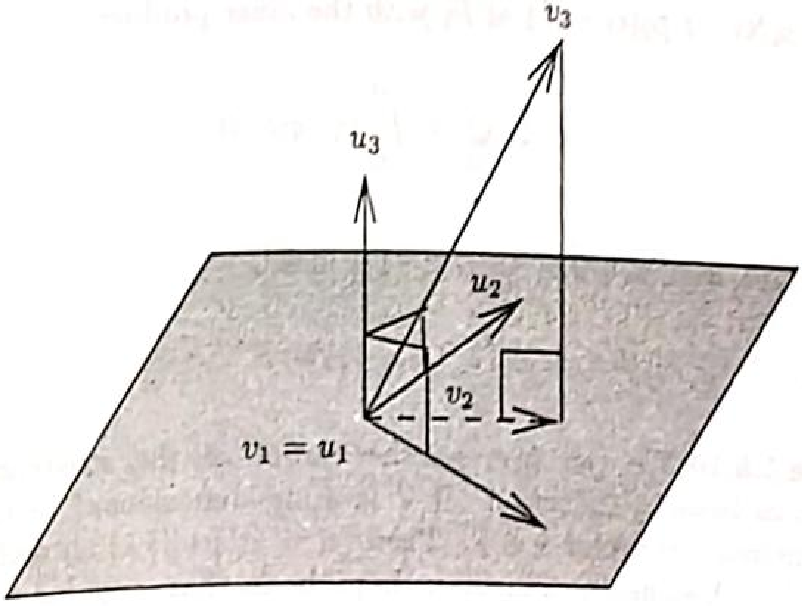


Figure 5.5.1 Gram-Schmidt process.

Proceeding as above by induction, define

$$u_k = v_k - \sum_{i=1}^{k-1} \frac{\langle v_k, u_i \rangle}{\langle u_i, u_i \rangle} u_i.$$

Then $\langle u_k, u_i \rangle = 0$ for all $1 \le i \le k-1$ and as before, $u_k \neq 0$. We have thus produced an orthogonal basis $\{u_1, \ldots, u_n\}$ of V. Then $\left\{ e_i = \frac{u_i}{\|u_i\|} \right\}$ is an orthonormal basis of V.

□

The above process of obtaining an orthogonal basis is known as the *Gram-Schmidt orthogonalization process.*

Exercise 5.5.7 Do you understand Exercise 5.2.3 now?

Exercise 5.5.8 Show that the Gram-Schmidt process does not disturb the initial r vectors if they already form an orthonormal set. That is, in the given basis $\{v_i\}_{i=1}^n$, if $\{v_1, \ldots, v_r\}$ is such that $\langle v_i, v_j \rangle = \delta_{ij}$ for $1 \le i, j \le r$ and if we apply the Gram-Schmidt process to get the orthonormal basis $\{e_i\}_{i=1}^n$, then $e_i = v_i$ for $1 \le i \le r$.

In particular, the Gram-Schmidt process when applied to an orthonormal basis returns it intact.

Exercise 5.5.9 Apply Gram-Schmidt process to obtain an orthonormal set:

(1) $\{(-1,0,1), (1,-1,0), (0,0,1)\}$ in $\mathbb{R}^3$.

(2) $\{1, p_1(t) = t, p_2(t) = t^2\}$ of $\mathcal{P}_2$ with the inner product

$$\langle p, q\rangle := \int_0^1 p(t)q(t)\,dt.$$

(3) $\{(1,1,1,1),(0,2,0,2),(-1,1,3,-1)\}$ in $\mathbb{R}^4$.

(4) $\{(1,-1,1,-1),(5,1,1,1),(2,3,4,-1)\}$ in $\mathbb{R}^4$.

Exercise 5.5.10 We can give another proof of the existence of an orthonormal basis by induction. If V is a one-dimensional vector space, choose any nonzero vector $v \in V$. Then $\{u := v/\|v\|\}$ is an orthonormal basis of V. Assume the induction hypothesis that *any* inner product space of dimension less than or equal to $n-1$ has an orthonormal basis. Let V be an n-dimensional inner product space. Choose any nonzero vector $v \in V$. Consider $W := v^\perp$. Then W is an inner product space of dimension $n-1$ (see Example 5.2.1). Thus, by induction, W has an orthonormal basis, say $\{w_1, \dots, w_{n-1}\}$. Let v_n be the unit vector along v. Then $\{w_1, \dots, w_{n-1}, v_n\}$ is an orthonormal basis of V.

Note that this proof is not constructive while Gram-Schmidt process gives us an algorithm to find an orthonormal basis.

5.6 Orthogonal Complements and Projections

Definition 5.6.1 Let $W \subset V$. Define

$$W^\perp = \{x \in V \mid \langle x, w\rangle = 0 \text{ for all } w \in W\}.$$

$W^\perp$ is called the *orthogonal complement* of W.

Exercise 5.6.1 $W^\perp$ is a vector subspace of V. (Exercise 5.2.8?)

Theorem 5.6.1 *Let W be a vector subspace of an inner product space V. Then $V = W \oplus W^\perp$. That is, any $x \in V$ is of the form $x = w + w'$, with $w \in W$ and $w' \in W^\perp$. Furthermore, this decomposition is unique.*

Proof Choose an orthonormal basis $\{w_1, \dots, w_r\}$ of W. Let $x \in V$. Let us define $w := \sum_{i=1}^{r} \langle x, w_i \rangle w_i \in W$. Let $w' = x - w$. Then $w' \in W^\perp$ as

$$\begin{aligned}
\langle w', w_k \rangle &= \langle x - w, w_k \rangle \\
&= \langle x, w_k \rangle - \langle w, w_k \rangle \\
&= \langle x, w_k \rangle - \left\langle \sum_{i=1}^{r} \langle x, w_i \rangle w_i, w_k \right\rangle \\
&= \langle x, w_k \rangle - \sum_{i=1}^{r} \langle x, w_i \rangle \langle w_i, w_k \rangle \\
&= \langle x, w_k \rangle - \sum_{i=1}^{r} \langle x, w_i \rangle \delta_{ik} \\
&= \langle x, w_k \rangle - \langle x, w_k \rangle \\
&= 0.
\end{aligned}$$

Thus w' is orthogonal to all w_k and hence to the vector subspace spanned by them, that is, to W. We can thus write $x = w + w'$ as required.

Now if $x = w_1 + w_1'$, $w_1 \in W$, and $w_1' \in W^\perp$, we then have

$$w + w' - w_1 - w_1' = 0 \qquad \text{or} \qquad w - w_1 = w' - w_1'.$$

The left hand side of this equation is in W and the right hand side is in $W^\perp$. So if $z = w - w_1 = w' - w_1'$, then $z \in W \cap W^\perp$. From Exercise 5.2.7 it follows that $z = 0$, that is, $w = w_1$ and $w' = w_1'$. In case, you have not solved Exercise 5.2.7, the solution follows: Since $z \in W^\perp$, $\langle z, y \rangle = 0$ for all $y \in W$. In particular, taking $y = z$, we get $\langle z, z \rangle = 0$. Hence $z = 0$. That is, $w - w_1 = 0 = w' - w_1'$ or $w = w_1$ and $w' = w_1'$. Thus the "decomposition" $x = w + w'$ is unique.

□

Definition 5.6.2 The decomposition of Theorem 5.6.1 is called the *orthogonal decomposition* of V with respect to the subspace W. The expression $x = w + w'$ in the theorem is called the orthogonal decomposition of the vector x with respect to W. The inner product space is said to be an *orthogonal direct sum* of W and $W^\perp$.

Exercise 5.6.2 When do you say V is the orthogonal direct sum of W_i, $1 \le i \le k$?

Exercise 5.6.3 Let W be a vector subspace of V. What is $(W^\perp)^\perp$?

The next definition generalizes Definition 5.4.2:

Definition 5.6.3 Let $W \subseteq V$ be a vector subspace of an inner product space V. Then the *orthogonal projection* P_W of V onto W is the map $P_W(x) = w$ where $x = w + w'$ is the orthogonal decomposition of x.

During the course of the proof of Theorem 5.6.1, we have derived an expression of P_W in terms of an orthonormal basis of W. If $\{w_1, \dots, w_r\}$ is an orthonormal basis of W, then $P_W(v) := \sum_{i=1}^{r} \langle v, w_i \rangle w_i$.

Lemma 5.6.2 *Let W be a vector subspace of V. Let $\{w_i\}_{i=1}^{r}$ be an orthonormal basis of W. Let $\{u_j\}_{j=1}^{s}$ be an orthonormal basis of $W^{\perp}$. Then $\{w_i\} \cup \{u_j\}$ is an orthonormal basis of V.*

Proof The set $\{w_i\} \cup \{u_j\}$ is certainly orthonormal. For, any pair is one of the forms: (w_i, w_k), (u_j, u_l), (w_i, u_j). The first two have inner products δ_{ik} and δ_{jl} respectively while the inner product of the third pair is 0. Hence, they are linearly independent. Also, they span V. For, given $v \in V$, by the orthogonal decomposition theorem, we can write $v = w + w'$ with $w \in W$ and $w' \in W^{\perp}$. But w (respectively w') is a linear combination of w_i's (respectively u_js). (Why?) Hence v is a linear combination of $\{w_i\} \cup \{u_j\}$. Thus they form a basis.

□

What is the geometric interpretation of P_W? If $v \in V$, then $P_w(v) \in W$ is the unique element of W which is nearest to v: $||v - w|| \geq ||v - P_w(v)||$ for all $w \in W$. In terms of the distance function d, we have

$$d(v, P_W(v)) \leq d(v, w)$$

for any $w \in W$. Let $w \in W$ be an arbitrary element. We denote $P_W(v)$ by x. Then

$$||v - w||^2 = ||v - x + x - w||^2 = ||v - x||^2 + ||x - w||^2 \qquad (5.6.1)$$

since $v - x \perp W$. Thus for all $w \in W$, $||v - w||^2 \geq ||v - x||^2$ and equality holds if and only if $||x - w||^2 = 0$, that is, $w = P_W(v)$. Further, the distance $d(v, W) := \inf_{w \in W} d(v, w)$ is $d(v, P_W(v)) = ||v - P_W(v)||$.

Proposition 5.6.3 *Any subspace W of $\mathbb{R}^n$ is the set of solutions (that is, a solution space) of a homogeneous system of linear equations.*

Proof Let $\{w_1, \dots, w_r\}$ (respectively $\{v_1, \dots, v_s\}$) be an orthonormal basis of W (respectively $W^{\perp}$). Then $\{w_1, \dots, w_r, v_1, \dots, v_s\}$ is an orthonormal basis of $\mathbb{R}^n$ by Lemma 5.6.2. Now $x \in \mathbb{R}^n$ lies in W if and only if

$$\langle x, v_i \rangle = 0, \quad 1 \leq i \leq s.$$

This is a homogeneous system of linear equations. For, if we write

$$v_i = \sum_{j=1}^{n} \alpha_{ij} e_j \text{ and } x = \sum_{j=1}^{n} x_j e_j$$

with respect to the standard basis, then $x \in W$ if and only if x is a solution of the homogeneous system

$$\sum_{j=1}^{n} \alpha_{ij} x_j = 0, \quad 1 \leq i \leq r.$$

□

The rest of the subsection may be omitted in the first reading.

Definition 5.6.4 Let V be an *arbitrary* vector space. V need not be an inner product space. A linear map $P: V \to V$ is said to be *idempotent* if $P^2 = P$.

Exercise 5.6.4 Show that the orthogonal projection P_W with respect to a vector subspace W is idempotent.

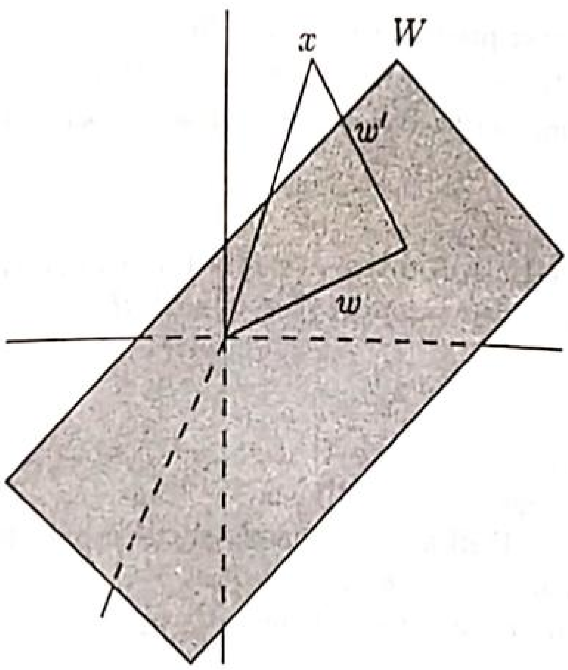


Figure 5.6.1 Orthogonal projection.

Exercise 5.6.5 Let V be an arbitrary vector space. Let W_i be vector subspaces such that $V = W_1 \oplus W_2$ (see Definition 2.3.8). If $v = w_1 + w_2$ is given, define $P_i(v) = w_i$ for $i = 1, 2$. Then P_i is idempotent. P_i is called the projection of V onto W_i with respect to the given direct sum decomposition $V = W_1 \oplus W_2$. (P_i depends on the factors W_i as it is possible to have $V = W_1 \oplus W_2$ and $V = W_1 \oplus W_2'$ with $W_1 \neq W_2$. The corresponding projections P_1 and P_1' will then be different. Find an example of this phenomenon in $\mathbb{R}^2$.)

Definition 5.6.5 If V is an inner product space, a linear map $T: V \to V$ is said to be *symmetric* (with respect to the given inner product) if $\langle Tx, y\rangle = \langle x, Ty\rangle$ for all $x, y \in V$.

Exercise 5.6.6 Check that the orthogonal projection P_W is symmetric.

Exercise 5.6.7 Show that an orthogonal projection is a projection and that a projection is an orthogonal projection if and only if it is symmetric.

5.7 Linear Functionals and Hyperplanes

V stands for an inner product space with $\dim V = n$.

We have already seen that for any fixed $a \in V$, the map $f_a : x \mapsto \langle x, a\rangle$ is a linear functional on V. The following theorem says these are the only ones.

Theorem 5.7.1 (Riesz Representation Theorem) *Given a linear form $f : V \to \mathbb{R}$, there exists a unique $y \in V$ such that $f(x) = \langle x, y\rangle$ for all $x \in V$.*

Proof Suppose there exists $y \in V$ such that $f(x) = \langle x, y\rangle$ for all $x \in V$. Choose an orthonormal basis $\{e_i\}_{i=1}^n$ of V. Then, $y = \sum_{i=1}^n \alpha_i e_i$ for some $\alpha_i \in \mathbb{R}$. Now, $f \in L(V, \mathbb{R})$ and f is completely determined if we know $f(e_i)$ for $1 \le i \le n$. Now $f(e_i) = \langle e_i, y\rangle = \alpha_i$ for $1 \le i \le n$. This suggests that we take $y = \sum f(e_i)e_i$. It is easy to check that $f(x) = \langle x, y\rangle$ for all $x \in V$. For, if $x = \sum \alpha_i e_i$, then

$$f(x) = \sum \alpha_i f(e_i). \tag{5.7.1}$$

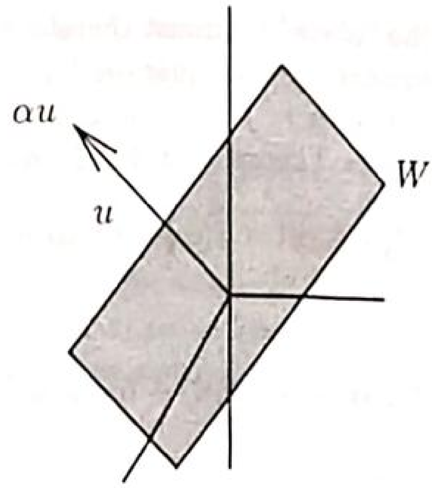


Figure 5.7.1 Riesz representation theorem.

Also,

$$\begin{aligned} \langle x, y\rangle &= \left\langle x, \sum f(e_i)e_i \right\rangle \\ &= \left\langle \sum \alpha_j e_j, \sum f(e_i)e_i \right\rangle \\ &= \sum_{i,j} f(e_i)\alpha_j \langle e_i, e_j\rangle \\ &= \sum_{i,j} f(e_i)\alpha_j \delta_{ij} \\ &= \sum f(e_i)\alpha_i. \end{aligned} \tag{5.7.2}$$

From Equations 5.7.1 and 5.7.2 it follows that $f(x) = \langle x, y\rangle$ for all $x \in \mathbb{R}^n$.
Now, suppose z is such that $f(x) = \langle x, z\rangle$ for all $x \in V$. Then,

$$f(x) = \langle x, z\rangle = \langle x, y\rangle .$$

Hence $\langle x, z - y\rangle = 0$ for all x. In particular, for $x = z - y$, we have $\langle z - y, z - y\rangle = 0$. But then $z - y = 0$ or $z = y$. Hence this y is unique.

□

Did you recognize Exercise 5.2.6 and its solution towards the end of this proof?

We now give a geometric proof of the *Riesz Representation Theorem.*

Proof If $f \equiv 0$, then the obvious choice is $y = 0$. If $f \neq 0$, then f is a linear form and $W = \ker f$ is of dimension $n - 1$, where $n = \dim V$. Thus there is a unit vector u perpendicular to W, for $V = W \oplus W^\perp$ (that is, u is a unit vector normal to the "plane"). y must therefore be a multiple αu of u. The choice of α is determined by the equation $f(u) = \langle u, y \rangle = \langle u, \alpha u \rangle = \alpha$. Thus we take $y = \alpha u$ where $\alpha = f(u)$. For $x \in V$, we have $x = w + tu$, where $w \in W$ and $t \in \mathbb{R}$ (see Theorem 5.6.1). Then

$$f(x) = f(w + tu) = f(w) + tf(u) = tf(u).$$

Also,

$$\langle x, y \rangle = \langle w + tu, \alpha u \rangle = \alpha \langle w, u \rangle + t\alpha \langle u, u \rangle = t\alpha = tf(u).$$

Hence the result.

□

5.7.1 Hyperplanes

This section is an extended discussion of the geometric idea introduced in the second proof of Riesz representation theorem.

We shall start with a geometric definition of a plane in $\mathbb{R}^3$.

Definition 5.7.1 A plane in $\mathbb{R}^3$ through a point p with normal N is the set of all lines passing through p and perpendicular to N. We denote such a plane by $\Pi(p, N)$ (see Figure 5.7.2).

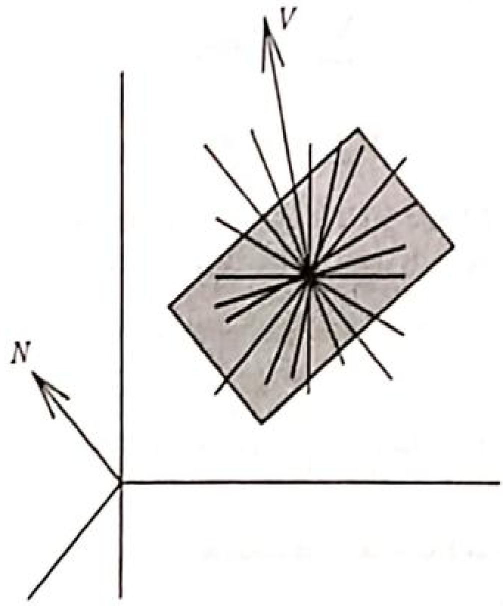


Figure 5.7.2 Planes in $\mathbb{R}^3$.

Let $X \in \Pi(p, N)$ be a point on the plane. Then the direction vectors $t(X - p)$ of the line $\ell(X, p)$ are perpendicular to N if and only if $\langle t(X-p), N\rangle = 0$ for all t, that is, if and only if $\langle (X-p), N\rangle = 0$. This happens if and only if $\langle X, N\rangle = \langle p, N\rangle$, or if and only if $\langle X, N\rangle = d$, where $d = \langle p, N\rangle$ is a constant.

In $\mathbb{R}^3$, let $N = (a, b, c) \neq 0$, $X = (x, y, z)$ and $p = (x_0, y_0, z_0)$. Then $\langle X, N\rangle = ax + by + cz$. Hence $\langle X, N\rangle = d$ is equivalent to the linear equation $ax + by + cz = d$, where $d = ax_0 + by_0 + cz_0$. Thus, from our geometric definition of a plane we see that any plane is given by a linear equation. Conversely, if a nontrivial linear equation $\sum_{i=1}^{3} a_i x_i = b$ is given, then it defines a plane whose normal is (a_1, a_2, a_3).

This shows us how to define analogous objects in higher dimensions.

Definition 5.7.2 Let V be an inner product space. Fix a nonzero vector $N \in V$ and a real number b. Then the set $\Pi := \{x \in V \mid \langle x, N\rangle = b\}$ is called a *hyperplane* in V with *normal* N.

Thus points in $\mathbb{R}$, lines in $\mathbb{R}^2$ and planes in $\mathbb{R}^3$ are hyperplanes.

However note that in Chapter 3 we were able to define a hyperplane in any vector space, not necessarily inner product spaces. Does this notion coincide with that we defined earlier in Chapter 3? The answer is yes.

First we show that any hyperplane according to this new definition is a hyperplane according to Chapter 3. There are two ways of seeing this. Let Π be a hyperplane in V. If we fix a point $p \in \Pi$ and take W to be $\{x \in V \mid \langle x, N\rangle = 0\}$, then $\Pi = p + W$. The other one is to show that the line joining any two points of Π is contained in Π. If $p_i \in \Pi$, for $i = 1, 2$, then $\langle p_i, N\rangle = b$. So $\langle tp_1 + (1-t)p_2, N\rangle = b$.

Conversely, if Π is a hyperplane according to our earlier definition, say, of the form $\Pi = x + W$ where W is a vector subspace of dimension $\dim V - 1$. By orthogonal decomposition theorem, we can find a nonzero vector $N \perp W$. If $b := \langle x, N\rangle$, we can easily show that

$$\Pi = \{v \in V \mid \langle v, N\rangle = b\}.$$

Thus our new definition gives a geometric characterization of hyperplanes in an inner product space. Another way of saying this is that any hyperplane is of the form $f_a^{-1}(d)$ for some nonzero $a \in V$ and $d \in \mathbb{R}$. Let u be the unit vector along a. Extend it to an orthonormal basis of V, say, $\{v_1, \ldots, v_{n-1}, v_n u\}$. Let x_i be the coordinates associated with this basis: $x_i(v) := \langle v, v_i\rangle$. Then the hyperplane $f_a^{-1}(d)$ has the following description: $\{v \in V \mid x_n(v) = \|a\|\, d\}$.

In the rest of this section, we want to find a formula for the distance $d(v, \Pi) := \inf\{d(v, w) \mid w \in W\}$ of a point v to a hyperplane Π. If the reader wishes, he may assume $V = \mathbb{R}^n$ (or even $\mathbb{R}^3$) with the dot product.

We shall first derive a formula for $d(v, W)$ if W is a vector subspace of dimension $n-1$. We have seen in Section 5.6 that if W is *any* vector subspace, then $d(v, W) = ||v - P_W(v)||$. We can say more if we assume that $\dim W = \dim V - 1$. Let $V = W \oplus W^\perp$ be the orthogonal decomposition. By Lemma 5.6.2, $\dim W^\perp = 1$. So, if $W^\perp = \mathbb{R}v$, choose a unit vector $N \in W^\perp$. Then N is of the form $\pm(v/\,||v||)$. Note that $v = P_W(v) + \langle v, N\rangle\, N$. (Why?) Thus

$$d(v, W) = ||v - P_W(v)|| = ||\langle v, N\rangle\, N|| = |\langle v, N\rangle|\,, \tag{5.7.3}$$

where N is a unit vector orthogonal to W.

We now specialize Equation (5.7.3). Let $V = \mathbb{R}^3$ and W be a plane passing through the origin. Then there exist $a, b, c \in \mathbb{R}$, not all zero, such that $W = \{(x, y, z) \in \mathbb{R}^3 \mid ax + by + cz = 0\}$. W is a vector subspace of V and we have $\mathbb{R}^3 = W \oplus \mathbb{R}(a, b, c)$. In geometric language, (a, b, c) is a normal to the plane W. We take as unit normal $N = (a, b, c)/\sqrt{a^2 + b^2 + c^2}$. Then Equation (5.7.3) reads

$$\begin{aligned} d((x, y, z), W) &= \langle (x, y, z), N\rangle \\ &= \left\langle (x, y, z), \frac{(a, b, c)}{\sqrt{(a^2 + b^2 + c^2)}} \right\rangle \\ &= \frac{ax + by + cz}{\sqrt{(a^2 + b^2 + c^2)}}. \end{aligned}$$

This is a well-known formula from analytic geometry of three dimensions.

To treat the general case, we shall exploit the translation invariant nature of the distance d on V (Exercise 5.1.11).

Let us now assume that the hyperplane is given by $\langle x, u\rangle = d$ for a unit vector u. Let $W := \ker f_u = \{x \in V \mid \langle x, u\rangle = 0\}$. Then W is an $n - 1$ dimensional vector subspace with $u \perp W$. Let $p \in \Pi$ be any arbitrary point. Then $W = \Pi - p$ (verify this). We know that $d(v - p, x - p) = d(v, x)$ for all $v, x, p \in V$. Hence, we see

$$\begin{aligned} \inf\{d(v, x) \mid x \in \Pi\} &= \inf\{d(v - p, x - p) \mid x \in W\} \\ &= \inf\{d(v - p, w) \mid w \in \Pi\} \\ &= |\langle v - p, u\rangle| \qquad \text{by Equation (5.7.3)} \\ &= |\langle v, u\rangle - \langle p, u\rangle|\,. \end{aligned} \tag{5.7.4}$$

If the hyperplane is given by $f_N^{-1}(d)$ for some nonzero N, then by taking $u = N/\,||N||$ in Equation (5.7.4), we see that

$$d(v, \Pi) = \frac{1}{||N||}\,|\langle v, N\rangle - \langle p, N\rangle|\,. \tag{5.7.5}$$

As earlier in the case of $\mathbb{R}^3$ and $\Pi = \{(x,y,z) \mid ax+by+cz=d\}$, Equation (5.7.5) becomes the well-known formula

$$d((x,y,z),\Pi) = \frac{|ax+by+cz-d|}{\sqrt{a^2+b^2+c^2}}.$$

The rest of the section may be omitted as it derives the formula for $d(p,\Pi)$ in two more different ways. They are included here just to show the different approaches possible to attack the same problem and also to establish the supremacy of the above approach.

Let Π be a plane given by $\langle X,N\rangle = d$. Let $q \in V\setminus\Pi$. Our aim is to compute $d(q,\Pi) := \inf_{p\in\Pi} d(q,p) = \inf_{p\in\Pi} ||q-p||$.

If $p \in \Pi$ is such that $d(q,p) = d(q,\Pi)$ we claim that $q-p \perp \Pi$. Grant this claim for a moment (see Figure 5.7.3). Then $q-p = \alpha N$ for some $\alpha \in \mathbb{R}$. Hence $\langle q-p,N\rangle = \alpha\langle N,N\rangle = \alpha\,||N||^2$ so that

$$|\langle q-p,N\rangle| = |\alpha|\,||N||^2 = ||q-p||\,||N||.$$

It follows that

$$d(p,q) = ||p-q|| = \frac{|\langle q-p,N\rangle|}{||N||}.$$

This reduces to the standard formula seen in analytic geometry. If we assume that the plane is given by the equation $ax+by+cz=d$, then we let $N=(a,b,c)$ and $q=(x,y,z)$. The above formula then becomes

$$d(q,\Pi) = \frac{|ax+by+cz-d|}{\sqrt{a^2+b^2+c^2}}.$$

It remains to show that $q-p \perp \Pi$. As was done earlier, we use calculus to prove this.

Let v be such that $\langle v,N\rangle = 0$. Consider $p+tv$. Then $p+tv \in \Pi$. For, $\langle p+tv,N\rangle = \langle p,N\rangle + t\langle v,N\rangle = d+0 = d$.

If we set $f(t) := ||p+tv-q||^2$, then

$$\begin{aligned} f(t) &= \langle p-q+tv, p-q+tv\rangle \\ &= \langle p-q,p-q\rangle + 2t\langle p-q,v\rangle + t^2\langle v,v\rangle. \end{aligned}$$

Since p is assumed to be nearest to q and $f(0) = ||p-q||^2$, we see that f has a minimum at $t=0$. Thus $f'(0)$ must be zero. Since

$$f'(t) = 2\langle p-q,v\rangle + 2t\langle v,v\rangle$$

we have $0 = f'(0) = 2\langle p-q,v\rangle$. Thus $p-q \perp v$. Since v is any vector perpendicular to N, we see that $p-q = \alpha N$ for some $\alpha \in \mathbb{R}$.

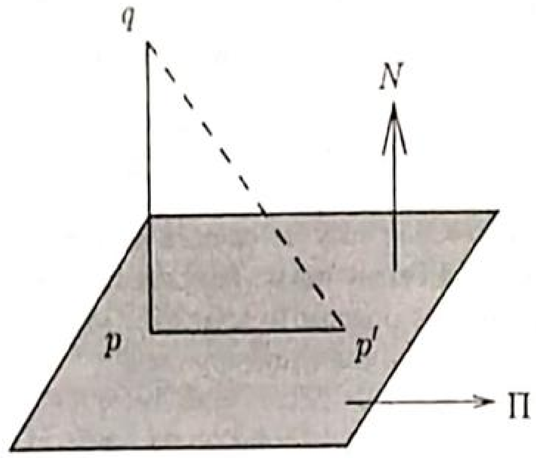


Figure 5.7.3 Distance between q and Π.

The crucial question is: How do we know that there exists one such $p \in \Pi$? Our entire analysis hinged on the existence of one such point. Here you need topology or analysis (depending on your preference) to answer this question affirmatively.

We now give a highly geometric proof which gives us the point and also lets us calculate the distance $d(q, \Pi)$.

By geometry (see Figure 5.7.3) we expect that the point of intersection of the perpendicular dropped from q to Π will be the required point. The perpendicular is the line through q with direction vector N.

Let us find the point of intersection p of the normal line $\ell(q; N)$ to the plane and the plane. Then p is of the form $q + t_0 N$ for some $t \in \mathbb{R}$. The point p lies on Π if and only if $\langle p, N \rangle = b$, if and only if $\langle q + t_0 N, N \rangle = b$, if and only if $\langle q, N \rangle + t_0 \langle N, N \rangle = b$ if and only if $t_0 = \frac{b - \langle q, N \rangle}{\langle N, N \rangle}$. Thus the required point is $p := q + t_0 N$.

Now if $p' \in \Pi$, then

$$\langle p' - p, p - q \rangle = \langle p', t_0 N \rangle - \langle p, t_0 N \rangle = t_0 b - t_0 b = 0.$$

Thus $p - p'$ and $p - q$ are perpendicular for any $p' \in \Pi$. Therefore, by Pythagoras theorem Lemma 5.2.1, we have

$$\|p' - p + p - q\|^2 = \|p - p'\|^2 + \|p - q\|^2.$$

This shows that $\|p - q\| \leq \|p' - q\|$ and equality holds if and only if $p = p'$.

Furthermore,

$$\begin{aligned} d(q,\Pi) = d(p,q) &= \|p-q\| \\ &= \left\| \frac{\langle q,N\rangle - b}{\langle N,N\rangle} N \right\| \\ &= \frac{|\langle q,N\rangle - b|}{\|N\|^2} \|N\| \\ &= \frac{|\langle q,N\rangle - b|}{\|N\|}. \end{aligned}$$

I hope you enjoyed seeing how we turn geometric ideas into rigorous proofs.

5.8 Orthogonal Transformations

V denotes an inner product space of dimension n unless stated otherwise.

We look for linear maps $T\colon V \to V$ which "preserve" the extra structure $\langle , \rangle$. So we make the following definition:

Definition 5.8.1 Let V be an inner product space. A linear transformation $T\colon V \to V$ is said to be *orthogonal* if $\langle Tx, Ty\rangle = \langle x, y\rangle$ for all $x, y \in V$.

Note that this definition implies that $\|Tx\| = \|x\|$ for all $x \in V$. What is surprising is that this is sufficient too:

Theorem 5.8.1 *Let $T\colon V \to V$ be linear. The following are equivalent:*

(1) *T is orthogonal.*

(2) *$\|Tx\| = \|x\|$ for all $x \in V$.*

(3) *T takes an orthonormal basis to an orthonormal basis. That is, if $\{e_i\}_{i=1}^n$ is an orthonormal basis, then $\{Te_i\}_{i=1}^n$ is an orthonormal basis.*

Proof As observed, (1) implies (2): If T is orthogonal, then

$$\langle Tx, Tx\rangle = \langle x, x\rangle .$$

The left side is $\|Tx\|^2$ and the right side is $\|x\|^2$ and they are equal. Hence (2) follows.

To show (2) implies (1) we need to prove that if $\|Tx\| = \|x\|$ for all $x \in V$, then $\langle Tx, Ty\rangle = \langle x, y\rangle$ for all $x, y \in V$. This is an often encountered idea. We know something about $\|x\|$ and we want to say something about

$\langle x, y \rangle$. To get these "cross-terms", the idea is to exploit what we know about $||x+y||$ or $||x-y||$. We have

$$\langle x+y, x+y \rangle = ||x||^2 + ||y||^2 + 2\langle x, y \rangle \qquad \text{for } x, y \in V. \tag{5.8.1}$$

We also have

$$\begin{aligned} \langle T(x+y), T(x+y) \rangle &= \langle Tx+Ty, Tx+Ty \rangle \quad \text{by linearity of } T \\ &= \langle Tx, Tx \rangle + \langle Ty, Ty \rangle + 2\langle Tx, Ty \rangle \\ &= ||Tx||^2 + ||Ty||^2 + 2\langle Tx, Ty \rangle . \end{aligned}$$

By hypothesis $||Tx|| = ||x||$, $||Ty|| = ||y||$ and $||T(x+y)|| = ||x+y||$ so that this yields

$$||x+y||^2 = ||T(x+y)||^2 = ||x||^2 + ||y||^2 + 2\langle Tx, Ty \rangle . \tag{5.8.2}$$

Comparing Equation (5.8.1) with Equation (5.8.2), we get $\langle Tx, Ty \rangle = \langle x, y \rangle$. Thus (1) is proved.

We prove that (3) implies (1): Let $\{e_i\}$ be an orthonormal basis of V. We are given that $\{Te_i\}$ is again an orthonormal basis of V. We are to show that T is orthogonal. Let $x = \sum x_i e_i$ and $y = \sum y_j e_j$. Then $Tx = \sum x_i Te_i$ and $Ty = \sum y_i Te_i$ (by linearity of T). Hence

$$\begin{aligned} \langle Tx, Ty \rangle &= \left\langle \sum_i x_i Te_i, \sum_j y_j Te_j \right\rangle \quad \text{by linearity of } T \\ &= \sum_{i,j} x_i y_j \langle Te_i, Te_j \rangle \\ &= \sum_{i,j} x_i y_j \delta_{ij} \quad \text{since } \{Te_i\} \text{ is orthonormal} \\ &= \sum_i x_i y_i \\ &= \langle x, y \rangle \quad \text{by Exercise 5.5.2.} \end{aligned}$$

Hence T is orthogonal.

Conversely, if T is orthogonal, then $\langle Te_i, Te_j \rangle = \delta_{ij}$. So $\{Te_i\}_{i=1}^n$ is an orthonormal set. By Lemma 5.5.2 it is linearly independent. Since it has n elements, it is a basis. Hence $\{Te_i\}$ is an orthonormal basis. Thus (1) implies (3).

We have shown that (1) is equivalent to (2) and also to (3). Thus (1), (2) and (3) are mutually equivalent.

□

Lemma 5.8.2 *If $f : V \to V$ is any map such that*

(1) $f(0) = 0$,

(2) $||f(x) - f(y)|| = ||x - y||$,

then f is an orthogonal linear transformation.

Remark 5.8.1 This result is really very amazing. Note that the second condition means that $d(f(x), f(y)) = d(x, y)$ for all $x, y \in V$, that is, f preserves distances. Thus a map $f: V \to V$ which preserves the distances and *maps the zero vector to itself* is necessarily a linear map and also orthogonal. Recall that distance is twice removed from the inner product. From the inner product, one gets the norm and from it the distance. The bond seems to be so strong that the "distant" cousin forces linearity. Note that the result is no longer true if we do not assume $f(0) = 0$. For, if we take $f = T_a$, the translation by a nonzero vector $a \in V$, then f preserves distances and it is not linear (see Exercise 4.1.2 and Exercise 5.1.11).

Proof From (1) and (2), we have

$$||f(x)|| = ||f(x) - 0|| = ||f(x) - f(0)|| = ||x - 0|| = ||x||. \tag{5.8.3}$$

that is, $||f(x)|| = ||x||$ for all $x \in V$. Hence

$$\begin{aligned} ||f(x) - f(y)||^2 &= \langle f(x), f(x)\rangle - 2\langle f(x), f(y)\rangle + \langle f(y), f(y)\rangle \\ &= ||f(x)||^2 + ||f(y)||^2 - 2\langle f(x), f(y)\rangle \\ &= ||x||^2 + ||y||^2 - 2\langle f(x), f(y)\rangle, \quad \text{by Equation (5.8.3).} \end{aligned}$$

Using (2) again, we get

$$||x - y||^2 = ||f(x) - f(y)||^2.$$

But $||x - y||^2 = ||x||^2 + ||y||^2 - 2\langle x, y\rangle$. It follows that $\langle f(x), f(y)\rangle = \langle x, y\rangle$. Thus f preserves the inner product. In particular, if $\{e_i\}_{i=1}^n$ is an orthonormal basis, $\{f(e_i)\}_{i=1}^n$ is an orthonormal basis too. Now, let $x = \sum x_i e_i \in V$. Since $\{f(e_i)\}_{i=1}^n$ is an orthonormal basis of V,

$$\begin{aligned} f(x) &= \sum \langle f(x), f(e_i)\rangle f(e_i) \quad \text{by Theorem 5.5.1} \\ &= \sum \langle x, e_i\rangle f(e_i) \\ &= \sum x_i f(e_i). \end{aligned}$$

This means that $f(\sum x_i e_i) = \sum x_i f(e_i)$. That is, f is linear. $\square$

The following corollary identifies all distance preserving maps $f: V \to V$. It says they are compositions of an orthogonal linear transformations and translations.

Corollary 5.8.3 *Let $g: V \to V$ be such that $\|g(x) - g(y)\| = \|x - y\|$ for all $x, y \in V$. Then there exists a unique $v \in V$ and an orthogonal linear transformation $A: V \to V$ such that $g(x) = Ax + v$ for all $x \in V$.*

Proof Take $v = g(0)$ and $f(x) = g(x) - g(0)$. Then one easily checks that f satisfies the hypothesis of Lemma 5.8.2 and hence $f(x) = Ax$ for some orthogonal linear transformation $A: V \to V$ and that $g(x) = Ax + v$.

□

Exercise 5.8.1 Let V be an arbitrary vector space. Let T_v denote the translation by v. Prove the following:

(1) T_v is a bijection for all $v \in V$ such that $T_v^{-1} = -T_v$.
(2) $T_{v+w} = T_v + T_w$ for all $v, w \in V$.
Translation is the device we employ to "shift the origin".

Exercise 5.8.2 Let A be an orthogonal linear map of V, T_v a translation. What is the inverse of $A \circ T_v$? What are $A \circ T_v$, $T_v \circ A$, $A \circ T_v \circ A^{-1}$, $T_v^{-1} \circ A \circ T_v$? Note that in general these are not linear maps.

If you know some group theory, the answers to these questions should tell you that the set

$$\{A \circ T_v \mid A \text{ is any orhogonal map of } V \text{ and } v \in V\}$$

is a group, called the group of *rigid motions*. This group is nonabelian. The set $\{T_v \mid v \in V\}$ is a normal subgroup.

Let V be an inner product space and let $T: V \to V$ be an orthogonal linear transformation. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of V. Then $\{Te_1, \ldots, Te_n\}$ is also an orthonormal basis of V. Let $T(e_i) = \sum \alpha_{ij} e_j$. We know how to find α_{ik}: $\langle Te_i, e_k\rangle = \sum_k \alpha_{ij} \langle e_j, e_k\rangle = \alpha_{ik}$. Therefore the matrix $M(T)$ with respect to this orthonormal basis is given by

$$\begin{aligned} M(T) &= \begin{pmatrix} \langle Te_1, e_1\rangle & \ldots & \langle Te_n, e_1\rangle \\ \langle Te_1, e_2\rangle & \ldots & \langle Te_n, e_2\rangle \\ \vdots & \ddots & \vdots \\ \langle Te_1, e_n\rangle & \ldots & \langle Te_n, e_n\rangle \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{21} & \ldots & \alpha_{n1} \\ \alpha_{12} & \alpha_{22} & \ldots & \alpha_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1n} & \alpha_{2n} & \ldots & \alpha_{nn} \end{pmatrix} \\ &= (C_1, \ldots, C_n), \end{aligned}$$

where C_i is the ith column of $M(T)$. We think of C_i as a column vector in $\mathbb{R}^n$.

Since T is orthogonal, $\{Te_i\}$ is an orthonormal basis. Hence

$$\langle Te_i, Te_j\rangle = \delta_{ij}.$$

Since $Te_i = \sum_j \alpha_{ij} e_j$, we have

$$\begin{aligned}
\langle Te_i, Te_j\rangle &= \left\langle \sum_r \alpha_{ir} e_r, \sum_s \alpha_{js} e_s \right\rangle \\
&= \sum_{r,s} \alpha_{ir}\alpha_{js} \langle e_r, e_s\rangle \\
&= \sum_{r,s} \alpha_{ir}\alpha_{js}\delta_{rs} = \sum_r \alpha_{ir}\alpha_{jr} = C_i \cdot C_j,
\end{aligned}$$

the dot product of the column vectors in $\mathbb{R}^n$.

This suggests the following definition:

Definition 5.8.2 An $n \times n$ matrix $A = (a_{ij})$ is said to be *orthogonal* if $C_i \cdot C_j = \delta_{ij}$, where C_i stands for the ith column of A considered as a column vector in $\mathbb{R}^n$.

Note Let V be an inner product space. If we start with any orthonormal basis $\{e_1, \ldots, e_n\}$ of V and an orthogonal transformation $T : V \to V$ and write down its matrix with respect to this orthonormal basis, then the matrix is orthogonal. This is just what we have seen and which motivated the definition of an orthogonal matrix.

Exercise 5.8.3 In fact, if an orthonormal basis is fixed, there exists a one-one correspondence between orthogonal transformations and orthogonal matrices. Can you prove this? If an orthogonal matrix $A = (a_{ij})$ is given, we know how to obtain a linear map of V to itself. It is then easy to show that this map is orthogonal.

Exercise 5.8.4 A matrix A is orthogonal if and only if $A^{-1} = A^t$ if and only if $AA^t = I$. Then we have

$$1 = \det I = \det(AA^t) = \det A \det A^t = (\det A)^2.$$

This means that $\det A = \pm 1$ for an orthogonal matrix A (see Theorem 6.4.1).

Exercise 5.8.5 Define an "orthogonal linear map" from an inner product space V into another. Prove a result analogous to Theorem 5.8.1.

Exercise 5.8.6 Show that the choice of an orthonormal basis $\{v_1, \ldots, v_n\}$ for V gives rise to an orthogonal linear map T from V to $\mathbb{R}^n$: $Tv_i := e_i$ for $1 \leq i \leq n$ and extended linearly.

Note the analogy. The choice of a basis of an arbitrary vector space gave rise to a linear isomorphism from the vector space onto some $\mathbb{R}^n$ while the choice of an orthonormal basis of V gives rise to an orthogonal linear isomorphism T from V onto $\mathbb{R}^n$.

5.8.1 Coordinates Associated with an Orthonormal Basis

Let V be a finite dimensional vector space over $\mathbb{R}$. Recall how the choice of a basis in V introduced a system of coordinates. Given any vector $v \in V$ we can write it in the form $v = \sum_i x_i v_i$. We call x_i the ith coordinate of v. This way we identify V with $\mathbb{R}^n$ via the linear isomorphism

$$v \mapsto (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Now assume that V is an inner product space and that $\{e_i\}_{i=1}^n$ is a fixed orthonormal basis of V. (If you desire you may assume that $V = \mathbb{R}^n$ with the dot product and that e_i's are the standard basis vectors.) Let x_i be the coordinates of a vector v with respect to this fixed ("standard") basis: $v = \sum x_i e_i$. Let $\{v_i\}_{i=1}^n$ be another orthonormal basis of V and $v = \sum y_i v_i$. We call y_i's the new coordinates. How are the old coordinates related to the new ones?

Recall that if $v = \sum x_i e_i$ then $x_i = \langle v, e_i \rangle$. Write $v_i = \sum_j v_{ij} e_j$ where $v_{ij} = \langle v_i, e_j \rangle$. If $v = \sum y_j v_j$, we then have

$$x_i = \left\langle \sum_j y_j v_j, e_i \right\rangle = \sum_j y_j \langle v_j, e_i \rangle = \sum_j y_j v_{ji}.$$

Do you see the pattern? If we form the matrix

$$K := \begin{pmatrix} v_{11} & v_{21} & \dots & v_{n1} \\ v_{12} & v_{22} & \dots & v_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1n} & v_{2n} & \dots & v_{nn} \end{pmatrix}$$

then we know that this matrix represents the orthogonal linear map which maps e_i to v_i. Then the coefficients of y_j's in the expression of x_i are the elements of the ith row in this matrix.

We shall put these observations into use in the section on classification of quadrics (Section 8.2).

5.9 Reflections and Orthogonal Maps of the Plane

V stands for an inner product space of dimension n.

5.9.1 Reflections

Let W be a subspace of V. Assume that $\dim W = n-1$. We wish to define *reflection* with respect to W. A little geometric thinking tells us that we must map each element of W to itself and the vectors perpendicular to W to their negatives. To put this in a precise form, we use the orthogonal decomposition: $V = W \oplus \mathbb{R}u$ where u is a unit vector orthogonal to W. Let $T : V \to V$ be a linear transformation defined by $T(w) = w$ for all $w \in W$ and $T(w') = -w'$ for all $w' \in \mathbb{R}u = W'$. If $v = w + w'$, where $w \in W$, $w' \in W'$, then $Tv = w - w'$ (see Figure 5.9.1).

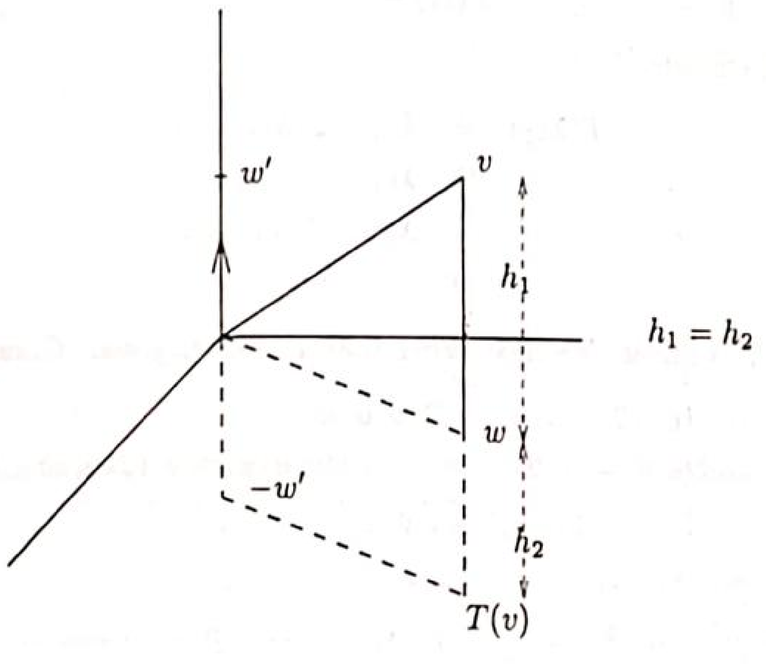


Figure 5.9.1 Reflection with respect to a hyperplane.

We now want to arrive at a neat formula for T. Note that $w' = \langle v, u\rangle u$. Hence

$$Tv = T(w + w') = w - w' = w + w' - 2w' = v - 2w' = v - 2\langle v, u\rangle u.$$

Definition 5.9.1 Let W be an $n-1$ dimensional vector subspace of an inner product space V. Let u be a unit vector perpendicular to W. Define $T : V \to V$ by $T(v) = v - 2\langle v, u\rangle u$. Then $R_W = T$ is called the *reflection* with respect to W.

Exercise 5.9.1 The definition of $R_W = T$ does not depend on the choice of the unit normal. (Recall that since $\dim W = n - 1$, $\dim W^\perp = 1$ and there are two unit vectors in $W^\perp$).

Proposition 5.9.1 *Let V be an inner product space of dimension n. Let W be a vector subspace of V with $\dim W = n - 1$. Let u be any unit vector orthogonal to W. Define T by $Tv = v - 2\langle v, u\rangle u$. Then T does not depend upon the choice of u. Further, T is an orthogonal linear transformation on V.*

Proof A routine verification. It is best if the reader does it on his own. Consider

$$\begin{aligned} T(v_1 + v_2) &= v_1 + v_2 - 2\langle v_1 + v_2, u\rangle u \\ &= v_1 + v_2 - 2\langle v_1, u\rangle u - 2\langle v_2, u\rangle u \\ &= v_1 - 2\langle v_1, u\rangle u + v_2 - 2\langle v_2, u\rangle u \\ &= T(v_1) + T(v_2). \end{aligned}$$

For $\lambda \in \mathbb{R}$, consider

$$\begin{aligned} T(\lambda v_1) &= \lambda v_1 - 2\langle \lambda v_1, u\rangle u \\ &= \lambda v_1 - 2\lambda\langle v_1, u\rangle u \\ &= \lambda(v_1 - 2\langle v_1, u\rangle u) \\ &= \lambda T(v_1). \end{aligned}$$

Therefore T is linear. We next prove that T is orthogonal. Consider

$$\begin{aligned} \langle Tv, Tv\rangle &= \langle v - 2\langle v, u\rangle u, v - 2\langle v, u\rangle u\rangle \\ &= \langle v, v\rangle - \langle v, 2\langle v, u\rangle u\rangle - \langle 2\langle v, u\rangle u, v\rangle + \langle 2\langle v, u\rangle u, 2\langle v, u\rangle u\rangle \\ &= \langle v, v\rangle - 2\langle v, u\rangle^2 - 2\langle v, u\rangle^2 + 4\langle v, u\rangle^2 \\ &= \langle v, v\rangle. \end{aligned}$$

Hence $||Tv||^2 = ||v||^2$. Hence $||Tv|| = ||v||$ and T is orthogonal.

□

The orthogonality of T can also be proved as follows: If $\{w_1, \ldots, w_{n-1}\}$ is an orthonormal basis of W, then $\{w_1, \ldots, w_{n-1}, u\}$ is an orthonormal basis of V and $Tw_i = w_i$, $1 \le i \le n-1$ and $Tu = -u$ so that T carries an orthonormal basis to an orthonormal basis.

□

We now find the matrix associated with the reflection ρ_x with respect to the subspace $\mathbb{R}e_1$ (that is, x-axis) in $\mathbb{R}^2$. Since

$$\rho_x(e_1) = e_1 \text{ and } \rho_x(e_2) = -e_2,$$

the matrix of ρ_x is

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Exercise 5.9.2 Show that $\rho_x(x, y) = (x, -y)$.

5.9.2 Orthogonal Maps of the Plane

We now look at the orthogonal linear maps of $\mathbb{R}^2$. If T is one such, let A denote its matrix with respect to the standard basis. Then A is a 2×2 orthogonal matrix. So, it suffices to find all orthogonal matrices in $M(2, \mathbb{R})$. Let

$$A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

be an orthogonal matrix. Therefore $AA^t = I$ and hence $a^2 + c^2 = 1$ and $b^2 + d^2 = 1$. Further $\langle (a, c), (b, d) \rangle = 0$. Now, $a^2 + c^2 = 1$ implies that there exists unique $\theta \in [0, 2\pi)$ such that $a = \cos\theta$, $c = \sin\theta$. Therefore $(a, c) = (\cos\theta, \sin\theta)$ and since $\langle (a, c), (b, d) \rangle = 0$, we get that $(b, d) = \pm(-\sin\theta, \cos\theta)$. Thus we have

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{5.9.1}$$

or

$$\rho_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}. \tag{5.9.2}$$

The transformation represented by Equation (5.9.1) is called a *rotation* by an angle θ (see Figure 5.9.2) and that represented by Equation (5.9.2) is called a *reflection* (see Figure 5.9.3).

The latter is called a reflection since it is the reflection with respect to the line $\mathbb{R}(\cos\frac{\theta}{2}e_1 + \sin\frac{\theta}{2}e_2) = \{(t\cos\frac{\theta}{2}, t\sin\frac{\theta}{2}) \mid t \in \mathbb{R}\}$. Let us denote this one-dimensional vector subspace by W. A unit normal u to this line is given by $u = (\sin\frac{\theta}{2}, -\cos\frac{\theta}{2})$ (a vector perpendicular to $(\cos\frac{\theta}{2}, \sin\frac{\theta}{2})$).

Let R_W be the reflection with respect to W. Let us compute

$$\begin{aligned} R_W(e_1) &= e_1 - 2\langle e_1, u\rangle u \\ &= (1, 0) - 2\sin\frac{\theta}{2}\left(\sin\frac{\theta}{2}, -\cos\frac{\theta}{2}\right) \\ &= \left(1 - 2\sin^2\frac{\theta}{2}, 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}\right) \\ &= (\cos\theta, \sin\theta). \end{aligned}$$

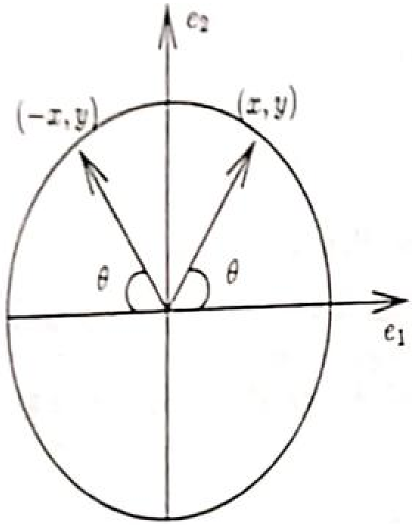


Figure 5.9.2 Rotation.

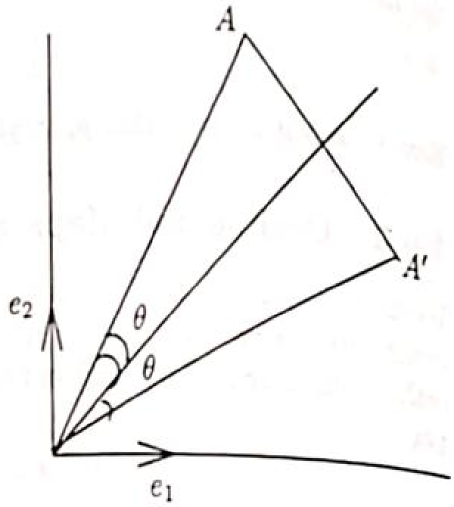


Figure 5.9.3 Reflection with respect to a line.

$$\begin{aligned}
R_W(e_2) &= e_2 - 2\langle e_2, u\rangle u \\
&= (0,1) + 2\cos\frac{\theta}{2}\left(\sin\frac{\theta}{2}, -\cos\frac{\theta}{2}\right) \\
&= \left(2\cos\frac{\theta}{2}\sin\frac{\theta}{2}, 1 - 2\cos^2\frac{\theta}{2}\right) \\
&= (\sin\theta, -\cos\theta).
\end{aligned}$$

Thus the matrix of R_W with respect to the standard orthonormal basis $\{e_1, e_2\}$ of $\mathbb{R}^2$ is $(R_W(e_1), R_W(e_2)) = \begin{pmatrix}\cos\theta & \sin\theta \\ \sin\theta & -\cos\theta\end{pmatrix}$. Hence the result.

We could have arrived at this result in a geometric way. To reflect with respect to the line W is same as the composition of the operations. Rotate $\mathbb{R}^2$ (and hence the line W) by an angle $-\theta/2$, reflect with respect to the x-axis and rotate back by an angle $\theta/2$. This composition is the product of the associated matrix

$$R_{\frac{\theta}{2}} \circ \begin{pmatrix}1 & 0 \\ 0 & -1\end{pmatrix} \circ R_{-\frac{\theta}{2}} = \rho_\theta.$$

6. Determinants

Recall that a determinant of order two is defined by the formula

$$\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = x_1 y_2 - x_2 y_1$$

In other words, the above determinant is a number assigned to the matrix

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}.$$

In this chapter, we define the determinant of a square matrix. This definition will be arrived at by treating the above concept in a geometric way.

6.1 2×2 Determinant as Area of a Parallelogram

From school geometry, one knows the areas of some elementary figures in the plane such as squares, rectangles, right-angled triangles etc. If one wants to find the areas of slightly more complicated subsets of $\mathbb{R}^2$, one appeals to intuitively satisfactory assumptions such as the following:

(1) The areas of rectangles, right-angled triangles are given by the well-known formulas.

(2) The area of any figure which consists only of line segments is zero. If A is the union of two subsets B and C such that $B \cap C$ is built up of line segments, then the area of A is the sum of the areas of B and C.

(3) If the given area is cut into elementary subsets of known area and rearranged, then the area of the original figure is the same as that of the rearranged figure.

Using such ideas we can arrive at the area of a parallelogram as in high school geometry. Consider $\mathbb{R}^2$. Let $x = (x_1, x_2)$, $y = (y_1, y_2) \in \mathbb{R}^2$ be any two nonzero vectors. We want to compute the area of the parallelogram spanned by these vectors. We compute its area in two ways. The first one will be of importance to us later.

Look at Figure 6.1.1. We can rearrange the given set into a rectangle by cutting away a right-angled triangle and pasting it on the opposite side. Hence the area of the parallelogram is "the base times the height". We find expressions for these in terms of the data given.

Let the (orthogonal) projection of y on x be z, that is, $z := y - \frac{\langle y, x\rangle}{\langle x, x\rangle} x$. Let $h = |z|$. From plane geometry, we have

$$\begin{aligned} h &= |y| \sin\theta \\ &= |y| \sqrt{1 - \cos^2\theta} \\ &= |y| \sqrt{1 - \frac{\langle x, y\rangle^2}{|x|^2 |y|^2}} \\ &= \frac{|y|}{|x|\,|y|} \sqrt{|x|^2 |y|^2 - \langle x, y\rangle^2} \\ &= \frac{1}{|x|} \sqrt{|x|^2 |y|^2 - \langle x, y\rangle^2}. \end{aligned}$$

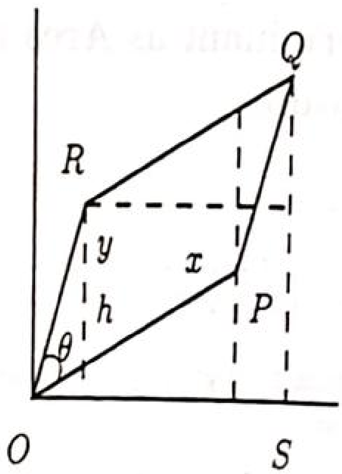


Figure 6.1.1 Determinant as signed area (1).

ɔw, the area of the parallelogram (see Figure 6.1.1)

$$A = (\text{base})h = |x| \frac{1}{|x|} \sqrt{|x|^2 |y|^2 - \langle x, y\rangle^2}.$$

Therefore,

$$\begin{aligned} A^2 &= |x|^2|y|^2 - \langle x, y\rangle^2 \\ &= (x_1y_2 - x_2y_1)^2 \\ &= (\det(x, y))^2 \end{aligned}$$

where (x, y) stands for the matrix whose first column is the column vector x and the second column is y. Hence the area of the parallelogram is $|x_1y_2 - x_2y_1|$. We may thus think of $\det(x, y)$ as the signed area of the parallelogram spanned by x and y.

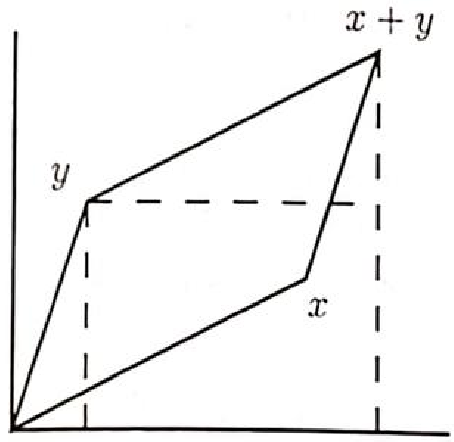


Figure 6.1.2 Determinant as signed area (2).

The second computation runs as follows: In Figure 6.1.2, the full area is

$$\frac{1}{2}y_1y_2 + y_2x_1 + \frac{1}{2}x_1x_2$$

where $x = (x_1, x_2)$, $y = (y_1, y_2)$.

In Figure 6.1.3, the area of the shaded portion is

$$\frac{1}{2}x_1x_2 + x_2y_1 + \frac{1}{2}y_1y_2.$$

Therefore the area of the parallelogram is $|x_1y_2 - x_2y_1| = \det(x, y)$.

Thus we may think of the determinant of a 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

as the *signed* area of the parallelogram spanned by the two column vectors

$$\begin{pmatrix} a \\ c \end{pmatrix} \text{ and } \begin{pmatrix} b \\ d \end{pmatrix} \text{ in } \mathbb{R}^2.$$

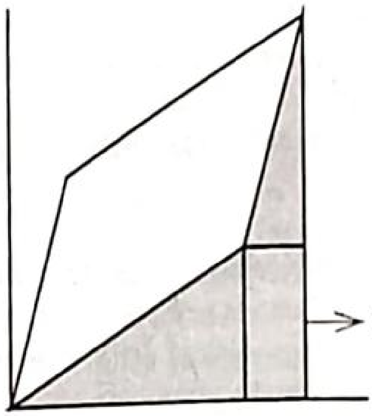


Figure 6.1.3 Area of a parallelogram.

Let us look at this view a little more closely. If the vectors are linearly dependent — in this case one is a multiple of the other — they "span" a one-dimensional area and hence the area is 0. This is true, since if we assume that $\begin{pmatrix} b \\ d \end{pmatrix}$ is a scalar multiple of $\begin{pmatrix} a \\ c \end{pmatrix}$ then

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = 0.$$

If the vectors are the standard basis, then they span the unit square whose area is 1. We immediately verify that this is so, as

$$\det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1.$$

If we dilate one of the sides, say, by a factor of δ, we expect that the area of the resulting parallelogram should be δ times that of the original figure:

$$\det \begin{pmatrix} \delta a & b \\ \delta c & d \end{pmatrix} = \delta \det \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

However, note that we have

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

Even though the parallelogram spanned by e_2 and e_1 is also the unit square, we get its area as -1! Thus it is clear that we are looking not only at the final geometric figure, namely, the parallelogram spanned by the vectors but also at the order in which the vertices are given.

Consider $\mathbb{R}^2$ with the standard basis $\{e_1, e_2\}$. Let A be the area of the unit square Q spanned by $\{e_1, e_2\}$. Let $v = \sum v_i e_i$ and $w = \sum w_i e_i$, $i = 1, 2$ be a basis of $\mathbb{R}^2$. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map such that $Te_1 = v$ and $Te_2 = w$. The matrix representation of T with respect to $\{e_1, e_2\}$ is

$$T = \begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}.$$

Then the area of the parallelogram $[v, w]$ spanned by v and w is given by $\text{Area}[v, w] = |\det(T)| = |\det(T)|\ A$. Thus the linear map T distorts the area of Q by the factor $|\det(T)|$, that is, Area $[v, w] = |\det(T)|$ Area Q (see Figure 6.1.4). If v_i are linearly dependent, then they span only a line segment not a two-dimensional object. Hence in this case the area of $T(Q)$ is zero. Thus $\det(T)$ is the factor by which the area of the unit square in $\mathbb{R}^2$ is multiplied to get the area of $T(Q)$. This is the geometric interpretation of the determinant of a square matrix of size 2. We shall return to this theme later.

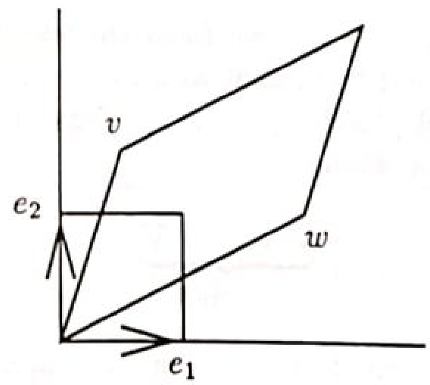


Figure 6.1.4 Geometric meaning of the determinant.

In the next section, we shall define the determinant based on these observations.

6.2 Determinant and its Properties

Let V be any vector space of dimension n. Let V^n stand for $V \times \cdots \times V$, the product of V with itself n times. The reader may assume that $V = \mathbb{R}^n$ if he so wishes.

We wish to define determinant as a function which attaches to any n-tuple of vectors $(v_1, \cdots, v_n) \in V^n$ a real number. This number is to be thought of as the signed volume of the parallelepiped spanned by v_i's. What is a parallelepiped in V?

Definition 6.2.1 Let $\{v_i\}_{i=1}^n \subseteq V$. The parallelepiped $[v_1, \dots, v_n]$ spanned by $\{v_i\}_{i=1}^n$ is defined as the set $\{v \in V \mid v = \sum_{i=1}^n \alpha_i v_i, 0 \le \alpha_i \le 1\}$.
The vectors v_i are called the *vertices* of the parallelepiped.

Note that all α_i are allowed to vary between 0 and 1. For instance, $v_1 + \dots + v_n \in [v_1, \dots, v_n]$.

When $n = 2$, a parallelepiped is a parallelogram. If $n = 3$, the picture is Figure 6.2.1.

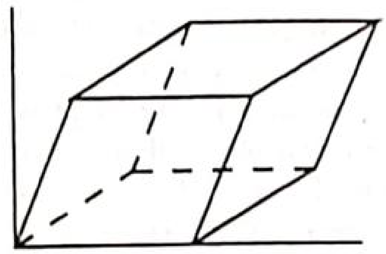


Figure 6.2.1 Parallelepiped in R^3.

Given n vectors $\{v_1, \dots, v_n\}$, we form the parallelepiped $[v_1, \dots, v_n]$. We want to think of the determinant as a function which assigns "signed" volume to each parallelepiped $[v_1, \dots, v_n]$. Thus the determinant should be a real valued function from

$$\underbrace{V \times \cdots \times V}_{n \text{ times}}.$$

In any kind of measurement we need a unit against which others are measured. In our case this means that we have to make a choice of a parallelepiped and declare its volume as 1. Will any n vectors $[v_1, \dots, v_n]$ do? No! For, if they are linearly dependent the parallelepiped lies in a vector subspace of dimension at most $n - 1$ and hence its n-dimensional volume must be zero. Thus, it behoves us to choose a basis $\{e_i\}_{i=1}^n$ of V. For the remaining part of this discussion, this basis will be fixed.

Based on our geometric intuition, we expect this function $\det : V^n \to \mathbb{R}$ to possess the following geometric properties:

(P1) For all $\alpha \in \mathbb{R}$, and for all i,

$$\det(v_1, v_2, \dots, \alpha v_i, \dots, v_n) = \alpha \det(v_1, \dots, v_i, \dots, v_n).$$

(P2) For $i \neq j$, $\det(v_1, \dots, v_i, v_j, \dots, v_n) = \det(v_1, \dots, v_i + v_j, v_j, \dots, v_n)$.

(P3) If $\{e_1, \dots, e_n\}$ is the chosen basis of V, then $\det(e_1, \dots, e_n) = 1$.

Remark 6.2.1 We make some remarks on the geometric contents of the conditions (P1) through (P3).

Condition (P1) is the mathematical rendition of the principle: The volume is magnified by a if any one side of the parallelepiped is magnified by a.

(P2) together with (P1) is the mathematical dressing of the principle: The volume or area is unaltered by cutting and rearranging to get a simpler geometric figure whose area or volume is easily determined. This is what we did while computing the area of the parallelogram geometrically (see (2) in Theorem 6.2.1).

(P3) is a normalization condition which is always needed in any kind of measurement.

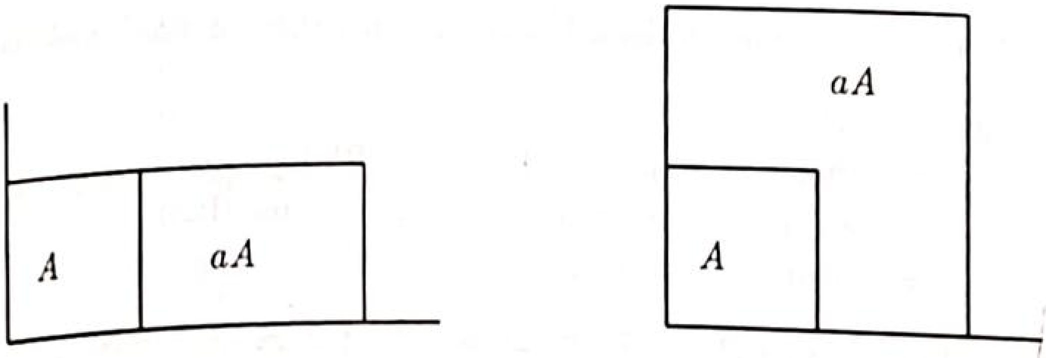


Figure 6.2.2 Magnification of geometric figures.

Assuming the existence of such a function "det", we derive some of its properties.

Theorem 6.2.1 *Assume that there exists a function* det: $V^n \to \mathbb{R}$ *with the properties (P1), (P2) and (P3) listed above. Then*

(1) $\det(x_1, \dots, x_i, \dots, x_j, \dots, x_n) = 0$ *if* $x_i = 0$ *for some* i.

(2) *For* $i \neq j$ *and* $\alpha \in \mathbb{R}$,

$$\det(v_1, \dots, v_i, v_j, \dots, v_n) = \det(v_1, \dots, v_i + \alpha v_j, \dots, v_n).$$

(2′) *More generally, we see that, for any* $\alpha_j \in \mathbb{R}$, $j \neq i$,

$$\det(v_1, \dots, v_i, \dots, v_n) = \det(v_1, \dots, v_{i-1}, v_i + \sum_{j \neq i} \alpha_j v_j, v_{i+1}, \dots, v_n). \tag{6.2.1}$$

(3) $\det(v_1, \dots, v_n) = 0$ *if* $\{v_1, \dots, v_n\}$ *is linearly dependent. In particular, if* $v_i = v_j$ *for some* $i \neq j$, *then* $\det(v_1, \dots, v_n) = 0$.

(4) *For any* $j \in \{1, \dots, n\}$ *and for any* v'_j, v''_j *we have*

$$\begin{aligned}\det(v_1, \dots, v'_j + v''_j, \dots, v_n) &= \\ &\det(v_1, \dots, v'_j, \dots, v_n) + \det(v_1, \dots, v''_j, \dots, v_n).\end{aligned}$$

(5) $\det(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -\det(v_1, \dots, v_j, \dots, v_i, \dots, v_n)$, *for* $i \neq j$.

Proof (1) is easy. For, in (P1) we can take $\alpha = 0$:

$$\det(x_1, \dots, \alpha x_i, \dots, x_n) = \alpha \det(x_1, \dots, x_i, \dots, x_n).$$

To prove (2), assume without loss of generality that $i < j$ for simplicity.

$$\begin{aligned}&\alpha \det(v_1, \dots, v_i, \dots, v_j, \dots, v_n) \\ &\quad = \det(v_1, \dots, \alpha v_j, \dots, v_n) \qquad \text{(by (P1))} \\ &\quad = \det(v_1, \dots, v_i + \alpha v_j, \alpha v_j, \dots, v_n) \qquad \text{(by (P2))} \\ &\quad = \alpha \det(v_1, \dots, v_i + \alpha v_j, \dots, v_j, \dots, v_n) \qquad \text{(by (P1))}.\end{aligned}$$

Since this is true for all $\alpha \in \mathbb{R}$, (2) follows. (2′) follows easily from (2).

We now prove (3). If $\{v_1, \dots, v_n\}$ is linearly dependent, then there exists i such that $v_i = \sum_{j \neq i} \alpha_j v_j$. Using (2) repeatedly, we get

$$\begin{aligned}&\det(v_1, \dots, v_i, \dots, v_n) \\ &\quad = \det(v_1, \dots, v_i - \alpha_j v_j, \dots, v_n) \qquad \text{for } j \neq i \text{ by (2)} \\ &\quad = \det(v_1, \dots, v_i - \alpha_j v_j - \alpha_k v_k, \dots, v_n) \qquad \text{for } k \neq i, j \text{ by (2)} \\ &\quad = \det(v_1, \dots, v_i - \sum_{j \neq i} \alpha_j v_j, \dots, v_n) \\ &\quad = \det(v_1, \dots, 0, \dots, v_n) \\ &\quad = 0 \qquad \text{by (1)}.\end{aligned}$$

We now prove (4). If v'_j and v''_j are both linearly dependent on

$$\{v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n\}$$

there is nothing to prove as both sides of the equation are zero by (3). Therefore assume that one of them is linearly independent of

$$\{v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n\}.$$

Assume without loss of generality that $\{v_1, \dots, v_{j-1}, v'_j, \dots, v_n\}$ is linearly independent. Hence it is a basis of V. So we can write $v''_j = \alpha_j v'_j + Y$ where

$$Y \in L(\{v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n\})$$

is the linear span of $\{v_i \mid 1 \leq i \leq n, i \neq j\}$. Therefore,

$$\begin{aligned}
&\det(v_1, \dots, v'_j + v''_j, \dots, v_n) \\
&\quad = \det(v_1, \dots, v'_j + \alpha_j v'_j + Y, \dots, v_n) \\
&\quad = \det(v_1, \dots, v'_j + \alpha_j v'_j, \dots, v_n) \text{ by } (2') \\
&\quad = \det(v_1, \dots, (1+\alpha_j) v'_j, \dots, v_n) \\
&\quad = (1+\alpha_j)\det(v_1, \dots, v'_j, \dots, v_n) \text{ by (P1)} \\
&\quad = \det(v_1, \dots, v'_j, \dots, v_n) + \alpha_j \det(v_1, \dots, v'_j, \dots, v_n) \\
&\quad = \det(v_1, \dots, v'_j, \dots, v_n) + \det(v_1, \dots, \alpha_j v'_j, \dots, v_n) \text{ by (2)} \\
&\quad = \det(v_1, \dots, v'_j, \dots, v_n) + \det(v_1, \dots, \alpha_j v'_j + Y, \dots, v_n) \text{ by } (2') \\
&\quad = \det(v_1, \dots, v'_j, \dots, v_n) + \det(v_1, \dots, v''_j, \dots, v_n).
\end{aligned}$$

We prove (5). We consider det on $(v_1, \dots, v_i + v_j, \dots, v_i + v_j, \dots, v_n)$ and use (4).

$$\begin{aligned}
0 &= \det(v_1, \dots, v_i + v_j, \dots, v_i + v_j, \dots, v_n) \\
&= \det(v_1, \dots, v_i + v_j, \dots, v_j, \dots, v_n) + \det(v_1, \dots, v_i + v_j, \dots, v_i, \dots, v_n) \\
&= \det(v_1, \dots, v_i, \dots, v_j, \dots, v_n) + \det(v_1, \dots, v_j, \dots, v_j, \dots, v_n) \\
&\quad + \det(v_1, \dots, v_i, \dots, v_i, \dots, v_n) + \det(v_1, \dots, v_j, \dots, v_i, \dots, v_n) \\
&= \det(v_1, \dots, v_i, \dots, v_j, \dots, v_n) + \det(v_1, \dots, v_j, \dots, v_i, \dots, v_n),
\end{aligned}$$

as $\det(v_1, \dots, v_i, \dots, v_i, \dots, v_n) = 0$ by (3).

□

Properties (P1) and (4) put together imply that the function det is "linear in each variable". This means that if we keep all variables v_j except the ith variable fixed, then it is a linear map in v_i. To be precise we have the following

Definition 6.2.2 Let V be any vector space. An r-linear map is a function $f\colon V^r \to \mathbb{R}$ such that for each i, $1 \leq i \leq r$, the following are true:

$$\begin{aligned}
f(v_1, \dots, v_i + w_i, \dots, v_r) &= f(v_1, \dots, v_i, \dots, v_r) + f(v_1, \dots, w_i, \dots, v_r) \\
f(v_1, \dots, \alpha v_i, \dots, v_r) &= \alpha f(v_1, \dots, v_i, \dots, v_r)
\end{aligned}$$

for all $v_j, w_i \in V$ and $\alpha \in \mathbb{R}$.

If f is 2-linear (respectively 3-linear), then we say that f is bilinear (respectively trilinear).

Example 6.2.1 Let V be an inner product space. Then the map

$$f(x, y) := \langle x, y \rangle$$

is a bilinear map.

Example 6.2.2 Let $V := \mathbb{R}^2$ and $f(x, y) = x_1y_2 - x_2y_1$. Then f is bilinear.

Proposition 6.2.2 *If* det: $V^n \to \mathbb{R}$ *exists satisfying the conditions P1 through P3, then* det *is n-linear.*

Proof If such a det exists, it enjoys the properties (1) through (5) listed in Theorem 6.2.1. The Proposition then follows from (P1) and (4).

□

A more general formulation of (5) is true and it involves certain facts about the group of permutations on n symbols. Let S_n denote the set of permutations (bijections) of the set $\{1, \ldots, n\}$. In order not to interrupt our discussion, we shall assume the following facts about S_n as known to the reader. Proofs may be found in any book on algebra or a book on group theory.

(1) The number of elements in S_n is $n!$.

(2) If σ and τ are in S_n, then their composition $\sigma \circ \tau$ also lies in S_n.

(3) If $\sigma \in S_n$ so does its inverse σ^{-1}.

(4) A standard way of writing $\sigma \in S_n$ is

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{pmatrix}.$$

(5) A permutation, which switches exactly two elements and leaves the rest unaffected is called a transposition. If σ is defined by $\sigma(k) = k$, if $k \neq i$ and $k \neq j$, $\sigma(i) = j$ and $\sigma(j) = i$, then in the above notation it is written as

$$\begin{pmatrix} 1 & \ldots & i & \ldots & j & \ldots & n \\ 1 & \ldots & j & \ldots & i & \ldots & n \end{pmatrix}.$$

However, in this case, we use a still shorter notation $\sigma := (ij)$.

(6) It is known that any permutation is a product (composition) of transpositions. While this product is not unique, the parity of the number of transpositions in any such product is well-defined — either it is always even or always odd.

(7) If we set $\operatorname{sign}(\sigma) := (-1)^r$ when σ is a product of r transpositions, then $\operatorname{sign}(\sigma)$ is well-defined. $\operatorname{sign}(\sigma)$ is called the *sign* of σ. In particular, the sign of a transposition is -1.

(8) If $\sigma, \tau \in S_n$, then $\operatorname{sign}(\sigma \circ \tau) = \operatorname{sign}(\sigma)\operatorname{sign}(\tau)$.

(9) Let $\tau \in S_n$ be fixed. Then the map $\sigma \mapsto \tau \circ \sigma$ is a bijection of S_n.

(10) The map $\sigma \mapsto \sigma^{-1}$ is a bijection of S_n.

An important consequence of (8) and (9) is the following observation which will be used many times in the sequel: Let $f\colon S_n \to \mathbb{R}$ be a function. Then

$$\sum_{\sigma \in S_n} f(\sigma) = \sum_{\sigma \in S_n} f(\sigma^{-1}) = \sum_{\sigma \in S_n} f(\sigma\tau), \quad \text{for any fixed } \tau \in S_n.$$

Definition 6.2.3 Let $f\colon V^r \to \mathbb{R}$ be an r-linear map. f is said to be *skew-symmetric* if $f(v_{\sigma(1)}, \cdots, v_{\sigma(r)}) = \operatorname{sign}(\sigma) f(v_1, \cdots, v_r)$ for all $\sigma \in S_r$.

Exercise 6.2.1 The function f in Example 6.2.2 is skew-symmetric and bilinear.

Proposition 6.2.3 *If* $\det\colon V^n \to \mathbb{R}$ *exists satisfying the conditions P1 through P3, then* $\det$ *is skew-symmetric. That is, for any permutation* $\sigma \in S_n$, *we have* $\det(v_1, \cdots, v_n) = \operatorname{sign}(\sigma)\det(v_{\sigma(1)}, \cdots, v_{\sigma(n)})$.

Proof This follows from property (5) in Theorem 6.2.1 and the facts (6) and (7) about S_n.

□

We have thus proved that if $f\colon V^n \to \mathbb{R}$ with the properties P1, P2 exists, then such an f is n-linear and skew-symmetric. (We have not used P3 so far.) The next result asserts the existence of such maps.

Theorem 6.2.4 *Fix a basis* $\{e_1, \cdots, e_n\}$ *of* V. *Then there exists a unique function*

$$g : \underbrace{V \times \cdots \times V}_{n \text{ times}} \to \mathbb{R}$$

such that

(1) *g is n-linear.*

(2) *g is skew-symmetric.*

(3) $g(e_1, \cdots, e_n) = 1$.

Proof Let $v_1, \dots, v_n \in V$. Write $v_i = \sum \alpha_{ij} e_j$. Then

$$\begin{aligned}
g(v_1, \dots, v_n) &= g\left(\sum_{j_1} \alpha_{1j_1} e_{j_1}, \dots, \sum_{j_n} \alpha_{nj_n} e_{j_n}\right) \\
&= \sum_{j_1} \alpha_{1j_1} g\left(e_{j_1}, \sum \alpha_{2j_2} e_{j_2}, \dots, \sum \alpha_{nj_n} e_{j_n}\right) \text{ by } n\text{-linearity} \\
&= \sum_{j_1, \dots, j_n} \alpha_{1j_1} \alpha_{2j_2} \cdots \alpha_{nj_n} g(e_{j_1}, e_{j_2}, \dots, e_{j_n}) \text{ by } n\text{-linearity} \\
&= \sum_{j_1, \dots, j_n} \alpha_{1j_1} \alpha_{2j_2} \cdots \alpha_{nj_n} \operatorname{sign} \begin{pmatrix} 1 & \dots & n \\ j_1 & \dots & j_n \end{pmatrix} g(e_1, \dots, e_n) \\
&= \left(\sum \operatorname{sign} \begin{pmatrix} 1 & \dots & n \\ j_1 & \dots & j_n \end{pmatrix} \alpha_{1j_1} \alpha_{2j_2} \cdots \alpha_{nj_n}\right) g(e_1, \dots, e_n).
\end{aligned}$$

This equation tells us that if a function g satisfies the properties (1) and (2) of the theorem, then it must be of the form

$$g(v_1, \dots, v_n) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \alpha_{1\sigma(1)} \cdots \alpha_{n\sigma(n)} g(e_1, \dots, e_n). \qquad (6.2.2)$$

We now show that if g is defined by setting

$$g(v_1, \dots, v_n) := \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \alpha_{1\sigma(1)} \cdots \alpha_{n\sigma(n)} \qquad (6.2.3)$$

where $v_i = \sum_{j=1}^n \alpha_{ij} e_j$, then g has the properties (1)–(3).

We show that $g(v_1 + v_1', v_2, \dots, v_n) = g(v_1, \dots, v_n) + g(v_1', \dots, v_n)$. Let $w_1 = v_1 + v_1'$ and $w_r = v_r$ for $r \geq 2$. If we write $w_i = \sum_j \beta_{ij} e_j$, then $w_{1j} = \alpha_{1j} + \alpha_{1j}'$ where $v_1' = \sum_j \alpha_{1j}' e_j$. We have

$$\begin{aligned}
g(w_1, \dots, w_n) &= \sum_{\sigma} \operatorname{sign}(\sigma) \beta_{1\sigma(1)} \cdots \beta_{n\sigma(n)} \\
&= \sum_{\sigma} \operatorname{sign}(\sigma) (\alpha_{1\sigma(1)} + \alpha_{1\sigma(1)}') \alpha_{2\sigma(2)} \cdots \alpha_{n\sigma(n)} \\
&= \sum_{\sigma} \operatorname{sign}(\sigma) \alpha_{1\sigma(1)} \cdots \alpha_{n\sigma(n)} + \sum_{\sigma} \operatorname{sign}(\sigma) \alpha_{1\sigma(1)}' \cdots \alpha_{n\sigma(n)} \\
&= g(v_1, \dots, v_n) + g(v_1', \dots, v_n).
\end{aligned}$$

A similar computation shows that $g(\alpha v_1, v_2, \dots, v_n) = \alpha g(v_1, \dots, v_n)$ for any $\alpha \in \mathbb{R}$. Thus g is linear in the first variable. One can either prove its linearity in the other variables in a similar way, or derive it from the fact that g satisfies (2) of the theorem.

To prove (2), let $\tau = (ij)$ be a transposition with $i < j$. Let $w_r := v_{\tau(r)}$. We must show that $g(w_1, \dots, w_n) = -g(v_1, \dots, v_n)$. Let us write

$$w_k = \sum_r \beta_{kr} e_r.$$

Then we have

$$\beta_{kr} = \begin{cases} \alpha_{kr} & \text{if } k \neq i \text{ and } k \neq j \\ \alpha_{jr} & \text{if } k = i \\ \alpha_{ir} & \text{if } k = j. \end{cases}$$

We use this below:

$$\begin{aligned} g(w_1, \dots, w_n) &= \sum_\sigma \operatorname{sign}(\sigma) \beta_{1\sigma(1)} \cdots \beta_{n\sigma(n)} \\ &= \sum_\sigma \operatorname{sign}(\sigma) \alpha_{1\sigma(1)} \cdots \alpha_{j\sigma(i)} \cdots \alpha_{i\sigma(j)} \cdots \alpha_{n\sigma(n)} \\ &= \sum_\sigma \operatorname{sign}(\sigma) \alpha_{1\sigma(1)} \cdots \alpha_{i\sigma(j)} \cdots \alpha_{j\sigma(i)} \cdots \alpha_{n\sigma(n)} \\ &= \sum_\sigma \operatorname{sign}(\sigma) \alpha_{1\sigma\tau(1)} \cdots \alpha_{i\sigma\tau(i)} \cdots \alpha_{j\sigma\tau(j)} \cdots \alpha_{n\sigma\tau(n)} \\ &= \sum_{\sigma\tau} \operatorname{sign}(\sigma\tau) \operatorname{sign}(\tau) \alpha_{1\sigma\tau(1)} \cdots \alpha_{i\sigma\tau(i)} \cdots \\ &\qquad \alpha_{j\sigma\tau(j)} \cdots \alpha_{n\sigma\tau(n)} \\ &= \operatorname{sign}(\tau) \sum_{\eta \in S_n} \operatorname{sign}(\eta) \alpha_{1\eta(1)} \cdots \alpha_{n\eta(n)} \\ &= -g(v_1, \dots, v_n). \end{aligned}$$

Thus g satisfies (2). (3) is easy and left to the reader.

We have thus proved that there exists a function g as required and it is given by Equation (6.2.3).

□

Definition 6.2.4 We call the g of Theorem 6.2.4 as the *determinant* and denote it by det.

For the purpose of easy reference we list the following properties of the function *det* in the form of a theorem:

Theorem 6.2.5 *Let V be a vector space. Fix a basis $\{e_1, \dots, e_n\}$ of V. Define* $\det: V^n \to \mathbb{R}$ *by*

$$\det(v_1, \dots, v_n) := \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \alpha_{1\sigma(1)} \cdots \alpha_{n\sigma(n)}$$

for $v_i = \sum_{j=1}^n \alpha_{ij} e_j$. *Then* $\det : V^n \to \mathbb{R}$ *satisfies the following properties:*

(1) det *is linear in each of its variables.*

(2) *If one interchanges* v_i *and* v_j *for* $i \neq j$, *then the determinants are of opposite sign. More generally,*

$$\det(v_{\sigma(1)}, \dots, v_{\sigma(n)}) = \operatorname{sign}(\sigma)\det(v_1, \dots, v_n) \textit{ for any } \sigma \in S_n.$$

(3) $\det(e_1, \dots, e_n) = 1$.

(4) $\det(v_1, \dots, v_n) = 0$ *if* v_i *are linearly dependent.*

Proof This is left as an instructive exercise to the reader.

□

Exercise 6.2.2 Show that any map $f\colon V^n \to \mathbb{R}$ which is n-linear and skew-symmetric is of the form $f = f(e_1, \dots, e_n)\det$. *Hint:* This is already solved in the proof of Theorem 6.2.4.

Definition 6.2.5 We define the determinant function on $M(n, \mathbb{R})$ as follows: Let $A \in M(n, \mathbb{R})$ and write $A = (C_1, \dots, C_n)$ where C_i is the ith column of A. We then define

$$\det A := \det(C_1, \dots, C_n)$$

where det is the n-linear skew-symmetric function on $\underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ times}} \to \mathbb{R}$ with $\det(e_1, \dots, e_n) = 1$. (Here $\{e_1, \dots, e_n\}$ is the standard basis of $\mathbb{R}^n$). Note that $\det A = \det(Ae_1, \dots, Ae_n)$ for $A \in M(n, \mathbb{R})$.

If V is assumed to be an inner product space, then there are natural choices of the basis, namely, we would like to choose an orthonormal basis $\{e_i\}$ so that $[e_1, \dots, e_n]$ is "the unit cube". In this case it is natural to demand that its volume be 1, which is nothing but P3!

In the next section we show how to compute the determinant using only the properties of det listed in Theorem 6.2.5.

6.3 Computation of Determinants

In this section, we illustrate how to compute the determinants of matrices. We consider vectors of $\mathbb{R}^n$ as column vectors. We write the given matrix A as $A = (C_1, \dots, C_n)$ where C_i is the ith column of A. Then $\det(A)$ is defined by the formula

$$\det(A) = \det(C_1, \dots, C_n) \quad \text{where } C_i \in \mathbb{R}^n.$$

In the first of the computations, we explain which property of the determinant (in Theorem 6.2.5) function is used.

Example 6.3.1 Let $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ be a 2×2 matrix. Then

$$\begin{aligned}
&\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \\
&= \det(a_{11}e_1 + a_{21}e_2, a_{12}e_1 + a_{22}e_2) && \text{(by definition)} \\
&= \det(a_{11}e_1, a_{12}e_1 + a_{22}e_2) + \det(a_{21}e_2, a_{12}e_1 + a_{22}e_2) && \text{by (1)} \\
&= \det(a_{11}e_1, a_{12}e_1) + \det(a_{11}e_1, a_{22}e_2) && \text{by (1)} \\
&\qquad + \det(a_{21}e_2, a_{12}e_1) + \det(a_{21}e_2, a_{22}e_2) && \text{by (1)} \\
&= \det(a_{11}e_1, a_{22}e_2) + \det(a_{21}e_2, a_{12}e_1) && \text{by (2)} \\
&= a_{11}a_{22}\det(e_1, e_2) + a_{21}a_{12}\det(e_2, e_1) && \text{by (1)} \\
&= a_{11}a_{22} - a_{21}a_{12} && \text{by (2) and (3)}
\end{aligned}$$

Example 6.3.2 We compute the determinant of a diagonal matrix:

$$\det \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} \begin{aligned} &= \det(a_{11}e_1, a_{22}e_2, \dots, a_{nn}e_n) \\ &= a_{11}a_{22}\cdots a_{nn}\det(e_1, \dots, e_n) \\ &= a_{11}a_{22}\cdots a_{nn}. \end{aligned}$$

Example 6.3.3 We now find the determinant of a triangular matrix:

$$\begin{aligned}
&\det \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \\
&= \det\,(a_{11}e_1 + \dots + a_{n1}e_n, a_{22}e_2 + \dots + a_{n2}e_2, \dots, \\
&\qquad a_{n-1\,n-1}e_{n-1} + a_{n-1\,n}e_n, a_{nn}e_n) \\
&= \det(v_1, \dots, v_n),
\end{aligned}$$

in an obvious notation.

If we expand the above using multilinearity, say, in the first variable then the only term which contributes is e_1:

$$\det(v_1, v_2, \dots, v_n) = \det(a_{11}e_1, v_2, \dots, v_n).$$

Proceeding this way, we see that

$$\begin{aligned}
\det(a_{11}e_1 + \dots + a_{n1}e_1, a_{22}e_2 + \dots + a_{n2}e_n, \dots, a_{n1}e_1 + \dots + a_{nn}e_n) \\
= \det(a_{11}e_1, \dots, a_{nn}e_n) \\
= a_{11}a_{22}\cdots a_{nn}\det(e_1, \dots, e_n) \\
= a_{11}\cdots a_{nn}.
\end{aligned}$$

Let us do a couple of numerical examples.

Example 6.3.4

$$
\begin{aligned}
\det\begin{pmatrix}1&0&0&0\\0&0&1&0\\0&0&0&1\\0&1&0&0\end{pmatrix} &= \det(e_1,e_4,e_2,e_3)\\
&= -\det(e_1,e_2,e_4,e_3)\\
&= \det(e_1,e_2,e_3,e_4)\\
&= 1.
\end{aligned}
$$

Example 6.3.5

$$
\begin{aligned}
\det\begin{pmatrix}1&5&0&0\\2&0&8&0\\3&6&9&0\\4&7&10&1\end{pmatrix} &= \det\left(e_1+2e_2+3e_3+4e_4, 5e_1+6e_3+7e_4,\right.\\
&\qquad \left. 8e_2+9e_3+10e_4, e_4\right)\\
&= \det(3e_3,5e_1,8e_2,e_4)+\det(2e_2,5e_1,9e_3,e_4)\\
&\qquad +\det(e_1,6e_3,8e_2,e_4)\\
&= 8\times 3\times 5\times 1\times\det(e_1,e_2,e_3,e_4)\\
&\qquad -2\times 5\times 9\times 1\times\det(e_1,e_2,e_3,e_4)\\
&\qquad -1\times 6\times 8\times 1\det(e_1,e_2,e_3,e_4)\\
&= 48-90+120=-18.
\end{aligned}
$$

Exercise 6.3.1 Compute the determinant of $\begin{pmatrix}a&b&c\\u&v&w\\x&y&z\end{pmatrix}$.

We look at slightly more sophisticated examples.

Example 6.3.6 We shall give the beginning of computation only, leaving the rest to the reader. It may be a good idea to look at special cases $n=2,3$ and gain insight into what is happening.

$$
\begin{aligned}
&\det\begin{pmatrix}1+a_1&a_2&\dots&a_n\\a_1&1+a_2&\dots&a_n\\\vdots&\vdots&\ddots&\vdots\\a_1&a_2&\dots&1+a_n\end{pmatrix}\\
&= \det\left(e_1+\sum_{j_1}a_1e_{j_1}, e_2+\sum_{j_2}a_2e_{j_2},\dots,e_n+\sum_{j_n}a_ne_{j_n}\right)\\
&\qquad\times(1+a_1+\dots+a_n).
\end{aligned}
$$

Example 6.3.7 The matrix of this example is from differential geometry and its computation is required while computing the volume element of a (hyper) surface given as a graph. The matrix is

$$\begin{pmatrix} 1+x_1^2 & x_1x_2 & \dots & x_1x_n \\ x_2x_1 & 1+x_2^2 & \dots & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 & x_nx_2 & \dots & 1+x_n^2 \end{pmatrix}.$$

The trick here is, as in the last case, to realize the ith column vector C_i, which is the vector $e_i + x_i x$, where

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_i x_i e_i \in \mathbb{R}^n.$$

6.4 Basic Results on Determinants

In this section we establish all the standard results one needs about determinants.

Recall our definition of $\det(A)$ for $A \in M(n, \mathbb{R})$. Let $A = (A_1, \dots, A_n)$ where A_i are columns of A and hence we may consider

$$A := (A_1, \dots, A_n) \in \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ times}}.$$

It is also worth noting that $A_i = Ae_i$ where the right side is the matrix multiplication of A with the column vector e_i. Thus $A = (Ae_1, \dots, Ae_n)$. These facts will be used below without explicit mention.

Theorem 6.4.1 $\det(AB) = (\det A)(\det B)$.

Proof The most elegant proof runs as follows: Consider

$$f : \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ times}} \to \mathbb{R}$$

defined by

$$f(B) = \det(AB). \tag{6.4.1}$$

Then one shows that f satisfies the properties (1) and (2) of Theorem 6.2.4. By Exercise 6.2.2 it follows that f is given by $f(B) = f(I) \det B$. But $f(I) = \det(AI) = \det A$. Therefore,

$$f(B) = \det A \det B. \tag{6.4.2}$$

From Equations (6.4.1) and (6.4.2), we get $\det(AB) = (\det A)(\det B)$.

We now give a slightly different proof.

Consider

$$\phi(x_1, \ldots, x_n) = \det A \det(x_1, \ldots, x_n) - \det(Ax_1, \ldots, Ax_n),$$

where $x_i \in \mathbb{R}^n$. Then

(1) ϕ is linear in each x_i. This is easily seen since det is linear in each of its variables and A is linear. For instance,

$$\begin{aligned}
&\phi(x_1 + x_1', \ldots, x_n) \\
&\quad = \det A \det(x_1 + x_1', \ldots, x_n) - \det(A(x_1 + x_1'), \ldots, Ax_n) \\
&\quad = (\det A \det(x_1, \ldots, x_n) + \det A \det(x_1', \ldots, x_n)) \\
&\qquad - (\det(Ax_1, \ldots, Ax_n) + \det(Ax_1', \ldots, Ax_n)) \\
&\quad = \det A \det(x_1, \ldots, x_n) - \det(Ax_1, \ldots, Ax_n) \\
&\qquad + \det A \det(x_1', \ldots, x_n) - \det(Ax_1', \ldots, Ax_n) \\
&\quad = \phi(x_1, x_2, \ldots, x_n) + \phi(x_1', x_2, \ldots, x_n).
\end{aligned}$$

(2) $\phi(e_{i_1}, \ldots, e_{i_n}) = 0$, where $e_{i_1}, \ldots, e_{i_n}$ are any n vectors from the set of standard basis vectors. For, if $e_{i_j} = e_{i_k}$ for $j \neq k$, then both $\det(e_{i_1}, \ldots, e_{i_n})$ and $\det(Ae_{i_1}, \ldots, Ae_{i_n})$ are 0. Otherwise, $\{e_{i_1}, \ldots, e_{i_n}\}$ is obtained from $\{e_1, \ldots, e_n\}$ by a permutation σ. Hence

$$\begin{aligned}
\det A \det(e_{i_1}, \ldots, e_{i_n}) &= \operatorname{sign}(\sigma) \det A \det(e_1, \ldots, e_n) \\
&= \operatorname{sign}(\sigma) \det A \\
&= \operatorname{sign}(\sigma) \det(Ae_1, \ldots, Ae_n) \\
&= \operatorname{sign}(\sigma) \operatorname{sign}(\sigma^{-1}) \det(Ae_{i_1}, \ldots, Ae_{i_n}) \\
&= \det(Ae_{i_1}, \ldots, Ae_{i_n}).
\end{aligned}$$

Consequently $\phi(e_{i_1}, \ldots, e_{i_n}) = 0$ in any case.

Write $B = (B_1, \ldots, B_n)$ where B_i are the columns of B. Then we claim that $\phi(B_1, \ldots, B_n) = 0$. For,

$$\begin{aligned}
\phi(B_1, \ldots, B_n) &= \phi\left(\sum b_{i_1 1} e_{i_1}, \ldots, \sum b_{i_n n} e_{i_n}\right) \\
&= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} b_{i_1 1} \cdots b_{i_n n} \phi(e_{i_1}, \ldots, e_{i_n}) \\
&= 0.
\end{aligned}$$

This implies

$$\begin{aligned}
0 = \phi(B_1, \ldots, B_n) &= \det A \det(B_1, \ldots, B_n) - \det(AB_1, \ldots, AB_n) \\
&= \det A \det B - \det(AB)
\end{aligned}$$

since $\det(AB) = \det(AB(e_1), \ldots, AB(e_n)) = \det(AB_1, \ldots, AB_n)$.

□

Corollary 6.4.2 *Let $A, X \in M(n, \mathbb{R})$. Let A be invertible. Then we have*

(1) $\det(A^{-1}) = \det(A)^{-1}$.

(2) $\det(AXA^{-1}) = \det(X)$.

Proof Since $AA^{-1} = I$, the identity, (1) follows from Theorem 6.4.1 since $\det(I) = 1$. (2) follows from this and Theorem 6.4.1.

□

Theorem 6.4.3 *Let $A \in M(n, \mathbb{R})$. Then*

(1) *A is invertible if and only if the columns A_i of A are linearly independent.*

(2) *$\det A = 0$ if and only if columns of A are linearly dependent.*

(3) *$\det A \neq 0$ if and only if A is invertible.*

Proof Let the columns A_i of A be linearly independent. Then

$$\{A_i = Ae_i \mid 1 \leq i \leq n\}$$

is a basis of $\mathbb{R}^n$ where $\{e_i\}_{i=1}^n$ is the standard basis of $\mathbb{R}^n$. Therefore there exist scalars β_{ij} such that, $e_i = \sum_j \beta_{ij} A_j$. We claim that $B = (\beta_{ij})$ is the inverse of A:

$$e_i = \sum_j \beta_{ij} A_j = \sum_j \beta_{ij} Ae_j = \sum_j \beta_{ij} \left(\sum_k a_{jk} e_k \right) = \sum_k \left(\sum_j \beta_{ij} a_{jk} \right) e_k.$$

Since $\{e_i\}$ is a basis, by uniqueness, we see that $\sum_j \beta_{ij} a_{jk} = \delta_{ik}$ and hence the claim. Therefore A is invertible with B as its inverse.

Conversely, if A is invertible, then A maps a basis of $\mathbb{R}^n$ to another basis. But $A_i = Ae_i$ and hence $\{A_i\}$ is a basis of $\mathbb{R}^n$. This proves (1).

From the properties of the determinant we know that if the columns of A are linearly dependent, $\det A = 0$. Conversely, assume that $\det A = 0$ and suppose columns of A are linearly independent. From (1) we know that A has an inverse B. Therefore $1 = \det I = \det(AB) = \det A \det B$. It follows that $\det A \neq 0$, a contradiction. Hence the columns of A are linearly dependent. This proves (2).

(3) is an immediate consequence of the first two assertions: $\det A \neq 0$ if and only if the columns A_i are linearly independent (by the second assertion) which is true if and only if A is invertible.

□

Theorem 6.4.4 *Let $A \in M(n,\mathbb{R})$. Then $Ax = 0$ has a nonzero solution $x \in \mathbb{R}^n$ if and only if* $\det A = 0$.

Proof Suppose $Ax = 0$ has a nonzero solution

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Then

$$(A_1 A_2 \cdots A_n)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

if and only if $\sum_{j=1}^{n} x_j A_j = 0$. This happens if and only if A_i's are linearly dependent which is if and only if $\det A = 0$, by Theorem 6.4.3.

□

Definition 6.4.1 Let $T\colon V \to V$ be a linear map. We fix a basis $\{v_i\}_{i=1}^{n}$ of V. We define $\det T := \det M_v^v(T)$. Is this well-defined? That is, if we choose another basis $\{u_1, \dots, u_n\}$ of V and set $\det T := \det M_u^u(T)$, we need to show that $M_v^v(T) = M_u^u(T)$. This is an easy consequence of Corollary 6.4.2. The matrices are related by a conjugation by an invertible matrix taking one of these bases to the other. (Exercise: Work out the details.)

Corollary 6.4.5 *$T\colon V \to V$ has a nonzero kernel if and only if* $\det T := 0$.

Proof This is an easy consequence of Theorem 6.4.4.

□

Lemma 6.4.6 *If A^t denotes the transpose of the matrix A,* $\det A = \det A^t$.

Proof Let $A = (\alpha_{ij})$ and let σ be a permutation. Since

$$\alpha_{1\sigma(1)}\alpha_{2\sigma(2)}\cdots\alpha_{n\sigma(n)} = \alpha_{\sigma^{-1}(1)1}\alpha_{\sigma^{-1}(2)2}\cdots\alpha_{\sigma^{-1}(n)n}.$$

Note that σ^{-1} runs through S_n as σ varies and that $\operatorname{sign}(\sigma) = \operatorname{sign}(\sigma^{-1})$. Hence

$$\begin{aligned} \det A &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma)\alpha_{1\sigma(1)}\alpha_{2\sigma(2)}\cdots\alpha_{n\sigma(n)} \\ &= \sum_{\sigma^{-1} \in S_n} \operatorname{sign}(\sigma^{-1})\alpha_{\sigma^{-1}(1)1}\alpha_{\sigma^{-1}(2)2}\cdots\alpha_{\sigma^{-1}(n)n} \\ &= \sum \operatorname{sign}(\tau)\alpha_{\tau(1)1}\cdots\alpha_{\tau(n)n} \\ &= \det A^t. \end{aligned}$$

6.4.1 Laplace Expansion

Laplace expansion shows how to reduce the evaluation of the determinant of an $n \times n$ matrix to that of an $(n-1) \times (n-1)$ matrix.

Let $A = (\alpha_{ij})_{1 \le i,j \le n}$ be given. Fix i. Let us write the ith row R_i as $R_i = \alpha_{i1}e_1 + \cdots + \alpha_{in}e_n$. We expand

$$\begin{aligned}\det A &= \det\left(R_1, \ldots, R_{i-1}, \sum_{j=1}^{n} \alpha_{ij}e_j, R_{i+1}, \ldots, R_n\right)\\ &= \sum_{j=1}^{n} \alpha_{ij} \det(R_1, \ldots, R_{i-1}, e_j, R_{i+1}, \ldots, R_n) = \sum_{j=1}^{n} \alpha_{ij}A^{\bullet}_{ij}, \quad \text{say.}\end{aligned}$$

Now, if we can subtract e_j from any of the R_k, $\det A$ remains unchanged. Thus $A^{\bullet}_{ij}$ is the determinant of the matrix obtained from A by changing the entry α_{ij} to 1, and all other entries in the ith row and the jth column to 0. Laplace's expansion says that $A^{\bullet}_{ij}$ is (upto sign) the determinant of the $(n-1) \times (n-1)$ matrix obtained from A by deleting the ith row and jth column.

Definition 6.4.2 Let A_{ij} denote the $(n-1)$ square matrix obtained by deleting the ith row and the jth column. $\det A_{ij}$ is called the *minor* of α_{ij} of A and the *cofactor* C_{ij} is by definition $C_{ij} := (-1)^{i+j} \det A_{ij}$.

Theorem 6.4.7 (Laplace Expansion) *For an $n \times n$ matrix $A = (\alpha_{ij})$,*

$$\det A = \alpha_{i1}C_{i1} + \alpha_{i2}C_{i2} + \cdots + \alpha_{in}C_{in}, \tag{6.4.3}$$

where C_{ij} is the cofactor of α_{ij}.

Proof Recall the explicit expression for $\det A$:

$$\det A = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma)\alpha_{1\sigma(1)}\alpha_{2\sigma(2)} \cdots \alpha_{n\sigma(n)}.$$

Each term in the right side above contains exactly one element from the ith row $(\alpha_{i1}, \ldots, \alpha_{in})$, of A. Hence we can write $\det A = \alpha_{i1}A^{\bullet}_{i1} + \cdots + \alpha_{in}A^{\bullet}_{in}$. Then $A^{\bullet}_{ij}$ is the sum of the terms having no entry from the ith row. We need only show $A^{\bullet}_{ij} = (-1)^{i+j} \det A_{ij}$.

Let us first look at a special case where $i = n$ and $j = n$. Then

$$\alpha_{nn}A^{\bullet}_{nn} = \alpha_{nn} \sum \operatorname{sign}(\sigma)\alpha_{1\sigma(1)} \cdots \alpha_{n-1\sigma(n-1)}$$

where the sum is over all permutations $\sigma \in S$ which leaves n fixed: $\sigma(n) = n$. But this is the same as summing over all permutations from S_{n-1} so that

$$A^{\bullet}_{nn} = \det A_{nn} = (-1)^{n+n} \det A_{nn}.$$

Now to the general case. We bring the ith row to the nth row by interchanging successively with succeeding rows. Similarly, the jth column is brought to the nth column. In this process, observe that $\det A_{ij}$ is not altered (as the positions of other rows and columns do not get affected). But the sign of $\det A$ (and hence that of A^*_{ij}) is changed by

$$(-1)^{n-i} \cdot (-1)^{n-j} = (-1)^{i+j}.$$

Hence $A^*_{ij} = (-1)^{i+j} \det A_{ij}$.

□

The expansion in Equation (6.4.3) is known as the expansion by the ith row. Similarly one can expand $\det(A)$ by its jth column.

Exercise 6.4.1 As an immediate exercise, prove the Laplace expansion by the jth column:

$$\det A = (-1)^{1+j} a_{1j} \det A_{1j} + \cdots + (-1)^{n+j} a_{nj} \det A_{nj}.$$

Theorem 6.4.7 yields many dividends.

Proposition 6.4.8 *If $A = (a_{ij})$ is an $n \times n$ matrix, then*

$$\begin{aligned} a_{i1}C_{k1} + a_{i2}C_{k2} + \cdots + a_{in}C_{kn} &= 0 \text{ if } i \neq k && (6.4.4)\\ &= \det A \text{ if } i = k && (6.4.5)\end{aligned}$$

and

$$\begin{aligned} a_{1j}C_{1k} + a_{2j}C_{2k} + \cdots + a_{nj}C_{nk} &= 0 \text{ if } j \neq k && (6.4.6)\\ &= \det A \ \text{ if } j = k. && (6.4.7)\end{aligned}$$

Proof Consider the matrix B obtained from A by replacing the ith row of A by its kth row. If $i = k$, then $B = A$ and the result follows from Equation (6.4.3). If $i \neq k$, $\det B = 0$ (by Theorem 6.4.3 and Lemma 6.4.6). The kth row of B is $(a_{i1}, \ldots, a_{in})$ and the cofactors are $A_{k1}, \ldots, A_{kn}$. We use Equation (6.4.3) to expand B by its kth row to get

$$\det B = a_{i1}C_{k1} + \cdots + a_{in}C_{kn}.$$

The second result is obtained using Exercise 6.4.1.

□

Let us write Equations (6.4.4) and (6.4.5) in a compact form.

$$\sum_j a_{ij}(-1)^{k+j} \det A_{kj} = \delta_{ik} \det A. \qquad (6.4.8)$$

This leads us to define a new matrix $\widetilde{A}$ whose (jk)th entry is $(-1)^{k+j}\det A_{kj}$. Thus Equation (6.4.8) says that the product $A\widetilde{A}$ is $(\det A)I$. Using the column versions of Equations (6.4.6) and (6.4.7), we get

$$\widetilde{A}A = \det AI.$$

This new matrix $\widetilde{A}$ is called the *adjunct* of A. It is denoted by $\mathrm{adj}\,(A)$. (We avoid naming it adjoint, as there is another more widely used concept which is also called adjoint.)

An important consequence of this is that we get a formula for the inverse of A:

$$A^{-1} = \frac{1}{\det A}\widetilde{A}$$

if $\det A \neq 0$ (which is same as saying A^{-1} exists).

Let us put the results we have learnt into use.

Example 6.4.1 Let

$$A := \begin{pmatrix} 0 & -1 & 3 \\ 2 & 5 & -4 \\ -3 & 7 & 1 \end{pmatrix}.$$

Then we find

$$A_{11} = \det\begin{pmatrix} 5 & -4 \\ 7 & 1 \end{pmatrix} = 33.$$

Similarly, we find

$$\begin{array}{lllllllll} A_{11} & = & 33 & A_{12} & = & -10 & A_{13} & = & 29 \\ A_{21} & = & -22 & A_{22} & = & 9 & A_{23} & = & -3 \\ A_{31} & = & -11 & A_{32} & = & 6 & A_{33} & = & 2. \end{array}$$

Hence the matrix

$$(-1)^{i+j}A_{ij} = \begin{pmatrix} 33 & 10 & 29 \\ 22 & 9 & 3 \\ -11 & 6 & 2 \end{pmatrix}.$$

Its transpose is

$$(-1)^{i+j}A_{ji} = \begin{pmatrix} 33 & 22 & -11 \\ 10 & 9 & 6 \\ 29 & 3 & 2 \end{pmatrix}.$$

Let us compute the determinant of A: We use the standard notation such as $C_i + \alpha C_j$ means that α times the jth column C_j is added to the ith column.

$$A = \begin{pmatrix} 0 & -1 & 3 \\ 2 & 5 & -4 \\ -3 & 7 & 1 \end{pmatrix} \xrightarrow{C_3+3C_2} \begin{pmatrix} 0 & -1 & 0 \\ 2 & 5 & 11 \\ -3 & 7 & 22 \end{pmatrix}.$$

We now expand the matrix on the right side by the first row. We get

$$\det A = (-1)^{1+2}(-1)(44+33) = 77.$$

Hence we get $A^{-1} = \frac{1}{77}\begin{pmatrix} 33 & 22 & -11 \\ 10 & 9 & 6 \\ 29 & 3 & 2 \end{pmatrix}$.

6.4.2 Cramer's Rule

Consider a system of three equations in three unknowns, $Ax = b$, where

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \text{ and } x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

We assume $\det A \neq 0$. Thus we want to solve

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

Let

$$A_i := \begin{pmatrix} a_{1i} \\ a_{2i} \\ a_{3i} \end{pmatrix}$$

for $i = 1, 2, 3$. Then

$$\begin{aligned} x_1 \det(A_1, A_2, A_3) &= \det(x_1 A_1, A_2, A_3) \\ &= \det(x_1 A_1 + x_2 A_2 + x_3 A_3, A_2, A_3) \\ &= \det(b, A_2, A_3). \end{aligned}$$

Therefore,

$$x_1 = \frac{\det(b, A_2, A_3)}{\det A}.$$

This proof can obviously be generalized for any n. We have thus the following theorem:

Theorem 6.4.9 (Cramer's Rule) *Let $A \in M(n, \mathbb{R})$ with $\det A \neq 0$. Let $b \in \mathbb{R}^n$ be a column vector. Then the solution of $Ax = b$ is given by*

$$x_j = \frac{\det(A_1, \ldots, b, \ldots, A_n)}{\det A}$$

where b is in the jth place and $\det A = \det(A_1, \ldots, A_n)$ where A_i is the ith column of A.

□

Proof 2. We want to solve $Ax = b$, where A is an $n \times n$ matrix, b is a fixed column vector. Write $A = (A_1, \dots, A_n)$, where A_i is the ith column. Set

$$X_k = (e_1, \dots, e_{k-1}, x, e_{k+1}, \dots, e_n)$$

where $I = (e_1, \dots, e_n)$. Then

$$\begin{aligned}
x_k &= \det X_k \\
&= \det\left(e_1, \dots, e_{k-1}, \sum x_k e_k, e_{k+1}, \dots, e_n\right) \\
&= \det(A^{-1}AX_k) \qquad (\text{since } \det(AB) = \det A \det B) \\
&= \det(AX_k)/\det A \qquad (\text{since } \det(A^{-1}) = (\det A)^{-1}) \\
&= \det(A_1, \dots, A_{k-1}, b, A_{k+1}, \dots, A_n)/\det A
\end{aligned}$$

since $AX_k = (Ae_1, \dots, Ae_{k-1}, Ax, Ae_{k+1}, \dots, Ae_n)$.

$\square$

Example 6.4.2 Let us solve the system

$$\begin{aligned}
x + y &= 0 \\
y + z &= 1 \\
z + x &= -1.
\end{aligned}$$

This can easily be solved. From the first equation, we see that $x = -y$. Substituting this value of y in the second equation, we get $z - x = 1$. This along with $x + y = 0$ gives us $z = 0$, $x = -1$ and $y = 1$. However, we shall solve this using Cramer's rule. The coefficient matrix A is

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Let us write the system as matrix equation $Ax = b$ where

$$b := \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

We compute the determinant of A:

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \xrightarrow{C_2 - C_1} A' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix}$$

$$\det(A') = (-1)^{1+1}2 = 2 \neq 0.$$

Now we can apply Cramer's rule. Let B_i denote the matrix obtained from A by replacing the ith column by the column vector b. Then we have

$$B_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix}.$$

Hence $\det B_1 = (-1)^{1+2}2 = -2$.

$$B_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix}.$$

Hence $\det B_2 = (-1)^{1+1}2 = 2$.

$$B_3 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix} \xrightarrow{C_2 - C_1} B_3' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix}.$$

Hence $\det B_3' = (-1)^{1+1}0 = 0$. Thus the solution is given by

$$\begin{aligned} x &= \frac{\det B_1}{\det A} = -1 \\ y &= \frac{\det B_2}{\det A} = 1 \\ z &= \frac{\det B_3}{\det A} = 0. \end{aligned}$$

Exercise 6.4.2 Let

$$A = \begin{pmatrix} 6 & 2 & 8 \\ -3 & 4 & 1 \\ 4 & -4 & 5 \end{pmatrix}.$$

(1) Find the adjunct $\tilde{A}$ of A.
(2) Compute $\det A$.
(3) Show that $\tilde{A}A = \det AI$.

Exercise 6.4.3 Compute the inverse of the following matrices if they exist:

$$\text{(i)} \quad \begin{pmatrix} 4 & 2 & 2 \\ 0 & 1 & 2 \\ 1 & 0 & 3 \end{pmatrix} \qquad \text{(ii)} \quad \begin{pmatrix} 4 & 0 & 2 \\ 0 & 3 & 4 \\ 0 & 1 & -2 \end{pmatrix}.$$

Exercise 6.4.4 Compute the determinant of

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 0 & -1 \\ 2 & 0 & 4 & 2 \\ 7 & 3 & 1 & -1 \end{pmatrix}.$$

Exercise 6.4.5 Compute the determinant of

$$\begin{pmatrix} 1-x & 1 & 1 \\ 1 & 1-x & 1 \\ 1 & 1 & 1-x \end{pmatrix}.$$

Can you generalize this?

Exercise 6.4.6 Find the inverse of the following matrices:

$$\text{(i)} \quad \begin{pmatrix} 1 & 2 & 2 \\ 3 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}. \quad \text{(ii)} \quad \begin{pmatrix} 2 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Exercise 6.4.7 Solve the system of equations:

$$\begin{aligned} 2x + y &= 0 \\ 3y + z &= 1 \\ 4z + x &= 2. \end{aligned}$$

Exercise 6.4.8 True or false: $\det(A+B) = \det(A) + \det(B)$?

Exercise 6.4.9 For what values of r we have $\det(\alpha A) = \alpha^r \det A$?

Exercise 6.4.10 (Block-diagonal matrix) Consider a matrix of the form

$$C = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

where A and B are square matrices and each 0 denotes a matrix of zeroes. C is called a block-diagonal matrix with two diagonal blocks A and B. Now $\det C = \det A \det B$ as

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix} \tag{6.4.9}$$

where I is the identity matrix. Consider the function:

$$f(A) = \det \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix}.$$

f satisfies the conditions (1) and (2) of Theorem 6.4.3. Therefore,

$$\det \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix} = \det A.$$

Similarly, we get

$$\det \begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix} = \det B.$$

Since $\det(ST) = \det(S)\det(T)$ the result follows from Equation (6.4.9).

Exercise 6.4.11 If

$$C = \begin{pmatrix} A & D \\ 0 & B \end{pmatrix}$$

then, $\det C = \det A \det B$.

6.4.3 Some Geometric Ideas

Let $A = (a_{ij}) \in M(n, \mathbb{R})$. $\det A$ can be thought of in two ways:

(1) As in the definition, $\det(A) = \sum_{\sigma \in S_n} \text{sign}(\sigma) \alpha_{1\sigma(1)} \cdots \alpha_{n\sigma(n)}$.

(2) Since $A_i = Ae_i$, $\det(A_1, A_2, \ldots, A_n)$ stands for the 'signed' volume of the image of the unit cube under the linear map A. In this case we can think of $\det A$ as the distortion factor for the volume under A of the volume in the domain space.

This geometric way of looking at the determinant is quite useful in differential geometry. Let us look at some examples.

Lemma 6.4.10 *Let $\{v_i\}_{i=1}^n$ be a basis for $\mathbb{R}^n$. Let $v_i = \sum_{i=1}^n v_{ji} e_j$, where $\{e_j\}_{j=1}^n$ is the standard basis for $\mathbb{R}^n$. Then*

$$\text{vol}\,([v_1, \ldots, v_n]) := \det(v_1, \ldots, v_n) = \sqrt{\det(\langle v_i, v_j \rangle)}.$$

Proof Let A be the linear operator which takes e_i to v_i. Then its matrix with respect to $\{e_i\}$ is given by $A = [v_1, \ldots, v_n]$, where

$$v_i = \begin{pmatrix} v_{1i} \\ \vdots \\ v_{ni} \end{pmatrix}.$$

We have

$$\det(A^t A) = \det([v_1, \ldots, v_n]^t [v_1, \cdots v_n]) = \det \begin{pmatrix} \langle v_1, v_1 \rangle & \ldots & \langle v_1, v_n \rangle \\ \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle & \ldots & \langle v_n, v_n \rangle \end{pmatrix}.$$

That is, $(\det A)^2 = \det(\langle v_i, v_j \rangle)$, $1 \le i, j \le n$. Therefore

$$\text{vol}\,([v_1, \ldots, v_n]) = |\det(\langle v_i, v_j \rangle)|^{\frac{1}{2}}.$$

□

Remark 6.4.1 Let V be an inner product space and $\{e_1, \ldots, e_n\}$ be an orthonormal basis of V. If the basis $\{u_1, \ldots, u_n\}$ is orthogonal, then the parallelepiped $[u_1, \ldots, u_n]$ is 'rectangular'. From Lemma 6.4.10, we see that

$$\det(u_1, \ldots, u_n)^2 = \prod_i ||u_i||^2.$$

This agrees with our intuition, namely the volume of a rectangular parallelepiped is the product of the lengths of its sides.

If $\{v_1, \ldots, v_n\}$ is a basis of V, then by Gram-Schmidt process we get an orthogonal basis $\{u_1, \ldots, u_n\}$. Let $A = (v_1, \ldots, v_n)$ where v_i is thought of as a column

$$v_i = \begin{pmatrix} v_{1i} \\ \vdots \\ v_{ni} \end{pmatrix}.$$

Since u_i is a linear combination of v_j for $1 \le j \le i$, it follows that if we replace v_i by u_i then $\det(v_1, \ldots, v_n) = \det(u_1, \ldots, u_n)$. This corresponds to the geometric idea of cutting the parallelepiped into pieces and rearranging them to get a rectangular parallelepiped as was done in the case of a parallelogram.

Remark 6.4.2 If we think of elements of $\mathbb{R}^n$ as column vectors then the dot product on $\mathbb{R}^n$ can be written as $\langle x, y\rangle = y^t x$, the matrix multiplication of $1 \times n$ matrix by $n \times 1$ matrix. A 1×1 matrix (α) is thought of as the real number α.

Lemma 6.4.11 *Let $A \in M(n, \mathbb{R})$ be given. Let $\{v_i\}_{i=1}^n$ be a basis of $\mathbb{R}^n$. Assume that we are given $\langle Av_i, v_j\rangle$ for all $1 \le i, j \le n$. Then*

$$\det(A) = \frac{\det(\langle Av_i, v_j\rangle)}{\det(\langle v_i, v_j\rangle)}.$$

Proof Let B be the matrix which takes e_i to v_i. (Recall that the ith column of B is the column vector v_i.) Now, we have

$$\langle Av_i, v_j\rangle = v_j^t A v_i = (Be_j)^t ABe_i = e_j^t B^t ABe_i = \left\langle B^t ABe_i, e_j\right\rangle.$$

Hence

$$\begin{aligned} \det(\langle Av_i, v_j\rangle) &= \det(\left\langle B^t ABe_i, e_j\right\rangle) \\ &= \det(B^t AB) \\ &= \det(B^t)\det(A)\det(B). \end{aligned} \tag{6.4.10}$$

Again,

$$\begin{aligned} \det(B^t B) &= \det(\left\langle B^t Be_i, e_j\right\rangle) \\ &= \det(\langle Be_i, Be_j\rangle) \\ &= \det(\langle v_i, v_j\rangle). \end{aligned} \tag{6.4.11}$$

The result follows from Equations (6.4.10) and (6.4.11).

□

6.5 Orientation and Vector Product

This section deals with two more uses of determinants, which may be skipped in the first reading.

6.5.1 Orientation

Let V be any vector space, $v_i \in V$, $1 \le i \le n$. We have emphasized that we may think of $|\det(v_1, \ldots, v_n)|$ as the volume of the parallelepiped $[v_1, \ldots, v_n]$. There is another use of determinants which employs the sign of the determinant. In $\mathbb{R}^2$ and $\mathbb{R}^3$ we have notions of orientation. To talk of orientation in higher dimensions is quite unintuitive unless based on some mathematical concept.

We fix a basis $\{e_1, \ldots, e_n\}$ of V. Let $\{v_1, \ldots, v_n\}$ be another basis. Let A be the matrix which takes e_i to v_i. Note that $\det(A) \neq 0$, thanks to Theorem 6.4.3. We say that $\{v_1, \ldots, v_n\}$ is *positively* (or *negatively*) *oriented* if $\det(A) > 0$ (respectively if $\det(A) < 0$). Note that the order in the listing of the basis is important, for instance, the basis $\{e_2, e_1, e_3, \ldots, e_n\}$ is negatively oriented.

One can see that this agrees with our intuition in the case of $\mathbb{R}^2$ and $\mathbb{R}^3$ where the fixed basis is taken as the standard basis.

One usually thinks of $\det(v_1, \ldots, v_n)$ as the *oriented volume* of the parallelepiped $[v_1, \ldots, v_n]$.

Remark 6.5.1 More abstractly, given two ordered bases B_1 and B_2 of a real vector space V, the unique linear isomorphism of V which takes B_1 to B_2 (preserving the order) has nonzero determinant. We say that B_1 and B_2 are equivalent if this determinant is positive. Thus, the set of ordered bases $\mathcal{B}$ of a real vector space V is the disjoint union of two subsets $\mathcal{B}_1$ and $\mathcal{B}_2$ where the nonsingular transformation which takes one basis of $\mathcal{B}_j$ to another in $\mathcal{B}_j$ has positive determinant whereas the linear isomorphism which takes one basis, say, from $\mathcal{B}_1$ to another in $\mathcal{B}_2$ has negative determinant. An orientation of V is nothing other than declaring one of $\mathcal{B}_j$ to be the set of positive bases of V and call the other as the set of negative bases.

6.5.2 Vector Product

We define a *cross product* on a three-dimensional real vector space V with an inner product: $(x, y) \mapsto \langle x, y \rangle$. We fix an orthonormal basis $\{e_i\}$ of V such that $\langle e_i, e_j \rangle = \delta_{ij}$. If you wish you may take $V = \mathbb{R}^3$ with the standard basis vectors and the Euclidean inner product $(x, y) \mapsto \langle x, y \rangle := \sum_{i=1}^{3} x_i y_i$. We also have the Riesz representation theorem: For any linear map $f : V \to \mathbb{R}$ there exists a unique $u \in V$ such that $f(x) = \langle x, u \rangle$. *Hint:* With basis vectors e_i we take $u := \sum_i f(e_i)e_i$. We now define the *cross product* or *vector product* on V as follows:

For $x, y \in V$, the map $z \mapsto \det(x, y, z)$ is linear map of V to $\mathbb{R}$ and hence by Riesz representation theorem (Theorem 5.7.1) there exists a unique vector v such that $\langle v, z\rangle = \det(x, y, z)$, for all $z \in V$. We denote this vector v by $x \times y$ and call it the *cross product* or the *vector product* of x and y. Let us record this defining property of $x \times y$:

$$\langle z, x \times y\rangle := \det(x, y, z), \quad \text{for all } z \in V.$$

Let us find the coordinates of $x \times y$ with respect to the orthonormal basis $\{e_i\}$. If we write $x \times y = \sum_i w_i e_i$, then

$$\begin{aligned}
w_1 := \langle x \times y, e_1\rangle &= \langle e_1, x \times y\rangle \\
&= \det(x, y, e_1) \\
&= \det(x_1e_1 + x_2e_2 + x_3e_3, y_1e_1 + y_2e_2 + y_3e_3, e_1) \\
&= \det(x_2e_2 + x_3e_3, y_2e_2 + y_3e_3, e_1) \\
&= \det(x_2e_2, y_3e_3, e_1) + \det(x_3e_3 + y_2e_2, e_1) \\
&= x_2y_3 \det(e_2, e_3, e_1) + x_3y_2 \det(e_3, e_2, e_1) \\
&= x_2y_3 - x_3y_2.
\end{aligned}$$

One similarly finds that $w_2 = x_3y_1 - x_1y_3$ and $w_3 = x_1y_2 - x_2y_1$.
Thus

$$\begin{aligned}
x \times y &= (x_2y_3 - x_3y_2)e_1 - (x_3y_1 - x_1y_3)e_2 + (x_1y_2 - x_2y_1)e_3 \\
&= \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix}.
\end{aligned}$$

This is the familiar expression for $x \times y$ in vector analysis.

Lemma 6.5.1 *The vector product has the following properties:*

(1) *$x \times y$ is orthogonal to x and y.*

(2) *$\lambda x \times y = \lambda(x \times y) = x \times \lambda y$, for $\lambda \in \mathbb{R}$.*

(3) *$y \times x = -x \times y$.*

(4) *$x \times y = 0$ if and only if x and y are linearly dependent.*

(5) *$\langle x \times y, z\rangle = \langle y \times z, x\rangle = \langle z \times x, y\rangle$.*

(6) *$\langle x, y \times z\rangle = \langle y, z \times x\rangle = \langle z, x \times y\rangle$.*

Proof These are immediate consequences of the properties of determinants. We give a sample of the arguments. Let us prove the first assertion. For instance, to prove that $\langle x, x \times y \rangle = 0$, we have by the very definition of vector product

$$\langle x, x \times y \rangle = \det(x, y, x) = 0,$$

since two terms are equal in the determinant. The rest of the assertions go on similar lines and we leave them to the reader.

□

Proposition 6.5.2 *For any three vectors $x, y, z \in V$, we have*

$$x \times (y \times z) = (\langle x, z \rangle)y - (\langle x, y \rangle)z. \tag{6.5.1}$$

Proof To show that these two vectors are equal, it is enough to show that their inner product with any vector of V (in fact, any vector in an orthonormal basis) are the same:

$$\langle v, x \times (y \times z) \rangle = \langle v, (\langle x, z \rangle)y - (\langle x, y \rangle)z \rangle .$$

In view of (4), it is enough to verify for an arbitrary vector v,

$$\langle v \times x, y \times z \rangle = \langle v, y \rangle \langle x, z \rangle - \langle x, y \rangle \langle v, z \rangle . \tag{6.5.2}$$

We first observe that both sides are linear in each of the variables. Hence it is enough to verify it on $\{e_i\}$. Due to symmetry we may take $y = e_1$, $z = e_2$ so that $y \times z = e_3$. Now it is easily checked that both sides of Equation (6.5.2) are equal to $(v_1 x_2 - v_2 x_1)$.

□

Note that the vector $x \times y$ is the vector orthogonal to x and y. If x and y are linearly independent, then $x \times y \neq 0$. In fact, from Equation (6.5.2), it follows that $||x \times y||^2 = \langle x, x \rangle \langle y, y \rangle - \langle x, y \rangle^2$, which is the square of the area of the parallelogram spanned by x and y. Thus, $\{x, y, x \times y\}$ is a basis of V. Also, we claim that this has the same orientation as the basis $\{e_1, e_2, e_3\}$. To show this we need to show that if A is the matrix such that $Ae_1 = x$, $Ae_2 = y$ and $Ae_3 = x \times y$, then $\det(A) > 0$. The matrix A has as its columns x, y and $x \times y$: $A = (x, y, x \times y)$, where x etc. are thought of column vector with respect to the basis $\{e_i\}$. Now,

$$\det(A) = \det(x, y, x \times y) = \langle x \times y, x \times y \rangle > 0,$$

since $x \times y \neq 0$.

The geometric meaning of the vector or cross product $x \times y$ is that it is the vector orthogonal to x and y with the property that $\{x, y, x \times y\}$

is a basis with the same orientation as $\{e_1, e_2, e_3\}$ and is of length equal to the area of the parallelogram spanned by x and y. It may be noted that the length is $||x|| \, ||y|| \sin\theta$, where θ is the angle between x and y (see Section 6.1).

7. Diagonalization

The simplest linear maps from a vector space V to itself are αI, for $\alpha \in \mathbb{R}$. Next come the linear maps of the form $v_i \mapsto \alpha_i v_i$ where $\{v_i\}$ is a basis of V and $\alpha_i \in \mathbb{R}$, $1 \le i \le n$. If we write the matrix of these maps with respect to this basis, it is of the form (a_{ij}) where

$$a_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ \alpha_i & \text{if } i = j \end{cases}$$

is a diagonal matrix. We denote this diagonal matrix by diag $(\alpha_1, \dots, \alpha_n)$. The main theme of this chapter is to prove that if we are given a symmetric $n \times n$ matrix, then we can find an orthonormal basis $\{v_i\}$ of $\mathbb{R}^n$ and scalars $\alpha_i \in \mathbb{R}$ such that $Av_i = \alpha_i v_i$ for $1 \le i \le n$.

7.1 Rotation of Axes of Conics

Let us start by reviewing the trick of rotation of axes so that a conic given by the quadratic expression $ax^2 + 2hxy + by^2$ is written in one of the standard forms of an ellipse, hyperbola or parabola.

Let

$$A = \begin{pmatrix} a & h \\ h & b \end{pmatrix}.$$

Then the given quadratic expression can be expressed as

$$ax^2 + 2hxy + by^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & h \\ h & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

A is called the *coefficient matrix* of the homogeneous quadratic polynomial. The rotation by the angle θ in the anticlockwise direction is given by

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

If we effect the coordinate transformation

$$\begin{pmatrix} x \\ y \end{pmatrix} = R_\theta \begin{pmatrix} u \\ v \end{pmatrix}$$

then the quadratic expression becomes

$$\begin{pmatrix} u \\ v \end{pmatrix} \mapsto (u, v) R_\theta^t A R_\theta \begin{pmatrix} u \\ v \end{pmatrix}.$$

Hence the coefficient matrix with respect to the new coordinates (u, v) is

$$R_\theta^t A R_\theta = R_{-\theta} A R_\theta = \begin{pmatrix} p & q \\ q & r \end{pmatrix}.$$

where

$$q = h(\cos^2\theta - \sin^2\theta) - (a-b)\sin\theta\cos\theta$$

and hence

$$2q = 2h\cos 2\theta - (a-b)\sin 2\theta.$$

We can make $q = 0$ by taking θ so that $\tan 2\theta = \dfrac{2h}{a-b}$ if $a \neq b$, or

$$\theta = \begin{cases} \frac{\pi}{4} & \text{if } a = b \text{ and } h \neq 0 \\ 0 & \text{if } h = 0. \end{cases}$$

We thus have the quadratic expression as $(u, v) \mapsto pu^2 + rv^2$.

In geometric terms, p and r are the principal (major/minor) axes of the conic. Note that if v is a point of the conic Q, then

$$v = \begin{pmatrix} x \\ y \end{pmatrix}$$

is orthogonal to the tangent space T_vQ if and only if $Av = pv$ or rv. For, the tangent line is given by

$$(h_1, h_2) A \begin{pmatrix} x \\ y \end{pmatrix} = 0$$

that is, by $(Ax)^\perp$. Hence $v \in Q$ is orthogonal to $(Av)^\perp$ or v is a scalar multiple of Av. This introduces the notion of eigenvector.

Definition 7.1.1 Let V be a vector space, $A\colon V \to V$ be linear. We say that a nonzero $v \in V$ is an *eigenvector* for A if there exists $\alpha \in \mathbb{R}$ such that $Av = \alpha v$. α is called an *eigenvalue*.

Example 7.1.1 Assume that there exists a basis $\{v_i\}_{i=1}^n$ such that the matrix of A with respect to this basis is diagonal: $M_v^v A = \operatorname{diag}(\alpha_1, \dots, \alpha_n)$. Then v_i is an eigenvector of A with eigenvalue α_i.

Example 7.1.2 Let

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then $v_1 = e_1 + e_2$ and $v_2 = e_1 - e_2$ are eigenvectors with eigenvalues 1 and -1 respectively.

Assume that there exists $\alpha \in \mathbb{R}$ and a nonzero $v \in V$ such that $Av = \alpha v$. Then $(A - \alpha I)v = 0$. By Theorem 6.4.4, $\det(A - \alpha I) = 0$. Conversely, if $\det(A - \alpha I) = 0$ for some $\alpha \in \mathbb{R}$, then again by Theorem 6.4.4, there exists a nonzero $v \in V$ such that $(A - \alpha I)v = 0$ or $Av = \alpha v$. Thus finding of (real) eigenvalues is equivalent to finding the real roots of the polynomial equation $\det(A - xI) = 0$. The polynomial $\det(A - xI)$ is called the *characteristic polynomial* of A.

We shall concentrate on $\mathbb{R}^2$ in the rest of the section. We shall identify any linear map $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ with its matrix $M(A)$ with respect to the standard basis of $\mathbb{R}^2$. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then the characteristic polynomial χ_A of A is given by

$$\det(A - xI) = x^2 - (a + d)x + (ad - bc) = x^2 - \operatorname{tr}(A) + \det(A).$$

Now this quadratic polynomial has real roots if and only if its discriminant "$b^2 - 4ac$" is non-negative.

We now assume that the matrix A is symmetric, say,

$$A = \begin{pmatrix} a & h \\ h & b \end{pmatrix}.$$

Then the characteristic polynomial is $x^2 - (a + b)x + (ab - h^2)$. Its discriminant is $(a+b)^2 - 4(ab - h)^2 = (a - b)^2 + h^2 \geq 0$. Thus, a symmetric matrix of order 2 has real eigenvalues.

We shall redo the earlier formula for the angle of a rotation that brings the matrix into diagonal form in a slightly different way which will generalize to higher dimensions.

Let us consider the map $f\colon [0, 2\pi] \to \mathbb{R}$ given by

$$f(t) = \left\langle A \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} \right\rangle = (\cos t \quad \sin t)\, A \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

the dot product of Av and v where v is the column vector which is on the unit circle in $\mathbb{R}^2$. A computation yields

$$f(t) = a\cos^2 t + 2h\sin 2t + b\sin^2 t.$$

Since f is clearly a continuous real valued function on the closed and bounded interval $[0, 2\pi]$, it attains its maximum and minimum in $[0, 2\pi]$. Since $f(0) = f(2\pi)$ both these extremum cannot be at the end points of these intervals (unless f is a constant). If f is a constant, all points in the open interval are points of extremum for f. Thus we may assume that either a minimum or a maximum occurs at θ in the open interval $(0, 2\pi)$. Then $f'(\theta) = 0$. We find $f'(t) = (b-a)\sin 2t + 2h\cos 2t$. So θ satisfies the equation

$$(b-a)\sin 2\theta + 2h\cos 2\theta = 0 \text{ or } \tan 2\theta = \frac{2h}{a-b}$$

provided $a \neq b$. If $a = b$, then $f'(\theta) = 0$ implies that we may take $\theta = \pi/4$.

Exercise 7.1.1 Investigate the case when f is a constant.

7.2 Eigenvalues and Eigenvectors

V denotes a finite dimensional vector space over $\mathbb{R}$. Let $A\colon V \to V$ be linear. We fix a basis $\{e_1, \ldots, e_n\}$ of V. We use the same symbol A to denote the matrix $M_e^e(A)$.

Definition 7.2.1 We say a real number α is an *eigenvalue* of A if there exists a nonzero vector $v \in V$ such that $Av = \alpha v$. Any nonzero vector $u \in V$ such that $Au = \lambda v$ is called an *eigenvector* of A corresponding to the eigenvalue λ.

Exercise 7.2.1 If v is an eigenvector with eigenvalue α then λv is an eigenvector with eigenvalue λ. In fact, the nonempty set

$$V_\beta := \{x \in V \mid Ax = \beta x\}$$

is a vector subspace of V for any $\beta \in \mathbb{R}$. (You can prove this directly. But can you think of a proof which uses some earlier result?)

The central problem is to find whether there is a basis $\{v_i\}$ of V consisting of eigenvectors of A: $Av_i = \alpha_i v_i$. We call such a basis an A-eigen basis of V. If there is no confusion, we shall simply say an eigen basis of V.

If $\{v_i\}$ is an eigen basis of V, then $M_v^v(A)$ is diagonal:

$$M_v^v(A) = \operatorname{diag}(\alpha_1, \ldots, \alpha_n).$$

For this reason, one calls finding an eigen basis of A as diagonalization of A. In fact, we have

Example 7.2.1 Assume that there exists a basis $\{v_i\}_{i=1}^{n}$ such that the matrix of A with respect to this basis is diagonal: $M_v^v A = \text{diag}\,(\alpha_1, \ldots, \alpha_n)$. Then v_i is an eigenvector of A with eigenvalue α_i.

Example 7.2.2 Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then $v_1 = e_1 + e_2$ and $v_2 = e_1 - e_2$ are eigenvectors with eigenvalues 1 and -1 respectively.

Exercise 7.2.2 Find by inspection the eigenvectors and eigenvalues of

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Lemma 7.2.1 *Let $\{v_i\}$ be an eigen basis for A. Let T be the linear map such that $Te_i = v_i$. As agreed earlier, let A and T denote the matrices of the linear maps with respect to the fixed basis $\{e_i\}$. Then $T^{-1}AT$ is the diagonal matrix $\text{diag}\,(\alpha_1, \ldots, \alpha_n)$.*

Proof Let T be the linear map such that $Te_i = v_i$. Recall that

$$M_e^e(T) = (v_1, \ldots, v_n)$$

where v_i is the column vector $(v_{1i}, \ldots, v_{ni})^t$ where $v_i = \sum_j v_{ji} e_j$. Let us compute:

$$Av_i = \alpha_i v_i \implies ATe_i = \alpha_i Te_i \implies T^{-1}ATe_i = \alpha_i e_i.$$

This immediately yields the result.

$\square$

Assume that there exists $\alpha \in \mathbb{R}$ and a nonzero $v \in V$ such that $Av = \alpha v$. Then $(A - \alpha I)v = 0$. By Theorem 4.4.4, $\det(A - \alpha I) = 0$. Conversely, if $\det(A - \alpha I) = 0$ for some $\alpha \in \mathbb{R}$, then again by Theorem 4.4.4, there exists a nonzero $v \in V$ such that $(A - \alpha I)v = 0$ or $Av = \alpha v$. Thus finding of (real) eigenvalues is equivalent to finding the real roots of the polynomial equation $\det(A - xI) = 0$. The polynomial $\det(A - xI)$ is called the *characteristic polynomial* of A.

Example 7.2.3 Find the eigenvectors and eigenvalues of

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 3 \end{pmatrix}.$$

The characteristic equation of A is $\det(A - \lambda I) = 0$. That is,

$$\begin{vmatrix} -\lambda & 0 & 2 \\ 0 & 2-\lambda & 0 \\ 2 & 0 & 3-\lambda \end{vmatrix} = 0.$$

On expanding, we get

$$\begin{aligned} -\lambda \begin{vmatrix} 2-\lambda & 0 \\ 0 & 3-\lambda \end{vmatrix} + 0 \begin{vmatrix} 0 & 0 \\ 2 & 3-\lambda \end{vmatrix} + 2 \begin{vmatrix} 0 & 2-\lambda \\ 2 & 3-\lambda \end{vmatrix} &= 0 \\ -\lambda(2-\lambda)(3-\lambda) + 2[(-2)(2-\lambda)] &= 0 \\ (2-\lambda)[\lambda(\lambda-3)-4] &= 0 \\ (2-\lambda)(\lambda^2-3\lambda-4) &= 0 \\ (2-\lambda)(\lambda+1)(\lambda-4) &= 0. \end{aligned}$$

Hence the eigenvalues are $\lambda = 2, -1, 4$. For $\lambda = 2$, by inspection we can see that e_2 is the eigenvector. For $\lambda = -1$, we need to find a vector

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

such that

$$\begin{pmatrix} 0 & 0 & 2 \\ 0 & 2 & 0 \\ 2 & 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = - \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

That is, we need to solve the system of linear equations

$$\begin{aligned} x + 2z &= 0 \\ 3y &= 0 \\ 2x + 4z &= 0. \end{aligned}$$

Thus we have, in fact, a single equation $x + 2z = 0$ so that we may take the vector

$$\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

or the unit vector

$$\begin{pmatrix} -2/\sqrt{5} \\ 0 \\ 1/\sqrt{5} \end{pmatrix}.$$

A unit vector perpendicular to this in the xz-plane is given by

$$\begin{pmatrix} 1/\sqrt{5} \\ 0 \\ -2/\sqrt{5} \end{pmatrix}.$$

This is an eigenvector with eigenvalue 4, as can be easily verified.

Definition 7.2.2 A linear map $A: V \to V$ is said to be *diagonalizable* if there exists an A-eigen basis of V. Equivalently, if there exists a basis $\{v_i\}$ of V with respect to which $M_v^v(A)$ is a diagonal matrix.

Example 7.2.4 Reflections in $\mathbb{R}^2$ are diagonalizable. Let R denote the reflection with respect to x-axis. Then $R(x, y) = (x, -y)$ so that $Re_1 = e_1$ and $Re_2 = -e_2$. Thus the matrix of R with respect to the standard basis is

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

which is already in the diagonal form.

More generally, let

$$\mathbb{R}_\theta = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$$

denote the reflection with respect to the line $\mathbb{R}(\cos\theta, \sin\theta)$. Then $\mathbb{R}_\theta$ is a diagonalizable. This is an immediate consequence of the last paragraph of Section 5.9. However, we repeat the proof in a slightly different form. R_θ maps the vector $(\cos\theta, \sin\theta)$ to itself and maps any normal $(-\sin\theta, \cos\theta)$, say, to the line to its negative. Thus, the basis

$$v_1 := \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \text{ and } v_2 := \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$

are eigenvectors of R_θ. Let T denote the orthogonal transformation which takes e_i to v_i, $i = 1, 2$. Then

$$T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Also, as we have

$$\begin{aligned} T^{-1} R_\theta T e_1 &= T^{-1} v_1 = e_1 \\ T^{-1} R_\theta T e_2 &= T^{-1}(-v_2) = -e_2. \end{aligned}$$

In matrix notation, we have

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Example 7.2.5 Let us investigate when

$$A = \begin{pmatrix} a & h \\ h & b \end{pmatrix}$$

has a nonzero eigenvector. If it has a nonzero eigenvector it has an eigenvector of unit length. So, we may assume that

$$v := \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \text{ for some } \theta \in \mathbb{R}$$

is an eigenvector of A with eigenvalue $\lambda \in \mathbb{R}$: $Av = \lambda v$. This is equivalent to the system of linear equations

$$\begin{aligned} a\cos\theta + h\sin\theta &= \lambda\cos\theta \\ h\cos\theta + b\sin\theta &= \lambda\sin\theta. \end{aligned}$$

Let us formally divide the first equation by $\cos\theta$ and the second by $\sin\theta$ without worrying about one of them being zero. (Only one of them could be zero!) We then get $a + h\tan\theta = \lambda$ and $h\cot\theta + b = \lambda$. Eliminating λ from these two equations, we get $a - b = h\cot 2\theta$ or $\cot 2\theta = \frac{a-b}{h}$, this time not worrying about h being zero! If h is zero, A is then already diagonal and hence the standard basis is also an eigen basis.

Now, if $\sin\theta = 0$, then $\cos\theta = \pm 1$ so that we may take $v = e_1$. In this case, working as above we find that $\lambda = a$ and $h = 0$. Thus, A is diagonal.

Before we go any further, it is important to realize that not all linear maps are diagonalizable.

Example 7.2.6 Consider the linear map $A\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $Ae_1 = e_2$ and $Ae_2 = 0$. Then

$$A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

If A has any nonzero eigenvector

$$v = \begin{pmatrix} x \\ y \end{pmatrix}$$

say, with eigenvalue λ, we then end up with the system of linear equations $\lambda x = 0$ and $\lambda y = x$. If $\lambda \neq 0$, then $x = 0$ so that $y = 0$. Thus the vector is 0, a contradiction. Note also that we have also shown that the only eigenvalue is 0. If $\lambda = 0$, then we may take $v_1 = e_2$. Is there a second nonzero eigenvector v_2 so that $\{v_1, v_2\}$ is an eigen basis? If

$$v_2 = \begin{pmatrix} x \\ y \end{pmatrix}$$

and we have only one (up to scalar multiple) eigenvector and hence there is no eigen basis for A.

We could have made a slick argument using the characteristic polynomial. Note that the characteristic polynomial of A is $X^2 = 0$. Hence the eigenvalues are 0 and 0. Thus we need two linearly dependent eigenvectors. Clearly e_2 is an eigenvector with eigenvalue 0. We look for a second one. If

$$v = \begin{pmatrix} x \\ y \end{pmatrix}$$

is one such then $Av = 0$ yields

$$\begin{pmatrix} 0 \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ or } x = 0.$$

Thus any eigenvector is a scalar multiple of e_2! There is no way we can find an eigen basis.

The purpose of giving a bare-handed approach (the first proof) and a more theoretic one (the second one) is to help the reader

(1) appreciate the power of building a theory, and

(2) strike his/her own path when there is no theory to work on.

Let us assume that $A: V \to V$ has a (real) eigenvalue λ with a nonzero eigenvector v: $Av = \lambda v$. We can rewrite this as $(A - \lambda I)v = 0$. By Theorem 6.4.4 we know that this happens if and only if $\det(A - \lambda I) = 0$. The crucial observation now is that the left side is a polynomial of degree n in λ as can be seen from the explicit formula for the determinant. Thus any (real) eigenvalue is a real root of the equation $\det(A - \lambda I) = 0$.

Definition 7.2.3 Let $A \in M(n, \mathbb{R})$. Then the *characteristic polynomial* $\chi_A(X) := \det(A - XI)$ where X is an indeterminate.

Thus a real λ is an eigenvalue of A if and only if it is a real root of the characteristic polynomial of A.

Now the fundamental theorem of algebra tells us that a polynomial equation of degree n in one indeterminate has n complex roots. Thus, we may have "complex eigenvalues" but no eigenvector in $\mathbb{R}^n$.

Example 7.2.7 Let the linear map $A: \mathbb{R}^2 \to \mathbb{R}^2$ be given by $Ae_1 = e_2$ and $Ae_2 = -e_1$. Then

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

It is the rotation in the anticlockwise direction by $\pi/2$. Its characteristic polynomial is $X^2 + 1 = 0$. It has complex roots $\pm\sqrt{-1}$. Thus there are no eigenvectors of A in $\mathbb{R}^2$.

This example illustrates the problems we may encounter if we want to find an eigen basis for a given $A: V \to V$:

(1) There may not be enough eigenvalues corresponding to a given (real) eigenvalue as in Example 7.2.6.

(2) There may not be any (real) eigenvalue as in Example 7.2.7.

Lemma 7.2.2 *Let $A: V \to V$ be linear. Assume that v_i is a (nonzero) eigenvector of A with eigenvalue α_i and that $\alpha_i \neq \alpha_j$ for $i \neq j$, $1 \leq i, j \leq r$. Then $\{v_i\}_{i=1}^r$ is a linearly independent set.*

Proof Let us first look at the case $r = 2$. If v_1 and v_2 are linearly dependent, then each is a multiple of the other. Assume that $v_1 = \lambda v_2$. Now let us operate A on both sides. We get

$$\alpha_1 v_1 = Av_1 = A(\lambda v_2) = \lambda A v_2 = \lambda \alpha_2 v_2 = \alpha_2 v_1.$$

Hence, $(\alpha_1 - \alpha_2)v_1 = 0$. Since $v_1 \neq 0$, we see that $\alpha_1 = \alpha_2$.

Now we wish to generalize this argument to all r by induction. The result is true for $r = 1$, as $v_1 \neq 0$. Let us assume the result for all $r \leq n-1$. We shall prove the result for $r = n$. Assume that

$$\sum_{i=1}^{n} \lambda_i v_i = 0. \tag{7.2.1}$$

We want to show that $\lambda_i = 0$ for all i. Let us operate A on both sides of the equation to get

$$0 = \sum_i \lambda_i \alpha_i v_i. \tag{7.2.2}$$

Multiply Equation (7.2.1) by α_n and subtract it from Equation (7.2.2). We get $\sum_{j=1}^{n-1} (\alpha_j - \alpha_n)\lambda_j v_j = 0$. Now $\{v_j\}_{i=1}^{n-1}$ is a set of nonzero eigenvectors with pairwise distinct eigenvalues and hence it is linearly independent by induction hypothesis. Thus we conclude that $(\alpha_j - \alpha_n)\lambda_j = 0$ for $1 \leq j \leq n-1$. Since $\alpha_j - \alpha_n \neq 0$ we conclude that $\lambda_j = 0$ for $1 \leq j \leq n-1$. Using this in Equation (7.2.1) yields that $\lambda_1 v_1 = 0$. Since $v_1 \neq 0$, it follows that $\lambda_1 = 0$. Thus $\lambda_i = 0$ for $1 \leq i \leq n$.

□

Remark 7.2.1 The above induction proof may also be rephrased in a different way which is useful in certain circumstances. The rephrasing goes as follows: Assume that $\{v_i\}_{i=1}^r$ is linearly dependent. We look for the minimum m such that $\{v_k\}_{k=1}^m$ is linearly dependent, say, $\sum_{k=1}^m \lambda_k v_k = 0$.

Applying A to both sides of this equation and arguing as above we deduce that $\{v_l\}_{l=1}^{m-1}$ is linearly dependent. This contradicts the minimality of m thereby establishing the linear independence of $\{v_i\}_{i=1}^{r}$.

Remark 7.2.2 Yet another proof, which is completely different from the earlier ones, uses van der Monde's determinant. We apply A^k, $0 \leq k \leq n-1$ to both sides of Equation (7.2.1) to get

$$\begin{aligned}
\lambda_1 v_1 + \cdots + \lambda_n v_n &= 0 \\
\alpha_1 \lambda_1 v_1 + \cdots + \alpha_n \lambda_n v_n &= 0 \\
\alpha_1^2 \lambda_1 v_1 + \cdots + \alpha_n^2 \lambda_n v_n &= 0 \\
\vdots \quad & \quad \vdots \\
\alpha_1^{n-1} \lambda_1 v_1 + \cdots + \alpha_n^{n-1} \lambda_n v_n &= 0.
\end{aligned}$$

This can be written as a matrix equation

$$(\lambda_1 v_1, \ldots, \lambda_n v_n) \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \ldots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \ldots & \alpha_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n & \alpha_n^2 & \ldots & \alpha_n^{n-1} \end{pmatrix}.$$

The square matrix is the *van der Monde determinant* whose value is

$$\prod_{1 \leq j < i \leq j} (\alpha_i - \alpha_j) \neq 0$$

since $\alpha_i \neq \alpha_j$ for $i \neq j$. Hence, we conclude that $(\lambda_1 v_1, \ldots, \lambda_n v_n)$ is the zero vector. Since $v_i \neq 0$, we deduce that $\lambda_i = 0$ for all i.

Exercise 7.2.3 Let $A : V \to V$ be linear and λ be an eigenvalue of A. Let $V_\lambda := \{v \in V \mid Av = \lambda v\}$. Show that V_λ is a nonzero vector subspace of V. (V_λ is called the *eigenspace* corresponding to the eigenvector λ).

7.2.1 Cayley-Hamilton Theorem

Let $f(X) = \sum_{k=0}^{n} a_k X^k$ be a polynomial in the indeterminate X with real coefficients a_k. Let A be a square matrix of size n. We then define a new matrix, denoted by $f(A)$ by setting

$$f(A) := a_n A^n + a_{n-1} A^{n-1} + \cdots + a_1 A + a_0 I.$$

Theorem 7.2.3 (Cayley-Hamilton Theorem) *Let A be an $n \times n$ square matrix. Let*

$$f(X) = X^n + c_{n-1} X^{n-1} + \cdots + c_1 X + c_0$$

be the characteristic polynomial of A. Then

$$f(A) := A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 = 0.$$

Thus, A satisfies its characteristic equation.

Proof Recall the adjunct $\operatorname{adj}(B)$ of B defined in Section 6.4 has the property that $B \operatorname{adj}(B) = \det(B)I$. We apply this result to the matrix $XI - A$ to get

$$(XI - A) \operatorname{adj}(XI - A) = \det(XI - A)I = f(X)I. \tag{7.2.3}$$

Now, $\operatorname{adj}(XI - A)$ is a matrix whose entries are determinants (up to sign) of $(n-1)$ square submatrices of $XI - A$. Hence $\operatorname{adj}(XI - A)$ is a matrix whose entries are polynomials in X of degree at most $n-1$:

$$\operatorname{adj}(XI - A) = B_{n-1}X^{n-1} + \cdots + B_1 X + B_0.$$

where B_i are matrices with real entries. Hence Equation (7.2.3) can be written as

$$(XI - A)(B_{n-1}X^{n-1} + \cdots + B_1 X + B_0) = X^n + \cdots + c_1 X + c_0. \tag{7.2.4}$$

Comparing the coefficients of like powers of X, we get

$$\begin{aligned}
B_{n-1} &= I \\
B_{n-2} - AB_{n-1} &= c_{n-1}I \\
B_{n-3} - AB_{n-2} &= c_{n-2}I \\
&\vdots \\
B_0 - AB_1 &= c_1 I \\
-AB_0 &= c_0 I.
\end{aligned}$$

Multiplying the first of these equations by A^n, the second by A^{n-1}, so on, the last but one by A and the last one by I, and adding them we get the desired result.

□

Exercise 7.2.4 What is wrong with the following "proof?" In the equation $f(X) = \det(XI - A)$ put $X = A$. We then get the result. *Hint:* The equation required to be proved involves matrices whereas you are getting a scalar equation.

There is a class of good linear maps for which we can always find an eigen basis. We introduce them next.

7.3 Diagonalization of Symmetric Matrices

Throughout this section, we let V denote a (finite dimensional) vector space (over $\mathbb{R}$) with an inner product.

Definition 7.3.1 A linear map $T : V \to V$ is said to be *symmetric* if for all $x, y \in V$, we have $\langle Tx, y\rangle = \langle x, Ty\rangle$.

Exercise 7.3.1 Fix an orthonormal basis $\{e_i\}_{i=1}^n$ of V. Let $T : V \to V$ be any linear map. Let A denote the matrix of T with respect to this basis. Then T is symmetric if and only if A is a symmetric matrix, that is, $a_{ij} = a_{ji}$ for all $1 \leq i, j \leq n$.

Lemma 7.3.1 *Let $T : V \to V$ be a symmetric linear map. Then the eigenvectors v_i with eigenvalues λ_i, $i = 1, 2$ with $\lambda_1 \neq \lambda_2$ are orthogonal to each other.*

Proof $\lambda_1 \langle v_1, v_2\rangle = \langle Tv_1, v_2\rangle = \langle v_1, Tv_2\rangle = \lambda_2 \langle v_1, v_2\rangle$.

□

The above lemma gives yet another proof of Lemma 7.2.2 in the case of a symmetric linear map on an inner product space.

We shall prove in this section that if T is a symmetric linear map, then there exists a basis of V consisting of eigenvectors of T. That is, there exists a basis $\{v_i\}_{i=1}^n$ of V such that there exist *real* numbers λ_i, $1 \leq i \leq n$, with $Tv_i = \lambda_i v_i$. Note that with respect to this basis, the matrix A of T will be diagonal with the eigenvalues as the diagonal entries. We offer two proofs of this result. Both results use some facts which the reader may not be familiar at this stage. The first one is more algebraic while the second is more geometric and analytic in character. In both the proofs, the major burden is to show the existence of an eigenvalue for a symmetric linear map. The result is then completed by induction on the dimension.

The key idea of the first proof is to use the characteristic polynomial $\det(A - XI)$ of A and the fundamental theorem of algebra, whose statement we recall.

Theorem 7.3.2 (Fundamental Theorem of Algebra) *Let*

$$p(X) := a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0$$

be a polynomial with coefficients $a_i \in \mathbb{C}$. Assume that $n \geq 1$ and that $a_n \neq 0$. Then p has a complex root, that is, there exists a complex number λ such that

$$p(\lambda) := a_n \lambda^n + a_{n-1} \lambda^{n-1} + \cdots + a_1 \lambda + a_0 = 0.$$

We do not prove this result. For a proof, the reader may consult any book on complex analysis.

Thus, by the fundamental theorem of algebra, the characteristic polynomial of T has a complex root, say, λ.

Definition 7.3.2 The characteristic polynomial of a linear map $T: V \to V$ is that of any of its matrix representations. Note that this is well-defined in the following sense. If A (respectively B) is the matrix of T with respect to the basis $\{v_1, \ldots, v_n\}$ (respectively $\{w_1, \ldots, w_n\}$) then A and B are conjugate: There exists a matrix C such that $A = CBC^{-1}$. Hence $A - \lambda I$ and $B - \lambda I$ are conjugate: $A - \lambda I = C(B - \lambda I)C^{-1}$. Hence, their determinants are the same, that is, the characteristic polynomials of A and B are the same.

A root of the characteristic equation $\det(A - XI) = 0$ is called the *characteristic value* of T. The characteristic equation is an invaluable tool in our understanding of linear transformations or matrices. Note that if α is an eigenvalue of T, then α is a characteristic root, but the converse is not true. See Example 7.2.7. However, if α is a real root of the characteristic equation of T, then α is an eigenvalue of T. For, this means that $\det(T - \alpha I) = 0$. Hence by Theorem 6.4.4, there exists a nonzero vector $v \in V$ such that $(T - \alpha I)v = 0$. That is, v is an eigenvector of T.

Before we proceed to the main result of this section, let us establish an easy result.

Proposition 7.3.3 *If the characteristic equation of T has n distinct real roots, then T is diagonalizable.*

Proof Let $v_1, \ldots, v_n$ be the eigenvectors corresponding to the n distinct roots of A. Then by Lemma 7.2.2, $\{v_1, \ldots, v_n\}$ forms a basis of V. We have already shown that with respect to this basis, the matrix A of T will be diagonal.

□

Theorem 7.3.4 (Spectral Theorem for Symmetric Linear Maps) *Let $T: V \to V$ be a symmetric linear map on a (finite dimensional real) inner product space. Then there exists an orthonormal basis of V consisting of eigenvectors of T.*

The crucial observation towards the proof of this theorem is the fact that any characteristic root of a symmetric linear map is real and hence is an eigenvalue.

Lemma 7.3.5 *All characteristic roots of a symmetric linear map are real. Equivalently, all characteristic roots of a symmetric matrix are real. In particular, there is an eigenvalue of T.*

Proof Let λ be a root of the characteristic polynomial of the symmetric matrix A. Suppose λ is not real, that is, $\operatorname{Im}\lambda \neq 0$. We have

$$\det(A - \lambda I) = 0.$$

Therefore $\det[(A - \lambda I)(A - \overline{\lambda} I)] = 0$ [1].

Writing $\lambda = \operatorname{Re}\lambda + i\operatorname{Im}(\lambda)$, the last relation comes to

$$\det[(A - \operatorname{Re}\lambda I)^2 + (\operatorname{Im}\lambda)^2 I] = 0.$$

Since this last matrix is real, there exists a nonzero vector x such that $[(A - \operatorname{Re}\lambda I)^2 + (\operatorname{Im}\lambda)^2 I]x = 0$ and, in particular,

$$\langle [(A - \operatorname{Re}\lambda I)^2 + (\operatorname{Im}\lambda)^2 I]\, x, x \rangle = 0.$$

We get

$$\langle (A - \operatorname{Re}\lambda I)x, (A - \operatorname{Re}\lambda I)x \rangle + \operatorname{Im}(\lambda)^2 \langle x, x \rangle = 0.$$

Since the left hand side is positive, this is impossible.

$\square$

Definition 7.3.3 Let $T: V \to V$ be any linear map. We say that a vector subspace W is *invariant* under T if $Tw \in W$ for all $w \in W$. We also say that W is an invariant subspace of T.

The second observation needed for the proof of Theorem 7.3.4 is the following lemma:

Lemma 7.3.6 *Let $T: V \to V$ be a symmetric linear map. Assume that W is a vector subspace of V invariant under T. Then*

$$W^{\perp} = \{v \in V \mid \langle v, w \rangle = 0\}$$

is also invariant under T.

Proof This is easily verified. Let $v \in W^{\perp}$. The result follows from the following:

$$\langle Tv, w \rangle = \langle w, Tv \rangle = 0.$$

The first equality is by the symmetry of T. The second is true, since $Tw \in W$ (as W is invariant under T) and $v \in W^{\perp}$.

$\square$

[1] We are using here the formula $\det(AB) = \det A \cdot \det B$ for matrices with complex entries.

Proof (of Theorem 7.3.4) We prove the main theorem by induction on the dimension of the inner product space.

Let P_n be the statement: If X is an n-dimensional inner product space and if $F: X \to X$ is a symmetric linear map, then X has an orthonormal basis consisting of eigenvectors of F.

P_1 is clearly true: For, if V is one dimensional inner product space, let v be any nonzero vector in V. Then $u := v/\|v\|$ is a unit vector and $\{u\}$ is an orthonormal basis of V. If $T: V \to V$ is any linear map, then we already know that there exists a real $\lambda \in \mathbb{R}$ such that $Tu = \lambda u$ (see Example 4.1.10 and Exercise 4.1.9). Hence P_1 is true.

Assume that P_{n-1} is true. Let $T: V \to V$ be a symmetric linear map on an n-dimensional inner product space. By Lemma 7.3.5, there exists an eigenvalue, say, λ of T. Let w be an eigenvector of T with eigenvalue λ. Then $W := \mathbb{R}w$ is a vector subspace of V invariant under W. Hence, $W^\perp$ is also an invariant subspace of T. Note that by restricting the inner product of V to $W^\perp$, $W^\perp$ becomes an inner product space. As $W^\perp$ is the kernel of the linear map $f_w: V \to \mathbb{R}$ given by $f_w(v) := \langle v, w\rangle$, by the rank-nullity theorem, $\dim W^\perp = n - 1$. The restriction T_W of T to W is obviously symmetric. Hence, by induction hypothesis, there exists an orthonormal basis, say, $\{v_1, \dots, v_{n-1}\}$ of W consisting of eigenvectors of T_W. Since $V = W \oplus W^\perp$ is an orthogonal direct sum, the set $\{v_1, \dots, v_{n-1}, v_n := w\}$ is an orthonormal basis of V such that each v_j is an eigenvector of T.

□

As a corollary, we obtain the following result.

Proposition 7.3.7 *Let A be a real symmetric $n \times n$ matrix. Then there exists an orthogonal matrix B such that BAB^{-1} is a diagonal matrix whose entries are the eigenvalues of A.*

Proof Let $T: \mathbb{R}^n \to \mathbb{R}^n$ denote the symmetric map whose matrix with respect to the standard basis of $\mathbb{R}^n$ is A. To wit,

$$Tx := A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Then, by the spectral theorem, there exists an orthonormal basis $\{v_i\}_{i=1}^n$ of $\mathbb{R}^n$ consisting of eigenvectors of T. Let

$$v_i := \sum_j b_{ji} e_j = \begin{pmatrix} b_{1i} \\ \vdots \\ b_{ni} \end{pmatrix}.$$

Let $B = (b_{ij})$. Then B is easily seen to be orthogonal. Note that when we view B as a linear map on $\mathbb{R}^n$, then $Be_i = v_i$ for all i. Hence,

$$B^{-1}AB = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

since

$$B^{-1}AB(e_i) = B^{-1}Av_i = B^{-1}(\lambda_i v_i) = \lambda_i e_i, \quad \text{for all } i.$$

□

As for the second proof of the main theorem, as observed earlier, it suffices to prove the existence of an eigenvalue of T. We warn the reader that this proof is quite demanding, as it requires much more background in diverse fields such as metric spaces, analysis, calculus etc. However, it is quite worthwhile to learn the proof, as it brings out the interplay between the various branches of mathematics and gives a glimpse of the essential unity of the subject.

Lemma 7.3.8 *Let V be a finite dimensional inner product space. Let T be a symmetric linear map on V. Then T has an eigenvalue.*

Proof Our proof has a simple geometric interpretation. What we are going to do is to look for the minor axis of the "ellipse" $\{\langle Tx, x\rangle = 1\}$ (see Exercise 9.6.27).

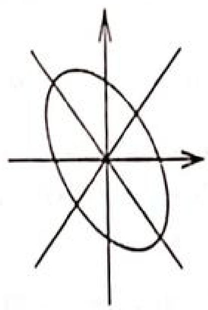


Figure 7.3.1 Axes of an ellipse and eigenvalues.

Let

$$S := \{x \in V \mid \|x\| = 1\}$$

be the unit sphere in V. We use the concepts from the theory of metric spaces. V is a metric space with $d(x,y) := \|x - y\|$. Since V is a finite dimensional inner product space, it is isometric to $\mathbb{R}^n$. Hence the isometric image of S in $\mathbb{R}^n$ is a compact subset of $\mathbb{R}^n$ as it is closed and bounded.

We consider the function $f(x) := \langle Tx, x\rangle$ on S. We now show that the function $T\colon V \to V$ is continuous. Let us fix an orthonormal basis $\{e_i\}_{i=1}^n$ (where $n = \dim V$) of V. Then for any $x \in V$, we can write $x = \sum x_i e_i$ with $\|x\|^2 = \langle x, x\rangle = \sum x_i^2$. Note that $|x_i| \le \|x\|$. We claim that $||Tx|| \leq C\,||x||$ for some constant $C > 0$ and for all $x \in V$.

$$\begin{aligned}||Tx|| &= \left\|\sum_i x_i Te_i\right\| \\ &\leq \sum_i |x_i|\ ||Te_i|| \\ &\leq \sum_i ||x||\ ||Te_i|| \\ &\leq \left(\sum_i ||Te_i||\right) ||x||.\end{aligned}$$

If we take $C = \sum_i ||Te_i||$, the claim obtains. From this it follows that

$$||Tx - Ty|| = ||T(x-y)|| \leq C\,||x-y||$$

and hence the (uniform) continuity of T.

Since T is continuous and the inner product is continuous, f is a real valued continuous function on V: For,

$$\begin{aligned}f(x+h) - f(x) &= \langle T(x+h), x+h\rangle - \langle Tx, x\rangle \\ &= 2\,\langle Tx, h\rangle + \langle Th, h\rangle\end{aligned}$$

where in the second equality we used the symmetry of T. It follows that

$$|f(x+h) - f(x)| \leq 2\,|Tx|\,h + |Th|\,h \to 0$$

as $h \to 0$, thanks to the continuity of T and of the norm function on V.

The function f is continuous on the compact set S. Hence it assumes a minimum, say, λ on S, at $x_0 \in S$.

Claim 1: λ is an eigenvalue and x_0 is an eigenvector.

This follows from

Claim 2: $\langle Tx_0, y\rangle = 0$ for all $y \in V$ with $y \perp x_0$.

Claim 2 $\Rightarrow$ *Claim 1*: *Claim 2* means that Tx_0 must lie in the one-dimensional space spanned by x_0, that is, $Tx_0 = \mu x_0$ for some scalar μ. But this scalar μ must be λ: $\mu = \langle Tx_0, x_0\rangle = \lambda$. Hence $Tx_0 = \lambda x_0$. Thus *Claim 1* and hence the theorem is proved.

We now prove *Claim 2:* The idea of the proof is simple. We consider a curve $g : \mathbb{R} \to S$ such that $g(0) = x_0$ and consider the one variable function $t \mapsto f(g(t))$. Since this function attains a minimum at $t = 0$, its derivative must be 0 at that point. Computing the derivative gives the result.

Now to get to work, let $y \in V$ be such that $\langle x_0, y\rangle = 0$. Let

$$x(t) := x_0 + ty.$$

Then $\|x(t)\|^2 = 1 + t^2\|y\|^2$. Let $u(t) := (1+t^2\|y\|^2)^{-1/2}(x_0+ty)$. Then clearly $u(t) \in S$ for all $t \in \mathbb{R}$. Consider the function $h : t \mapsto \langle Tu(t), u(t)\rangle$. By our assumption on x_0, this function attains a minimum at $t = 0$ and hence $h'(0) = 0$. We compute the derivative of h:

$$\begin{aligned}
h'(t)|_{t=0} &= \frac{d}{dt}\langle Tu(t), u(t)\rangle|_{t=0} \\
&= \frac{d}{dt}((1+t^2\|y\|^2)^{-1}\langle T(x_0+ty), x_0+ty\rangle)|_{t=0} \\
&= \frac{d}{dt}((1+t^2\|y\|^2)^{-1})|_{t=0}\,(\langle T(x_0+ty), x_0+ty\rangle)|_{t=0} \\
&\qquad +(1+t^2\|y\|^2)^{-1}|_{t=0}\,\frac{d}{dt}(\langle T(x_0+ty), x_0+ty\rangle)|_{t=0} \\
&= (-(1+t^2\|y\|^2)^{-2}2t\|y\|^2)|_{t=0}\,\lambda \\
&\qquad +\frac{d}{dt}(\langle Tx_0, x_0\rangle + t\langle Tx_0, y\rangle + t\langle Ty, x_0\rangle + t^2\langle Ty, y\rangle)|_{t=0} \\
&= 0 + \langle Tx_0, y\rangle + \langle Ty, x_0\rangle.
\end{aligned}$$

Since T is symmetric, the last term on the right side is $2\langle Tx_0, y\rangle$. Hence $h'(0) = 0$ if and only if $2\langle Tx_0, y\rangle = 0$. This completes the proof of *Claim 2.*

We may also consider another curve (in place of $x(t)$ above) which arises more geometrically as follows: Let $x_0 \in S$ be as above. Let $y \in S$ with $x \perp y$. Then, x_0 and y span a two-dimensional vector subspace (a plane through the origin) which intersects the sphere S along a great circle (see Figure 7.3.2). This curve on S is nothing other than the unit circle on the

Figure 7.3.2 Sphere.

plane $\mathbb{R}x_0 + \mathbb{R}y$. Since $||x_0|| = 1 = ||y||$ and $\langle x_0, y\rangle = 0$, this curve is given by

$$c(t) = \cos t\, x_0 + \sin t\, y.$$

(We invite the reader to check that $c(t) \in S$.) Proceeding as earlier, we again get the result $\langle Tx_0, y\rangle = 0$.

$\square$

I hope that the reader enjoyed the second proof even though it could have been a little overwhelming. I suggest that he goes through this proof a couple of times more to relish it.

8. Classification of Quadrics

8.1 Conics and Qaudrics

Recall the definition of a conic in coordinate geometry of the plane. A *conic* is the locus of the points in $\mathbb{R}^2$ satisfying a quadratic equation in two variables of the form

$$f(x_1, x_2) := a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + b_1x_1 + b_2x_2 + c = 0. \quad (8.1.1)$$

It is convenient to write the above equation in matrix notation. Let

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}, \; B = (b_1, b_2), \; \text{and} \; X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Then Equation (8.1.1) can be written as

$$f(X) := X^t AX + BX + C = 0. \quad (8.1.2)$$

A *quadric* is the analogue of conics in higher dimension. It is defined as the locus of a quadratic equation in n variables given by

$$\begin{aligned} f(x) &= \sum_i \sum_j a_{ij}x_ix_j + \sum b_kx_k + c = 0 \\ &= X^t AX + BX + C \end{aligned} \quad (8.1.3)$$

where $A = (a_{ij})$ is an $n \times n$ symmetric matrix and $B = (b_1, \dots, b_n)$.

We simplify Equation (8.1.3) using an orthogonal transformation to diagonalize A and then a translation to eliminate as much as possible the linear term $BX + C$.

We suggest that the reader works out the case $n = 2$ and $n = 3$ in the following computations.

First a general computation: A translation leaves the quadratic coefficient matrix A invariant. Let $X = Y + K$, K a constant vector

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

Then we have

$$\begin{aligned} f(X) &= (Y+K)^t A(Y+K) + B(Y+K) + C \\ &= Y^t AY + Y^t AK + K^t AY + K^t AK + BY + BK + C \\ &= Y^t AY + 2K^t AY + BY + BK + C' \quad (C' = C + K^t AK) \\ &= Y^t AY + (2K^t A + B)Y + BK + C' \end{aligned}$$

since $Y^t AK$ is a scalar and hence equals its transpose

$$(Y^t AK)^T = K^t A^t Y = K^t AY.$$

On the other hand, how does a linear change of variables affect the coefficient matrix A? Let $X = PY$. Then

$$\begin{aligned} f(X) &= (PY)^t A(PY) + BPY + C \\ &= Y^t P^t APY + BPY + C \end{aligned} \qquad (8.1.4)$$

Thus the coefficient matrix A changes into $P^t AP$.

We can now effect an orthogonal change of variable so that P is an orthogonal matrix which diagonalizes A, that is, $P^{-1}AP$ is diagonal. Thus under the orthogonal transformation Equation (8.1.3) becomes

$$f(X) = \lambda_1 y_1^2 + \cdots + \lambda_h y_n^2 + b_1 y_1 + \cdots + b_n y_n + c. \qquad (8.1.5)$$

We now eliminate the b_i's associated with nonzero λ_i's by the standard trick of "completing the square". Use the translation. Put $z_i = y_i + \frac{b_i}{2\lambda_i}$. Then the term $\lambda_i y_i^2 + b_i y_i$ becomes

$$\lambda_i^2 z_i^2 - \frac{b_i^2}{4\lambda_i}$$

for $\lambda_i \neq 0$.

Given Equation (8.1.5) we may permute the variables so that $\lambda_1, \ldots, \lambda_r$ are nonzero and also $\lambda_1 \geq \cdots \geq \lambda_r$. Now completing the squares as described above in the indices $1 \leq i \leq r$, we can write Equation (8.1.5) as

$$f(X) = \lambda_1^2 z_1^2 + \cdots + \lambda_r z_r^2 + b_{r+1} z_{r+1} + \cdots + b_n z_n + c' \qquad (8.1.6)$$

where $z_k = y_k$ for $k > r$ and $c' = c - \sum_{i=1}^{r} \frac{b_i^2}{4\lambda_i}$. The linear part can be changed by a rigid motion into the form $d\xi_{r+1}$, for some scalar 'd' without affecting the first r variables as follows: Consider the linear form

$$(z_1, \dots, z_n) \mapsto b_{r+1}z_{r+1} + \cdots + b_n z_n + c'. \tag{8.1.7}$$

By assumption $B = (0, \dots, 0, b_{r+1}, \dots, b_n) \neq 0$. Let $d := ||B^t|| = \sqrt{\sum b_j^2}$ be the norm.

We first use translation to kill the constant term c'. Let k be such that $b_k \neq 0$. Let

$$\eta_i = z_i, \quad i \neq k \quad \text{and} \quad \eta_k = z_i + \frac{c'}{b_k}.$$

Then in the η variables, the linear form Equation (8.1.7) takes the form

$$(\eta_1, \dots, \eta_n) \mapsto b_{r+1}\eta_{r+1} + \cdots + b_n\eta_n. \tag{8.1.8}$$

Now consider the unit vector

$$\frac{1}{d}(0, \dots, 0, b_{r+1}, \dots, b_n).$$

This vector is orthogonal to the unit vectors

$$(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, 0, 1, 0, \dots, 0) \qquad (r \text{ terms }).$$

Hence there exists an orthogonal transformation which takes these vectors to themselves and $\frac{b}{d}(0, \dots, 0, b_{r+1}, \dots, b_n)$ into $(0, \dots, 0, 1, 0, \dots, 0)$. Let the coordinates with respect to this new orthonormal basis be $\xi_1, \dots, \xi_n$. (Note that $\xi_i = \eta_i$ for $1 \leq i \leq r$). Then the linear form Equation (8.1.7) takes the form

$$(\xi_1, \dots, \xi_n) \mapsto d\xi_{r+1}. \tag{8.1.9}$$

Notice that the rigid motion effected so far does not affect the first r-coordinates at all.

Thus Equation (8.1.6) looks like one of the following after all the change of coordinates.

$$f(\xi) = \lambda_1 z_1^2 + \cdots + \lambda_r \xi_r^2 + d\xi_{r+1} \tag{8.1.10}$$

$$f(\xi) = \lambda_1 z_1^2 + \cdots + \lambda_r \xi_r^2 + c' \tag{8.1.11}$$

where $\lambda_1 \geq \cdots \geq \lambda_r$, λ_i nonzero, $d > 0$.

We have thus proved the following result.

Theorem 8.1.1 *Under the Euclidean group of rigid motions, any quadratic form in n variables such as*

$$f(x) = \sum_i \sum_j a_{ij} x_i x_j + \sum b_k x_k + c = 0$$
$$= X^t A X + BX + C$$

can be brought into one of the forms,

$$f(\xi) = \lambda_1 x_1^2 + \cdots + \lambda_r \xi_r^2 + d\xi_{r+1}$$

or

$$f(\xi) = \lambda_1 x_1^2 + \cdots + \lambda_r \xi_r^2 + c'$$

8.1.1 Classification of Quadrics

We present the standard forms of conics in $\mathbb{R}^2$ and those of quadrics in $\mathbb{R}^3$ in the following tables.

Canonical Forms of Conics

No.	Equation	Name of the Conic
1.	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$	Ellipse
2.	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1$	Imaginary ellipse
3.	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$	Point ellipse
4.	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$	Hyperbola
5.	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$	Intersecting lines
6.	$x^2 = 4cy$	Parabola
7.	$x^2 = a^2$	Parallel lines
8.	$x^2 = -a^2$	Imaginary lines
9.	$x^2 = 0$	Coincident lines

Canonical Forms of Quadrics

No.	Equation	Name of the Quadric
1.	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$	Ellipsoid
2.	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=-1$	Imaginary ellipsoid
3.	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$	Hyperboloid of one sheet
4.	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1$	Hyperboloid of two sheets
5.	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=2z$	Elliptic paraboloid
6.	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=2z$	Hyperbolic paraboloid
7.	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=0$	Point (imaginary elliptic cone)
8.	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0$	Elliptic cone
9.	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$	Elliptic cylinder
10.	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=-1$	Imaginary elliptic cylinder
11.	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$	Hyperbolic cylinder
12.	$\frac{x^2}{a^2}=4py$	Parabolic cylinder
13.	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=0$	Line (imaginary intersecting plane)
14.	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$	Intersecting plane
15.	$x^2=a^2$	Parallel planes
16.	$x^2=-a^2$	Imaginary parallel planes
17.	$x^2=0$	Coincident planes

The figures of some quadric surfaces can be found at the end of thi chapter.

8.2 Computational Examples

We shall illustrate the above theoretical results in some concrete cases.

Example 8.2.1 Let us consider the conic defined by $2xy = 1$. The matrix A is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

By (geometric) inspection we see that $e_1 + e_2$ is an eigenvector with eigenvalue 2. Hence the other eigenvector must be perpendicular to this vector. Thus $\pm(e_2 - e_1)$ is an eigenvector with eigenvalue -2. We choose as the orthonormal basis the basis consisting of eigenvectors $\{(e_1 + e_2)/\sqrt{2}, (e_2 - e_1)/\sqrt{2}\}$. (This choice is made so that the new axes are got by a rotation from the standard ones). Let the coordinates with respect to this new basis be denoted by (u, v). Then the relation between the old and the new ones is given by $x = (u - v)/\sqrt{2}$ and $y = (u + v)/\sqrt{2}$ (see the section on the coordinates with respect to an orthonormal basis). Thus the equation reads in the new coordinates as $u^2 - v^2 = 1$.

Example 8.2.2 Consider the conic given by the equation

$$2x^2 - 73xy + 23y^2 + 140x - 20y + 50 = 0.$$

The matrix A is

$$\begin{pmatrix} 2 & -36 \\ -36 & 23 \end{pmatrix}.$$

The characteristic equation is given by

$$\det \begin{pmatrix} 2 - \lambda & -36 \\ -36 & 23 - \lambda \end{pmatrix} = 0.$$

Thus the eigenvalues are the roots of the equation $\lambda^2 - 25\lambda - 1250 = 0$. The eigenvalues are 50 and -25. We wish to find the corresponding eigenvectors. We thus have to solve for the system $Ax = \lambda x$. That is, solve

$$\begin{pmatrix} 2 & -36 \\ -36 & 23 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 50 \begin{pmatrix} x \\ y \end{pmatrix}.$$

This results in the system of equations

$$\begin{aligned} 2x - 36y &= 50x \\ -36x + 23y &= 50y. \end{aligned}$$

The first equation becomes $-48x - 36y = 0$ so that the column vector $(3,-4)^T$ is an eigenvector. The vector perpendicular to this with the "correct orientation" is $(4,3)^T$. Thus the orthogonal matrix which diagonalizes A is

$$\begin{pmatrix} 3/5 & 4/5 \\ -4/5 & 3/5 \end{pmatrix}.$$

The given equation becomes

$$50u^2 - 25v^2 + 140x - 20y + 50 = 0.$$

The coordinates are related by $x = (3u + 4v)/5$ and $y = (-4u + 3v)/5$. Using this substitution, we get $50u^2 - 25v^2 + 100u + 100v + 50 = 0$. We now complete the squares

$$50(u+1)^2 - 50 - 25(v-2)^2 + 100 + 50 = 0.$$

We now effect a translation $X = u + 1$ and $Y = v - 2$ to get the equation in the form: $50X^2 - 25Y^2 + 100 = 0$. This can be cast in the standard form $X^2/2 - Y^2/4 = -1$. Note that the standard coordinates and the last coordinates are related by

$$(x,y) = \left(\frac{4}{5}X - \frac{3}{5}Y + 1, \frac{3}{5}X + \frac{4}{5}Y + 2\right).$$

We can use this information to draw the conic.

Example 8.2.3 Consider the quadric in $\mathbb{R}^3$ given by

$$f(x,y,z) := 4xz + 4y^2 + 8y + 8 = 0.$$

The matrix of coefficients of the second degree terms is

$$A := \begin{pmatrix} 0 & 0 & 2 \\ 0 & 4 & 0 \\ 2 & 0 & 0 \end{pmatrix}.$$

A (geometric) look at the matrix shows that e_2 is an eigenvector with eigenvalue 4. (For, observe that the second column is Ae_2.) Also, the plane perpendicular to the y-axis, namely, the xz-plane is mapped by this operator A to itself. Another look shows that $e_1 + e_3$ is an eigenvector with eigenvalue 2 and $e_1 - e_3$ is an eigenvector with eigenvalue -2. Thus as an orthonormal basis consisting of eigenvectors of A, we take

$$\{(e_1 + e_3)/\sqrt{2}, e_2, (e_1 - e_3)/\sqrt{2}\}.$$

The orthogonal matrix which takes the standard basis to this eigen basis is, of course, given by

$$\begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}.$$

Now let x_1, y_2, z_1 be the new coordinates associated with this eigen basis. The old and new coordinates are related by $x = (1/\sqrt{2})(x_1 - z_1)$, $y = y_1$ and $z = (1/\sqrt{2})(x_1 - z_1)$ (see Section 5.8). Hence the given polynomial becomes in the new coordinates

$$2x_1^2 + 4y_1^2 - 2z_1^2 + 8y_1 + 8 = 0.$$

We now complete the squares in the y_1 variables to get

$$2x_1^2 + 4(y_1 + 1)^2 - 4 - 2z_1^2 + 8y_1 + 8 = 0.$$

We now effect a change of coordinates $x_2 = x_1$, $y_2 = (y_1 + 1)$ and $z_2 = z_1$ so that the given polynomial becomes

$$2x_2^2 + 4y_2^2 - 2z_2^2 + 4 = 0.$$

Thus the given conic section is a hyperboloid of two sheets as can be seen from the table.

Example 8.2.4 Consider the quadric surface defined by

$$3x^2 + 2xy + 4yz + 2xz - 2x - 14y - 2z - 9 = 0.$$

Then

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix} \text{ and } b := \begin{pmatrix} -1 \\ -7 \\ -1 \end{pmatrix}.$$

This time we look for a centre of the quadric. Thus we wish to find solutions of $Ax = b$. Solving the system of equations we get $(-3/2, 5/4, 17/4)$ as the centre of the quadric. The eigenvalues are easily found to be 1, 4 and -2. Using the translation $x = x' - 3/2$, $y = y' + 5/4$ and $z = z' + 17/4$, we eliminate the first degree terms without affecting the second degree terms. Thus to find the transformed equation we need only commute the constant term with respect to the above substitution:

$$3\left(-\frac{3}{2}\right)^2 + 2\left(-\frac{3}{2}\right)\left(\frac{5}{4}\right) + 2\left(-\frac{3}{2}\right)\left(\frac{17}{4}\right) + 4\left(\frac{5}{4}\right)\left(\frac{17}{4}\right)$$
$$-2\left(-\frac{3}{2}\right) - 14\left(\frac{5}{4}\right) - 2\left(\frac{17}{4}\right) - 9 = 20.$$

Thus the standard form of the given quadric is $x'^2 + 4y'^2 - 2z'^2 + 20 = 0$.

Exercise 8.2.1 Reduce the following into standard forms:

(1) $11x^2 + 6xy + 19y^2 - 80$.

(2) $2x^2 - 5y^2 + 3x + 10y$.

(3) $16x^2 - 24xy + 9y^2 - 30x - 40y$.

(4) $8x^2 - 12xy + 17y^2 - 80$.

(5) $3x^2 + 2xy + 3y^2 - 4$.

(6) $5x^2 - 8xy + 5y^2 - 9$.

(7) $2x^2 + 3xy - 2y^2 - 10$.

(8) $2x^2 + 4xy + 2y^2 - 64$.

Exercise 8.2.2 Reduce the following into standard forms:

(1) $4xz + 4y^2 + 8y + 8$.

(2) $9x^2 - 4xy + 6y^2 + 3z^2 + 2\sqrt{5}x + 4\sqrt{5}y + 12z + 16$.

(3) $x^2 + y^2 - 7z^2 - 2xy - 4xz - 4yz + 8y + 14z - 6$.

(4) $x^2 + 4y^2 + 4z^2 + 4xy - 4xz - 8yz + 2x + 8y + 7$

(5) $8x^2 - 4xy + 4xz - 2yz + 2x + 2y - 6z - 20$.

(6) $x^2 + 6xy - 2y^2 - 3xz + z^2$.

(7) $-2x^2 - 11y^2 - 5z^2 + 4xy + 16yz + 20xz$

(8) $3x^2 - y^2 - 3z^2 + 3t^2 - 4xy - 10yz$.

Figure 8.1.1 Hyperboloid of one sheet.

Figure 8.1.2 Ellipsoid.

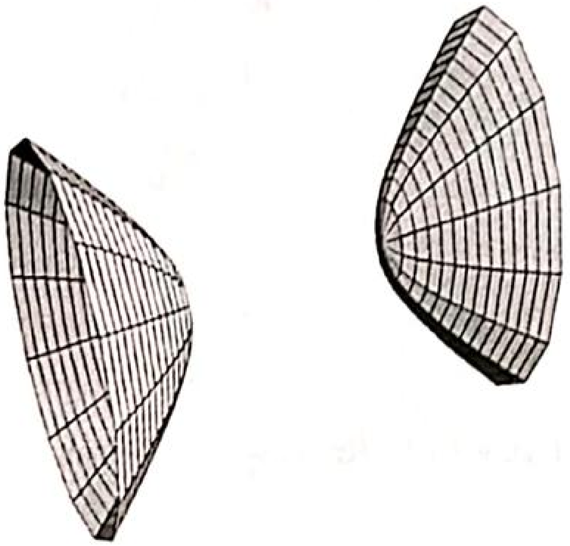


Figure 8.1.3 Hyperboloid of two sheets.

Figure 8.1.4 Elliptic paraboloid.

Figure 8.1.5 Hyperbolic paraboloid.

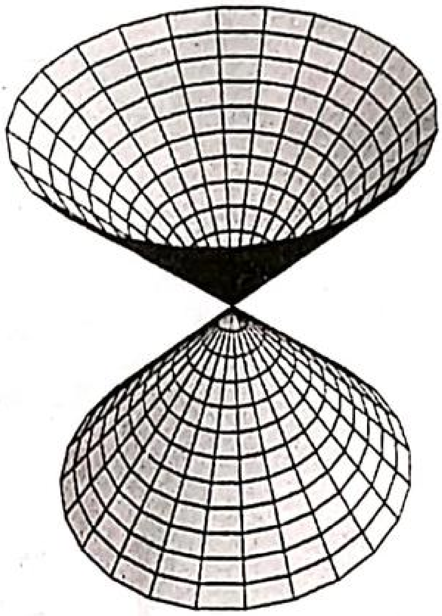


Figure 8.1.6 Elliptic cone.

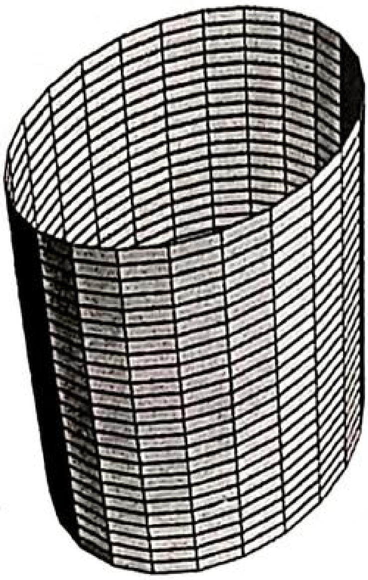


Figure 8.1.7 Elliptic cylinder.

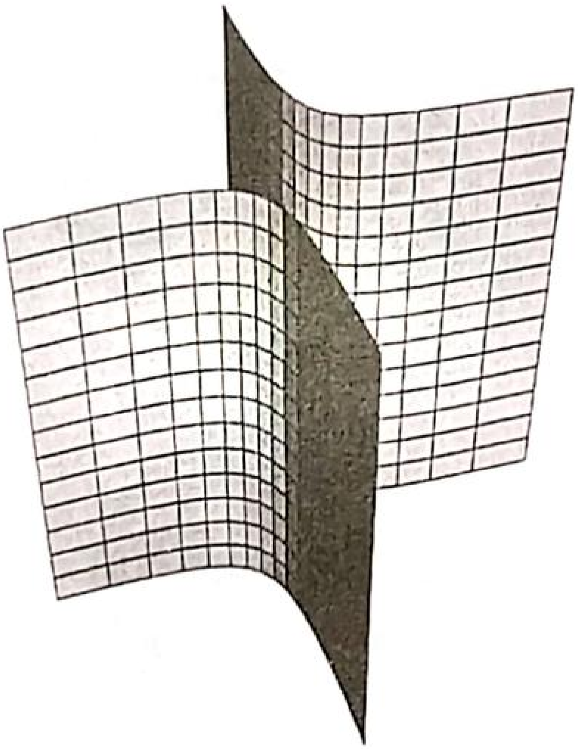


Figure 8.1.8 Hyperbolic cylinder.

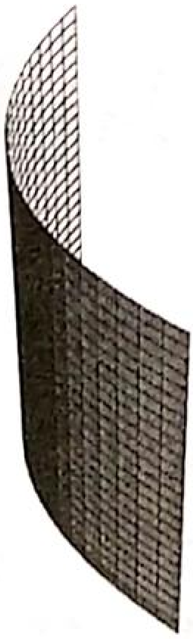


Figure 8.1.9 Parabolic cylinder.

9. Review Problems

In this chapter, we give lots of problems for practice. Some of the problems below appeared either as lemmas/theorems or were listed as exercises earlier. The point of giving the collection here is to help the reader to assess his overall understanding of linear algebra.

Unless specified otherwise, V stands for a finite dimensional vector space over $\mathbb{R}$, $\mathbb{R}^n$ is always equipped with the dot product or the Euclidean inner product $\langle x, y\rangle := \sum_{i=1}^{n} x_i y_i$. Here $x = (x_1, \ldots, x_n) = \sum_i x_i e_i$ where e_i are the standard basis vectors.

If a problem is just a statement, you are asked to provide a proof for it.

9.1 Linear Equations

Exercise 9.1.1 Find a necessary and sufficient condition for either the sum of two solutions or the scalar multiple by a number α ($\alpha \neq 1$), to be a solution again of the same system of equations.

Exercise 9.1.2 Under what conditions will a given linear combination of any solutions of a given non-homogeneous system of linear equations be again a solution of the same system?

Exercise 9.1.3 Consider all possible cases encountered in solving systems of linear equations involving two or three unknowns. Give the geometric interpretation in each case.

9.2 Linear Dependence

Exercise 9.2.1 Prove that a set of vectors containing the null vector is linearly dependent.

Exercise 9.2.2 Prove that a set of vectors, two of whose vectors differ only by a scalar multiple is linearly dependent.

Exercise 9.2.3 Prove that if, in a set of vectors, some subset is linearly dependent, then the full set is linearly dependent.

Exercise 9.2.4 Prove that any subset of a linearly independent set is linearly independent.

Exercise 9.2.5 Suppose $\{x_i\}_{i=1}^{m}$ be linearly independent, but $\{y\} \cup \{x_i\}$ is not. Then y can be written uniquely as a linear combination of x_i.

Exercise 9.2.6 Is there a converse of Exercise 9.2.5?

Exercise 9.2.7 Let a, b, c be distinct real numbers. Is the following set of polynomials linearly independent?

$$\{(X-a)(X-b), (X-b)(X-c), (X-c)(X-a)\}.$$

Exercise 9.2.8 Prove that in $\mathcal{P}_n$, any finite set consisting of polynomials of different degrees, not containing the zero polynomial is linearly independent.

Exercise 9.2.9 Determine whether the following sets are linearly dependent.

(1) $\{x_1 = (-3, 1, 5), x_2 = (6, -2, 15)\}$.

(2) $\{x_1 = (-1, 2, 3), x_2 = (2, 5, 7), x_3 = (3, 7, 10 + \epsilon), \epsilon \neq 0\}$.

(3) $\{x_1 = (4, -12, 28), x_2 = (-7, 21, -49)\}$.

9.3 Basis and Dimension

Exercise 9.3.1 Prove that

(1) Any nonzero vector can be enlarged into a basis.

(2) Any linearly independent set can be enlarged to a basis of the vector space.

Exercise 9.3.2 Find two different basis of $\mathbb{R}^3$ having $e_1 = (1, 0, 0)$ and $e_2 = (0, 1, 0)$ in common.

Exercise 9.3.3 Prove that in the space $\mathcal{P}_n$, any set of nonzero polynomials containing one polynomial of each degree k, $k = 0, 1, \ldots, n$ is a basis.

Exercise 9.3.4 Show that any basis is a maximal linearly independent set and a minimal set of generators.

Exercise 9.3.5 Find the coordinates of the polynomial $t^5-t^4+t^3-t^2+t+1$ in each of the following bases of $\mathcal{P}_5$.

(1) $\{1, t, t^2, t^3, t^4, t^5\}$.

(2) $\{1, t+1, t^2+1, t^3+1, t^4+1, t^5+1\}$.

(3) $\{1+t^3, t+t^3, t^2+t^3, t^3, t^4+t^3, t^5+t^3\}$.

Exercise 9.3.6 Prove that the span of an arbitrary subset of a vector space V is a vector subspace.

Exercise 9.3.7 Let $W \subset V$ be a subspace. Show that $\dim W \leq \dim V$. When does equality hold?

Exercise 9.3.8 Prove that in an n-dimensional vector space V, a vector subspace W of dimension k can be found for any $k = 1, 2, \ldots, n$.

Exercise 9.3.9 Construct a basis of $\mathcal{P}_5$ consisting of polynomials of degree 5. Can you construct a basis in which the degree of its members ≤ 4?

Exercise 9.3.10 Find a basis and the dimension of the linear subspace of $\mathbb{R}^n$ given by $x_1 + x_2 + \cdots + x_n = 0$.

Exercise 9.3.11 In $\mathcal{P}_n$, each of $W_1 = \{f(0) = 0\}$, $W_2 = \{f(1) = 0\}$, $W_3 = \{f(a) = 0\}$, $W_4 = \{f(0) = f(1) = 0\}$ is a vector subspace. Find their dimensions.

Exercise 9.3.12 Find the coordinates of the polynomial $f(X) = \sum_{i=1}^{n} a_i X^i$ with respect to the bases:

(1) The basis $\{1, X, X^2, \ldots, X^n\}$.

(2) The basis $\{1, (X-\alpha), (X-\alpha)^2, \ldots, (X-\alpha)^n\}$.

Exercise 9.3.13 Prove that if the sum of the dimensions of two vector subspaces of an n-dimensional vector space exceed n, then the subspaces have a nonzero vector in common.

In $\mathbb{R}^3$, is it possible to have two subspaces W_1, and W_2 such that $\dim W_1 = \dim W_2 = 2$, $W_1 \cap W_2 = \{0\}$?

Give the geometric meaning of the above. Can you generalize this?

Exercise 9.3.14 Prove that the following set of vectors in $\mathbb{R}^n$ form a linear subspace and find a basis and the dimension of each.

(1) All n-vectors whose first and the last coordinates are equal.

(2) All n-vectors whose even entries are zero.

(3) All n-vectors whose even entries are equal.

Exercise 9.3.15 Prove that

$$\dim W_1 + \dim W_2 = \dim(W_1 + W_2) + \dim(W_1 \cap W_2).$$

9.4 Linear Transformations

Exercise 9.4.1 Find all linear transformations on a vector space having dimension 1.

Exercise 9.4.2 Prove that any linear transformation maps a linearly dependent set to a linearly dependent set.

Exercise 9.4.3 Is it true that any linearly independent set is mapped to another linearly independent set under a linear transformation?

Exercise 9.4.4 If $W \subseteq V$, show that $T(W)$ is a subspace and also that $\dim T(W) \leq \dim W$.

Exercise 9.4.5 Show that a linear transformation is determined once we know its effect on a basis.

Exercise 9.4.6 Let $\{e_i\}_{i=1}^n$ be a basis of V. Also let $\{y_i\}_{i=1}^n \subseteq W$, another vector space. Show that there exists a unique linear transformation T such that $Te_i = y_i$.

Exercise 9.4.7 Let $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_n\}$ be arbitrary subsets of V. Does there exist a unique linear transformation T such that $Tx_i = y_i$?

Exercise 9.4.8 Let W be a subspace of V and $T\colon W \to X$ be a linear transformation. Show that there exists $\tilde{T}\colon V \to X$ such that $\tilde{T}(w) = T(w)$ for all $w \in W$. Is this $\tilde{T}$ unique?

Exercise 9.4.9 Show that the kernel and image of a linear transformation are linear subspaces.

Exercise 9.4.10 If W is a subspace of V, is there a linear transformation $T\colon V \to Y$ (Y is given) such that $\ker T = W$? (Answer depends on whether $\dim Y \geq \dim V$ or not!).

Exercise 9.4.11 Find two different linear transformations having the same kernel and image.

Exercise 9.4.12 Show that the multiplication of 2×2 matrices by $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ on the left (right) is a linear transformation. Find its matrix with respect to the basis

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\} \text{ of } M(2, \mathbb{R}).$$

Exercise 9.4.13 Find the matrix of $\frac{d}{dx}$ on $\mathcal{P}_n$ with respect to the basis
(1) $\{1, X, X^2, \ldots, X^n\}$. (2) $\left\{1, (X-\alpha), \frac{(X-\alpha)^2}{2!}, \cdots \frac{(X-\alpha)^n}{n!}\right\}$.

Exercise 9.4.14 What change will the matrix of a linear transformation undergo if two vectors $\{e_i, e_j\}$ of the basis $\{e_1, \ldots, e_n\}$ are interchanged?

Exercise 9.4.15 Prove that the matrices of the same linear transformation with respect to two different bases coincide if and only if the transition matrix from the bases commute with the matrix of the linear transformation with respect to one of the bases.

Exercise 9.4.16 Find those subspaces of $\mathcal{P}_n$ which remain invariant under $\frac{d}{dx}$.

Exercise 9.4.17 Find the kernel of the following linear transformations:

(1) $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ x-y \end{pmatrix}.$$

(2) $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix}.$$

(3) $T\colon \mathbb{R}^4 \to \mathbb{R}^2$ given by

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ z+w \end{pmatrix}.$$

Exercise 9.4.18 Let $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Show that $Tx = y$ has a solution only if $z - x - y = 0$. Is T onto? Find a basis for the range. What is the kernel of T?

Exercise 9.4.19 Let $T: \mathbb{R}^2 \to \mathbb{R}^3$ be given by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ x+y \\ y \end{pmatrix}.$$

Find the kernel of T. Is T one-one? Is T onto?

Exercise 9.4.20 Find the kernel and range of T, and their bases and dimensions when T is given by

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 3 & -1 \\ -1 & 1 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Is T one-one?

Exercise 9.4.21 If T is given by

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x-y \\ x+2y \\ z \end{pmatrix}$$

find a basis for the kernel and range of T. Verify the dimension formula.

Exercise 9.4.22 True or false: If $\{T(x_i)\}_{i=1}^{n}$ is linearly independent then $\{x_1, \dots, x_n\}$ is so?

Exercise 9.4.23 Prove that the linear map $T: V \to W$ is one-one if and only if $\dim(\text{Im } T) = \dim V$.

Exercise 9.4.24 Construct a linear transformation $T: \mathbb{R}^2 \to \mathbb{R}^3$ such that $\text{Im }(T) = \{x+y+z=0\}$.

Exercise 9.4.25 Construct a linear transformation $T: \mathbb{R}^2 \to \mathbb{R}^4$ such that $\text{Im }(T) = \{\sum_{i=1}^{4} x_i = 0\}$.

Exercise 9.4.26 Can you formulate an exercise of which Exercises 9.4.24 and 9.4.25 are special cases?

Exercise 9.4.27 Find the linear transformations $F: \mathbb{R} \to \mathbb{R}$ such that

(1) $F(-1) = 2$.

(2) $F(3) = 0$.

(3) $F(0) = -2$.

(4) $F(1) = 2$ and $F(2) = 3$.

Exercise 9.4.28 Find the linear transformations $F : \mathbb{R}^2 \to \mathbb{R}^2$ such that

(1) $F(-1,1) = (0,1)$.

(2) $F(1,1) = (2,0)$, $F(-1,1) = (0,1)$.

(3) $F(A) = F(B)$ where $A = \{(x,y) \mid y = 2x\}$, $B = \{(x,y) \mid x = 0\}$.

Write the matrix relative to the standard basis in each case.

Exercise 9.4.29 Define

$$F \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ y \end{pmatrix}.$$

What is the matrix of F? Let S be the square with vertices $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$. What is $F(S)$?

Exercise 9.4.30 Let $\ell_1 = \mathbb{R}(0,0,1)^t$ and $\ell_2 = \mathbb{R}(2,1,0)^t$. Find a map $F : \mathbb{R}^3 \to \mathbb{R}^3$ such that $F(\ell_1) = F(\ell_2)$.
Also, find F such that $F(0,0,1)^t = (2,1,0)^t$.

Exercise 9.4.31 Find $F : \mathbb{R}^3 \to \mathbb{R}^3$ such that $\{z = 0\}$ goes to $\{y = 0\}$.

Exercise 9.4.32 Let $\mathcal{P}_1$ be the plane $x + y - z = 0$ and $\mathcal{P}_2$ be the plane $2y + z - x = 0$. Find $F : \mathbb{R}^3 \to \mathbb{R}^3$ such that

(1) $F(\mathcal{P}_1) = \mathcal{P}_2$.

(2) $F(\mathcal{P}_1) = \mathcal{P}_2$ and $F(\mathcal{P}_2) = \mathcal{P}_1$.

Exercise 9.4.33 Let $\{x = 0\}$ be the yz-coordinate plane. Find a linear transformation $F : \mathbb{R}^3 \to \mathbb{R}^3$ which maps it to a parallel plane $\{x = 1\}$.

9.5 Euclidean Spaces

Exercise 9.5.1 Prove that in a Euclidean space the zero vector is the only one which is orthogonal to all vectors. Show that if $\langle a, x\rangle = \langle b, x\rangle$ for all $x \in V$, then $a = b$.

Exercise 9.5.2 If $\{x_1, \cdots, x_n\}$ is an orthogonal set, then $\{\alpha_1 x_1, \cdots, \alpha_n x_n\}$ is an orthogonal set for all $\alpha_i \in \mathbb{R}$.

Exercise 9.5.3 If $x \perp y_i$, $1 \le i \le n$, then x is perpendicular to any linear combination of y_i.

Exercise 9.5.4 Prove that an orthogonal set of nonzero vectors is linearly independent.

Exercise 9.5.5 Apply the Gram-Schmidt process to

$$x_1 = (1, -2, 2), x_2 = (-1, 0, 1), x_3 = (5, -3, -7)$$

in $\mathbb{R}^3$ with the dot product.

Exercise 9.5.6 Prove that the inner product of any two vectors x and y of a Euclidean space is expressed in terms of their coordinates with respect to a certain fixed basis $\{e_i\}$ by the formula $\langle x, y \rangle := \sum x_i y_i$ if and only if $\{e_i\}$ is an orthonormal basis.

Exercise 9.5.7 Find the dimension of the subspace formed by all vectors x such that $\langle x, a \rangle = 0$ for a fixed vector a.

Exercise 9.5.8 Let V have a basis $\{e_i\}$ over $\mathbb{R}$. We can then define an inner product on V such that $\{e_i\}$ becomes an orthonormal basis with respect to this inner product.

Exercise 9.5.9 Define an inner product on $\mathcal{P}^n$ such that

$$P_k(t) = \frac{t^k}{k!}, \ k = 0, 1, \ldots$$

becomes orthonormal.

Exercise 9.5.10 On $\mathcal{P}^n$ define $\langle p, q \rangle := \int_0^1 p(t)q(t)\,dt$. Is this an inner product?

Exercise 9.5.11 Problems on orthogonal complements:

(1) $(L^\perp)^\perp = L$.

(2) $V \subseteq W$ implies $V^\perp \supseteq W^\perp$.

(3) $(V + W)^\perp = V^\perp \cap W^\perp$.

(4) $E = V \oplus W$ implies $E = V^\perp \oplus W^\perp$.

Exercise 9.5.12 In $\mathcal{P}^n$, define $\langle f, g \rangle := \sum_{i=0}^n a_i b_i$. Here $f(x) = \sum a_i x_i$, $g(x) = \sum b_i x_i$. Find the orthogonal complement of all polynomials satisfying the condition $f(1) = 0$ and do the same for the subspace of all polynomials of even degree.

Exercise 9.5.13 Prove the cosine law for triangles given by x and y.

Exercise 9.5.14 Prove the Pythagoras theorem and its converse, namely, that two vectors x and y are orthogonal if and only if $||x-y||^2 = ||x||^2 + ||y||^2$.

Exercise 9.5.15 Prove that $||x|| = ||y||$ if and only if $x+y \perp x-y$. What is the geometric meaning underlying this?

Exercise 9.5.16 Let $x \in \mathbb{R}^n$ be such that $||x|| = 1$. Let $\cos\alpha_i := (x, e_i)$, $\{e_i\}$ an orthonormal basis. Then $\sum \cos^2\alpha_i = 1$. Do you understand the meaning of this?

Lines

Let $\ell(p;d) := \{p + td \mid t \in \mathbb{R}\} = p + \mathbb{R}d$ for $p, d \in \mathbb{R}^n$ fixed and $d \neq 0$.

Exercise 9.5.17 Two lines $\ell(p;d_1)$ and $\ell(p;d_2)$ are the same if and only if $d_1 = \alpha d_2$ for some nonzero $\alpha \in \mathbb{R}$.

Exercise 9.5.18 Two lines $\ell(p;d)$ and $\ell(q;d)$ are equal if and only if $q \in \ell(p;d)$.

Exercise 9.5.19 Two lines $\ell(p;d_1)$ and $\ell(p;d_2)$ are said to be parallel if and only if their direction vectors are parallel (that is, $d_1 = \alpha d_2$, $\alpha \neq 0$).

Given a line ℓ, and $q \notin \ell$, there exists a unique line ℓ' with $q \in \ell'$ and $\ell' \parallel \ell$.

Exercise 9.5.20 Two distinct points p, q determine a line. In fact the line is $\{p + t(q-p) \mid t \in \mathbb{R}\}$.

Exercise 9.5.21 Two vectors are linearly dependent if and only if they lie on the same line through the origin.

Exercise 9.5.22 Given two parallel lines $\ell(p;d)$ and $\ell(q;d')$, then either $\ell(p;d) = \ell(q;d')$ or $\ell(p;d) \cap \ell(q;d') = \emptyset$.

Exercise 9.5.23 Given two lines $\ell(p;d)$ and $\ell(q;d)$ which are not parallel, prove that their intersection is either empty or consists of exactly one point.

9.6 Problems in Linear Geometry

Exercise 9.6.1 Find the angle between the vectors $(1,1,1)$ and $(1,0,1)$ in $\mathbb{R}^3$. Find a vector of length π perpendicular to both these vectors.

Exercise 9.6.2 Find the orthogonal projection of (2,0) in the direction of the vector (1,1).

Exercise 9.6.3 Let $x, y \in V$.

(1) What can you say when $||x|| + ||y|| = ||x + y||$?

(2) Show that $|\,||x|| - ||y||\,| \le ||x - y||$.

(3) Assume that $||x|| = ||y|| \ne 0$. What can you say about $x - y$ and $x + y$?

Exercise 9.6.4 In V, we define $d(x, y) := ||x - y||$. Show that d is a distance function, that is, a metric. Do you recognize Exercise 9.6.3 (2) now?

Exercise 9.6.5 Find the distance between $(1, 2, 0)$ and $(-1, 3, 4)$ in $\mathbb{R}^3$.

Exercise 9.6.6 Let the elements of $\mathbb{R}^2$ be written with respect to the standard basis $e_1 = (1, 0)$ and $e_2 = (0, 1)$. Thus any $u \in \mathbb{R}^2$ is written as $u = (x, y)$ if $u = xe_1 + ye_2$. Define a map $\langle\ ,\rangle$ from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}$ as follows:

$$\langle u, v\rangle := (2x + y)x_1 + (x + 5y)y_1$$

where $u = (x, y)$ and $v = (x_1, y_1)$. Show that $\langle\ ,\rangle$ defines a new inner product on $\mathbb{R}^2$.

If you are curious as to how I thought of this crazy definition, the clue lies in the spectral theorem and the characterization of positive definite matrices in terms of their "principal minors." Here the matrix is $\begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}$.

The above exercise is simple but highly instructive; it needs perhaps some high school algebra!

Exercise 9.6.7 Let $\mathcal{P}_n$ be the space of all polynomials of degree $\le n$. What is the dimension of $\mathcal{P}_n$? Define

$$\langle p, q\rangle := \int_0^1 p(x)q(x)dx \text{ and } (p, q) := \int_{-1}^1 p(x)q(x)dx.$$

What is the length of $p(x) = x$ with respect to $\langle\ ,\rangle$ and $(\ )$? In $\mathcal{P}_2$, apply the Gram-Schmidt process to the basis $\{1, X, X^2\}$ with respect to the above inner products.

Exercise 9.6.8 Find the matrix of the linear map $A : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$A(x, y) = (2x + 3y, x - y)$$

with respect to the orthonormal basis $\{(1, 1)/\sqrt{2}, (1, -1)/\sqrt{2}\}$.

Exercise 9.6.9 Write down all the orthogonal transformations of $\mathbb{R}^2$.

Exercise 9.6.10 Let W be a vector subspace of a finite dimensional inner product space V. Show that there exists a unique vector subspace $W^\perp$ of V such that any $x \in V$ can be written uniquely as $x := y + z$ with $y \in W$ and $z \in W^\perp$ with $\langle W, W^\perp \rangle = 0$.

Exercise 9.6.11 Let V be any vector space over $\mathbb{R}$ not necessarily with an inner product. A *norm* on V by definition is a function $|| \ || : V \to \mathbb{R}$ satisfying the following conditions:

(1) $||x|| \geq 0$ and $||x|| = 0$ if and only if $x = 0$.

(2) $||\alpha x|| = |\alpha| \, ||x||$ for all $\alpha \in \mathbb{R}$ and $x \in V$.

(3) $||x + y|| \leq ||x|| + ||y||$, for $x, y \in V$.

Thus in an inner product space we have a naturally defined norm $||x|| := \langle x, x \rangle^{\frac{1}{2}}$. However there may exist other norms on a vector space.

(a) Show that if V is a finite dimensional vector space over $\mathbb{R}$ and $\{v_i\}_{i=1}^n$ is a basis of V, then for $x = \sum_i x_i v_i$, the functions

$$\begin{aligned} ||x||_1 &:= \sum_{i=1}^n |x_i|, \\ ||x||_\infty &:= \max_{1 \leq n} \{|x_i|\}, \\ ||x||_2 &:= \left(\sum_{i=1}^n {x_i}^2 \right)^{1/2}, \end{aligned}$$

are norms on V.

(b) Let $|| \ ||$ be a norm on V. Show that $d(x, y) := ||x - y||$ is a metric on V (see Exercise 9.6.4).

(c) Let $B(x, r) := \{y \in V \mid ||x - y|| < r\}$. Then $B(x, r)$ is called an open ball of radius r centred at x. Let $V = \mathbb{R}^n$. Let $v_i = e_i$ be the standard basis vectors. Sketch the balls of radius 1 centred at the origin with respect to various norms of (a).

(d) Show that there exists constants C_1 and C_2 such that

$$C_1 \, ||x|| \leq ||x||_1 \leq C_2 \, ||x||,$$

where the above are any two norms of (a) on $\mathbb{R}^n$.

(e) Show that $(x_k) \in \mathbb{R}^n$ converges to $x \in \mathbb{R}^n$ if and only if x_i^k converges to x_i for $1 \leq i \leq n$. Here $x_k := \sum_i x_i^k v_i$ and the convergence is with respect to the metric defined by *any* of the above norms.

(f) Let $M(n, \mathbb{R})$ be the set of all $n \times n$ matrices with real entries. Then $M(n, \mathbb{R})$ is a vector space of dimension n^2 over $\mathbb{R}$ under "natural operations". Let $||A|| := \max_{1 \leq i,j \leq n} \{|a_{ij}|\}$.

(i) Show that $|| \; ||$ is a norm on $M(n, \mathbb{R})$.

(ii) For any $x \in \mathbb{R}^n$ where $\mathbb{R}^n$ is endowed with the Euclidean norm, we have $||Ax|| \leq n \, ||A|| \; ||x||$.

(iii) Can you think of more natural norms which use the fact that A is a (linear) map on $\mathbb{R}^n$ rather than being just a vector in $\mathbb{R}^{n^2}$?

Exercise 9.6.12 Recall the definition of a line in a vector space V not necessarily with an inner product. $d \in V$ is a nonzero vector and $p \in V$ any point. Then the line $\ell(p, d)$ *through* p having the *direction* d (or with direction d) is the set

$$\ell(p, d) := \{x \in V \mid x := p + td, \text{ for some } t \in \mathbb{R}\}.$$

Prove the following theorems completing them if necessary:

(1) $\ell(p, a)$ and $\ell(p, b)$ are equal if and only if their directions are *parallel*, that is, if and only if there exists a nonzero real α such that $\alpha a = b$.

(2) $\ell(p, a) = \ell(q, a)$ if and only if ... .

(3) **Definition:** Two lines are said to be *parallel* if and only if their directions are parallel. Given a line ℓ and a point $q \notin \ell$, then there exists a unique line ℓ' such that $q \in \ell'$ and ℓ is parallel to ℓ'. (Euclid's parallel axiom proved!)

(4) Two distinct points p and q of V determine a line ℓ where

$$\ell = \ell(p, ?) = \ell(q, ?).$$

(5) $\ell(p, a)$ and $\ell(q, b)$ intersect if and only if $p - q$ lies in the span of ... .

Note that none of the above theorems needed the notion of an inner product. However we have

(6) In $\mathbb{R}^2$, a line can be described as $\{v \in \mathbb{R}^2 \mid \langle v - p, N \rangle = 0\}$ for a point p on the line. N is said to be *normal* to the line.

(7) Let $x(t) := p + td$ be a line ℓ in $\mathbb{R}^n$ and $q \notin \ell$. Show that

$$f(t) := ||q - x(t)||^2$$

takes a minimum value exactly at one point $t = t_0$. Prove that $q - x(t_0)$ is perpendicular to d (draw pictures).

Can you generalize this to the case of Exercise 9.6.10?

Exercise 9.6.13 Find the parametric equations of lines through the following pairs of points and find the midpoint of the segment between the pairs:

(1) $(-5, -6, 8)$ and (1,3,7).

(2) $(2, 4, 6)$ and (1,2,3).

(3) $(1, 3, 10)$ and $(-3, 6, -2)$.

(4) $(10, 3, 1)$ and $(6, -2, -3)$.

We recall the definition of a plane in $\mathbb{R}^3$. Given a point P and a *nonzero* vector N, the plane $\Pi(P; N)$ is the set of all lines through the point P which are perpendicular to N. Thus,

$$\Pi(P, N) := \left\{X \in \mathbb{R}^3 \mid (X - P) \cdot N = 0\right\} = \left\{X \in \mathbb{R}^3 \mid X \cdot N = P \cdot N\right\}$$

(Note the similarity between this and (6) of Exercise 9.6.12). Since P and N are given, $P \cdot N$ is a constant, say, d. If $X = (x, y, z)$ and $N = (a, b, c)$ then $X \cdot N = ax + by + cz$ so that we get the equation of the plane $ax + by + cz = d$.

Exercise 9.6.14 For each of the following equations find the normal vector to the corresponding plane and find *any* point on the plane:

(1) $x + y + z = 1$.

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

(3) $(x - 2) + 3(y - 5) - 4(z + 1) = 0$.

Exercise 9.6.15 Find the equation of a plane through the three given points:

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

(2) $(1, 0, 0), (-1, 0, 0), (0, 1, 1)$.

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

Exercise 9.6.16 Show that the x-axis in $\mathbb{R}^3$ is a line.

Exercise 9.6.17 Find the point of intersection of the lines

$$(1, -5, 2) + t(-3, 4, 0) \text{ and } (3, -13, 1) + t(4, 0, 1).$$

Exercise 9.6.18 Prove that the line $(1, 3, -1) + t(0, 3, 5)$ lies on the plane $2x - 5y + 3z = -16$.

Exercise 9.6.19 Let p_1 and p_2 lie on the plane $\{p \in \mathbb{R}^3 \mid \langle p, N\rangle = d\}$. Prove that any point of the line joining p_1 and p_2 lies in the plane.

Exercise 9.6.20 Find all points of intersection of the given line and the plane:

(1) $t(1, -3, 6)$; $x + 3y + z = 2$.

(2) $(1, -3, 6) + t(1, 0, 0)$; $z = 6$.

(3) $(1, -3, 6) + t(1, 0, 0)$; $z = 0$.

Exercise 9.6.21 Prove the following:

(1) If $B \cdot N \neq 0$, the line $A + tB$ intersects the plane $P \cdot N = P_0 \cdot N$ exactly at one point.

(2) If $B \cdot N = 0$ and A is in the plane $P \cdot N = P_0 \cdot N$, then the entire line $A + tB$ lies in the plane $P \cdot N = P_0 \cdot N$.

(3) If $B \cdot N = 0$ and A is not in the plane $P \cdot N = P_0 \cdot N$, then the line $A + tB$ does not intersect the plane $P \cdot N = P_0 \cdot N$ at any point.

Exercise 9.6.22 (1) Find the line through the given point (x_1, y_1, z_1) (say P_1) and normal to the plane $\Pi := \{(x, y, z) \mid ax + by + cz = d\}$.

(2) Find the point P_0 in which the line and the plane in (1) intersect.

(3) For any point P on the plane and P_0 as in (2), show that $P_1 - P_0 \perp P - P_0$.

(4) The distance between P_1 to the given plane is given by

$$d(P_1, \Pi) = \frac{|ax_1 + by_1 + cz_1 - d|}{\sqrt{a^2 + b^2 + c^2}}.$$

Compare this with (7) of Exercise 9.6.12. Can you generalize this to a hyperplane $\Pi := x + W$ where W is an $(n-1)$-dimensional vector subspace of an n-dimensional inner product space?

Exercise 9.6.23 (1) Let P_0 lie on the sphere $\{x \mid \|x\| = r\}$. Prove that the line $P_0 + tB$ intersects the sphere in two distinct points unless $B \cdot P_0 = 0$.

(2) The plane $P \cdot P_0 = P_0 \cdot P_0$ intersects the sphere only at P_0. *Hint:* $P = P_0 + (P - P_0)$.

(3) Let $P \cdot N = P_0 \cdot N$ be a plane through P_0. Prove that there is a point P_1 which lies on the sphere, on this plane and on the line $-P_0 + tN$. Show also that $P_1 = P_0$ if and only if N is parallel to P_0.

Exercise 9.6.24 Let $A = (1, 8, 2)$ and $B = (-3, 1, 1)$. Show that (x, y, z) lies on the line $A + tB$ if and only if $x = 7 - 3z$ and $6y + z$. What is the geometric interpretation of this result?

Exercise 9.6.25 Prove that every line is the intersection of two planes.

Exercise 9.6.26 Show that all three medians of a triangle meet at a point which divides the median"in the ratio 1:2".

Exercise 9.6.27 Let $\mathbf{E} = \mathbb{R}^2$ be with the dot product. Let

$$A := \begin{pmatrix} a & 0 \\ 0 & c \end{pmatrix}$$

Then the locus of the points $\{\langle Ax, x\rangle = 1\}$ can be considered as a conic section. It is an ellipse if $a > 0$ and $c > 0$ and it is a hyperbola if $a > 0$ and $c < 0$, for example. In the case of an ellipse, the eigenvalues of A are obtained by the minor and major axes of the ellipse. This may help you understand the proof of Theorem 7.3.8.

9.7 Miscellaneous Problems

Exercise 9.7.1 Answer true or false:

(1) If the vector 0 is among $\{v_i\}_{i=1}^k$, then $\{v_i\}_{i=1}^k$ are linearly dependent.

(2) If $v_1, v_2, v_3 \in \mathbb{R}^4$ are linearly dependent, then some v_i is a scalar multiple of some v_j.

(3) If $\{v_1, \dots, v_n\} \subseteq V$ is linearly dependent and i is given, $1 \le i \le n$, then v_i is a linear combination of other v_j's, $j \neq i$.

(4) The set of vectors $\{x \in \mathbb{R}^2 \mid x^2 - y^2 = 0\}$ is a vector subspace.

(5) If S_i, $i = 1, 2$ are subsets of V, and $L(S_1) = L(S_2)$, then $S_1 \cap S_2 \neq \emptyset$.

(6) The set of solutions of $a_1X_1 + \cdots + a_kX_k = 5$ in $\mathbb{R}^n$ is a vector subspace for any $a_1, \ldots, a_k \in \mathbb{R}$.

Exercise 9.7.2 What are the vector subspaces of $\mathbb{R}$?

Exercise 9.7.3 If W_1 and W_2 are vector subspaces of V, are $W_1 \cap W_2$ and $W_1 \cup W_2$ vector subspaces of V?

Exercise 9.7.4 Let $W \subseteq V$ and $x \in V \backslash W$. Can you find a vector subspace W_1 such that $W_1 \supseteq W$ and $x \in W$?

Exercise 9.7.5 Consider $\mathcal{P}_n$. Let θ be a real number. Show that

$$W := \{f \in \mathcal{P}_n \mid f(\theta) = 0\}$$

is a vector subspace.

Exercise 9.7.6 Show that $W := \left\{f \in C[0,1] \mid f'(\frac{1}{2}) \text{ exists}\right\}$ is a vector space.

Exercise 9.7.7 Let W_1 and W_2 be vector subspaces of V. Show that $W_1 + W_2$ is a vector subspace of V.

Exercise 9.7.8 Find the vector subspaces of $\mathbb{R}^2$.

Exercise 9.7.9 Let $S = \left\{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\right\}$. What is the subspace $L(S)$ spanned by S?

Exercise 9.7.10 Prove that if $T : V \to W$ is a linear map and

$$\dim \ker T = \dim V,$$

then $T = 0$.

Exercise 9.7.11 If $T, S : V \to V$, then show that $\ker S \subseteq \ker TS$.

Exercise 9.7.12 If $T : V \to W$ is a linear map with $\ker T \neq \{0\}$, then there are vectors v_1 and v_2 in V such that $Tv_1 = Tv_2$.

Exercise 9.7.13 Show that if $T \in L(V,W)$ with $\dim V > \dim W$, then there is a nonzero vector $v \in V$ with $Tv = 0$.

Exercise 9.7.14 Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear map with

$$Te_1 = e_2,\ Te_2 = e_3 \text{ and } Te_3 = 0.$$

Then $T \neq 0$, $T^2 \neq 0$ but $T^3 = 0$.

Exercise 9.7.15 If V is an inner product space, and W is a vector subspace of V, then $(W^{\perp})^{\perp} = W$.

Exercise 9.7.16 Let V be an n-dimensional inner product space. If $v \in V$ is perpendicular to n linearly independent vectors, then $v = 0$.

Exercise 9.7.17 Let v_i, $1 \le i \le n$ be vectors in an inner product space such that $\langle v_i, v_j \rangle = 0$ if $i \neq j$. Show that $\{v_i \mid 1 \le i \le n\}$ is linearly independent.

Exercise 9.7.18 The matrix representation of the identity transformation of a vector space is always the identity matrix.

Exercise 9.7.19 Show that an $n \times n$ matrix A commutes with every diagonal matrix if and only if A is diagonal.

Exercise 9.7.20 Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by

$$T(x, y, z) = (x + y, y + z, z + x).$$

Find a similar formula for T^{-1}.

Exercise 9.7.21 If $T : V \to V$ is linear and T^{-1} exists, then T^{-1} is linear.

Exercise 9.7.22 Let a linear map $T : V \to V$ be invertible. If $v_1, \dots, v_r$ are linearly independent so are $\{Tv_1, \dots, Tv_r\}$.

Exercise 9.7.23 Which of the following are vector subspaces?

(1) $\{(x, y, z) \mid x = 2\} \subseteq \mathbb{R}^3$ (2) $\{(x, y, z) \mid x = z = 0\} \subseteq \mathbb{R}^3$
(3) $\{(x, y, z) \mid z > 0\} \subseteq \mathbb{R}^3$ (4) $\{(x, y, z) \mid z = x + y\} \subseteq \mathbb{R}^3$.

Exercise 9.7.24 What is the geometric description of the subspace of $\mathbb{R}^3$ spanned by $\{(1, 0, 1), (1, -1, 0)\}$?

Exercise 9.7.25 Find the subspace spanned by $\{1, x - \alpha, (x - \alpha)^2\}$ in $\mathcal{P}_3$.

Exercise 9.7.26 Which of the following sets span $\mathbb{R}^3$?

(1) $\{(1, -1, 2), (0, 0, 1)\}$.

(2) $\{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 2)\}$.

(3) $\{(1, 0, 0), (0, 1, 1), (0, 1, -1)\}$.

Exercise 9.7.27 Let $\{x, y, z\}$ be a linearly independent subset of V. Let $u = x$, $v = x + y$ and $w = x + y + z$. Prove that $\{u, v, w\}$ is linearly independent.

Exercise 9.7.28 Let $S_1 \subseteq S_2$. Then show that

(1) If S_1 is linearly independent so is S_2.

(2) If S_2 is linearly independent so is S_1.

Exercise 9.7.29 $\left\{ \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\}$ is a basis of $\mathbb{R}^2$ if and only if

$$\det \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} \neq 0.$$

Exercise 9.7.30 No set of $n-1$ vectors can span an n-dimensional vector space.

Exercise 9.7.31 Which of the following sets are bases of $\mathbb{R}^2$?

(1) $\{(1,2),(1,-1)\}$ (2) $\{(1,0),(0,1),(0,0)\}$
(3) $\{(1,1),(1,2),(1,0)\}$ (4) $\{(1,-1)\}$.

Exercise 9.7.32 Find a basis for the following subspaces of $\mathbb{R}^3$.

(1) $\{(x,y,z) \mid z = x+y\}$.

(2) $\{(x,y,z) \mid x = y\}$.

(3) $\{(x,y,z) \mid x = 0\}$.

(4) $\{(x,y,z) \mid ax+by+cz = 0, a \neq 0\}$.

Exercise 9.7.33 Find an orthonormal basis of $\mathbb{R}^3$ containing the vectors

$$\frac{(1,0,1)}{\sqrt{2}} \text{ and } \frac{(1,-1,0)}{\sqrt{2}}.$$

Exercise 9.7.34 Apply Gram-Schmidt process to obtain an orthonormal basis from

$$\{(1,0,1),(1,-1,0),(1,1,1)\}.$$

Exercise 9.7.35 Which of the following are linear maps?

(1) $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x+y \\ x \\ x-z \end{pmatrix}$ from $\mathbb{R}^3 \to \mathbb{R}^3$.

(2) $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x^2 \\ x+y \end{pmatrix}$ from $\mathbb{R}^2 \to \mathbb{R}^2$.

(3) $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x+z \\ y+z \end{pmatrix}$ from $\mathbb{R}^3 \to \mathbb{R}^2$.

Exercise 9.7.36 Let $T : V \to W$ be a linear map. If $\{v_1, \ldots, v_k\}$ spans V, then $\{Tv_i \mid 1 \leq i \leq k\}$ spans Im T.

Exercise 9.7.37 If $T : V \to W$ is linear and $\{v_1, \ldots, v_r\}$ is such that $\{Tv_1, \ldots, Tv_r\}$ is linearly independent, so is $\{v_1, \ldots, v_r\}$.

Exercise 9.7.38 Let $T : V = \mathbb{R}^2 \to \mathbb{R}^2 = W$ be given by

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

Let $\{e_1, e_2\}$ be the standard basis of $\mathbb{R}^2$ and $\{v_1 = (1,1), v_2 = (1,-1)\}$ be another basis of $\mathbb{R}^2$. Compute the matrix representation of T with respect to:

(1) The natural basis of $\mathbb{R}^2$.

(2) The standard basis of V and $\{v_1, v_2\}$ of W.

(3) The basis $\{v_1, v_2\}$ of V and $\{e_1, e_2\}$ of W.

(4) The basis $\{v_1, v_2\}$ of V and W.

Index