CHAPTER III

Vector Spaces

As usual, a collection of objects will be called a **set**. A member of the collection is also called an **element** of the set. It is useful in practice to use short symbols to denote certain sets. For instance we denote by $\mathbf{R}$ the set of all numbers. To say that "x is a number" or that "x is an element of $\mathbf{R}$" amounts to the same thing. The set of n-tuples of numbers will be denoted by $\mathbf{R}^n$. Thus "X is an element of $\mathbf{R}^n$" and "X is an n-tuple" mean the same thing. Instead of saying that u is an element of a set S, we shall also frequently say that u **lies in** S and we write $u \in S$. If S and S' are two sets, and if every element of S' is an element of S, then we say that S' is a **subset** of S. Thus the set of rational numbers is a subset of the set of (real) numbers. To say that S is a subset of S' is to say that S is part of S'. To denote the fact that S is a subset of S', we write $S \subset S'$.

If S_1, S_2 are sets, then the **intersection** of S_1 and S_2, denoted by $S_1 \cap S_2$, is the set of elements which lie in both S_1 and S_2. The **union** of S_1 and S_2, denoted by $S_1 \cup S_2$, is the set of elements which lie in S_1 or S_2.

III, §1. Definitions

In mathematics, we meet several types of objects which can be added and multiplied by numbers. Among these are vectors (of the same dimension) and functions. It is now convenient to define in general a notion which includes these as a special case.

A **vector space** V is a set of objects which can be added and multiplied by numbers, in such a way that the sum of two elements of V is again an element of V, the product of an element of V by a number is an element of V, and the following properties are satisfied:

VS 1. *Given the elements u, v, w of V, we have*

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

VS 2. *There is an element of V, denoted by O, such that*

$$O + u = u + O = u$$

for all elements u of V.

VS 3. *Given an element u of V, the element $(-1)u$ is such that*

$$u + (-1)u = O.$$

VS 4. *For all elements u, v of V, we have*

$$u + v = v + u.$$

VS 5. *If c is a number, then $c(u + v) = cu + cv$.*

VS 6. *If a, b are two numbers, then $(a + b)v = av + bv$.*

VS 7. *If a, b are two numbers, then $(ab)v = a(bv)$.*

VS 8. *For all elements u of V, we have $1 \cdot u = u$ (1 here is the number one).*

We have used all these rules when dealing with vectors, or with functions but we wish to be more systematic from now on, and hence have made a list of them. Further properties which can be easily deduced from these are given in the exercises and will be assumed from now on.

The algebraic properties of elements of an arbitrary vector space are very similar to those of elements of $\mathbf{R}^2$, $\mathbf{R}^3$, or $\mathbf{R}^n$. Consequently it is customary to call elements of an arbitrary vector space also **vectors**.

If u, v are vectors (i.e. elements of the arbitrary vector space V), then the sum

$$u + (-1)v$$

is usually written $u - v$. We also write $-v$ instead of $(-1)v$.

Example 1. Fix two positive integers m, n. Let V be the set of all $m \times n$ matrices. We also denote V by $\mathrm{Mat}(m \times n)$. Then V is a vector space. It is easy to verify that all properties **VS 1** through **VS 8** are satisfied by our rules for addition of matrices and multiplication of matrices by numbers. The main thing to observe here is that addition of matrices is defined in terms of the components, and for the addition of components, the conditions analogous to **VS 1** through **VS 4** are satisfied. They are standard properties of numbers. Similarly, **VS 5** through **VS 8** are true for multiplication of matrices by numbers, because the corresponding properties for the multiplication of numbers are true.

Example 2. Let V be the set of all functions defined for all numbers. If f, g are two functions, then we know how to form their sum $f + g$. It is the function whose value at a number t is $f(t) + g(t)$. We also know how to multiply f by a number c. It is the function cf whose values at a number t is $cf(t)$. In dealing with functions, we have used properties **VS 1** through **VS 8** many times. We now realize that the set of functions is a vector space.

The function f such that $f(t) = 0$ **for all** t is the **zero function**. We emphasize the condition for all t. If a function has some of its values equal to zero, but other values not equal to 0, then it is **not** the zero function.

In practice, a number of elementary properties concerning addition of elements in a vector space are obvious because of the concrete way the vector space is given in terms of numbers, for instance as in the previous two examples. We shall now see briefly how to prove such properties just from the axioms.

It is possible to add several elements of a vector space. Suppose we wish to add four elements, say u, v, w, z. We first add any two of them, then a third, and finally a fourth. Using the rules **VS 1** and **VS 4**, we see that it does not matter in which order we perform the additions. This is exactly the same situation as we had with vectors. For example, we have

$$\begin{aligned}((u + v) + w) + z &= (u + (v + w)) + z \\ &= ((v + w) + u) + z \\ &= (v + w) + (u + z), \quad \text{etc.}\end{aligned}$$

Thus it is customary to leave out the parentheses, and write simply

$$u + v + w + z.$$

The same remark applies to the sum of any number n of elements of V.

We shall use 0 to denote the number zero, and O to denote the element of any vector space V satisfying property **VS 2**. We also call it zero, but there is never any possibility of confusion. We observe that this zero element O is uniquely determined by condition **VS 2**. Indeed, if

$$v + w = v$$

then adding $-v$ to both sides yields

$$-v + v + w = -v + v = O,$$

and the left-hand side is just $O + w = w$, so $w = O$.

Observe that for any element v in V we have

$$0v = O.$$

Proof.

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

Similarly, if c is a number, then

$$cO = O.$$

Proof. We have $cO = c(O + O) = cO + cO$. Add $-cO$ to both sides to get $cO = O$.

Subspaces

Let V be a vector space, and let W be a subset of V. Assume that W satisfies the following conditions.

(i) If v, w are elements of W, their sum $v + w$ is also an element of W.

(ii) If v is an element of W and c a number, then cv is an element of W.

(iii) The element O of V is also an element of W.

Then W itself is a vector space. Indeed, properties **VS 1** through **VS 8**, being satisfied for all elements of V, are satisfied also for the elements of W. We shall call W a **subspace** of V.

Example 3. Let $V = \mathbf{R}^n$ and let W be the set of vectors in V whose last coordinate is equal to 0. Then W is a subspace of V, which we could identify with $\mathbf{R}^{n-1}$.

Example 4. Let A be a vector in $\mathbf{R}^3$. Let W be the set of all elements B in $\mathbf{R}^3$ such that $B \cdot A = 0$, i.e. such that B is perpendicular to A. Then W is a subspace of $\mathbf{R}^3$. To see this, note that $O \cdot A = 0$, so that O is in W. Next, suppose that B, C are perpendicular to A. Then

$$(B + C) \cdot A = B \cdot A + C \cdot A = 0,$$

so that $B + C$ is also perpendicular to A. Finally, if x is a number, then

$$(xB) \cdot A = x(B \cdot A) = 0,$$

so that xB is perpendicular to A. This proves that W is a subspace of $\mathbf{R}^3$.

More generally, if A is a vector in $\mathbf{R}^n$, then the set of all elements B in $\mathbf{R}^n$ such that $B \cdot A = 0$ is a subspace of $\mathbf{R}^n$. The proof is the same as when $n = 3$.

Example 5. Let $\mathrm{Sym}(n \times n)$ be the set of all symmetric $n \times n$ matrices. Then $\mathrm{Sym}(n \times n)$ is a subspace of the space of all $n \times n$ matrices. Indeed, if A, B are symmetric and c is a number, then $A + B$ and cA are symmetric. Also the zero matrix is symmetric.

Example 6. If f, g are two continuous functions, then $f + g$ is continuous. If c is a number, then cf is continuous. The zero function is continuous. Hence the continuous functions form a subspace of the vector space of all functions.

If f, g are two differentiable functions, then their sum $f + g$ is differentiable. If c is a number, then cf is differentiable. The zero function is differentiable. Hence the differentiable functions form a subspace of the vector space of all functions. Furthermore, every differentiable function is continuous. Hence the differentiable functions form a subspace of the vector space of continuous functions.

Example 7. Let V be a vector space and let U, W be subspaces. We denote by $U \cap W$ the **intersection** of U and W, i.e. the set of elements which lie both in U and W. Then $U \cap W$ is a subspace. For instance, if U, W are two planes in 3-space passing through the origin, then in general, their intersection will be a straight line passing through the origin, as shown in Fig. 1.

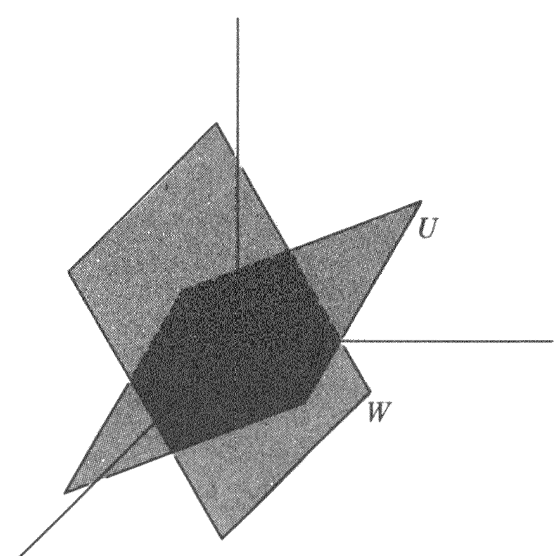


Figure 1

Example 8. Let U, W be subspaces of a vector space V. By

$$U + W$$

we denote the set of all elements $u + w$ with $u \in U$ and $w \in W$. Then we leave it to the reader to verify that $U + W$ is a subspace of V, said to be **generated by** U and W, and called the **sum** of U and W.

Exercises III, §1

1. Let $A_1, \dots, A_r$ be vectors in $\mathbf{R}^n$. Let W be the set of vectors B in $\mathbf{R}^n$ such that $B \cdot A_i = 0$ for every $i = 1, \dots, r$. Show that W is a subspace of $\mathbf{R}^n$.

2. Show that the following sets of elements in $\mathbf{R}^2$ form subspaces.
   (a) The set of all (x, y) such that $x = y$.
   (b) The set of all (x, y) such that $x - y = 0$.
   (c) The set of all (x, y) such that $x + 4y = 0$.

3. Show that the following sets of elements in $\mathbf{R}^3$ form subspaces.
   (a) The set of all (x, y, z) such that $x + y + z = 0$.
   (b) The set of all (x, y, z) such that $x = y$ and $2y = z$.
   (c) The set of all (x, y, z) such that $x + y = 3z$.

4. If U, W are subspaces of a vector space V, show that $U \cap W$ and $U + W$ are subspaces.

5. Let V be a subspace of $\mathbf{R}^n$. Let W be the set of elements of $\mathbf{R}^n$ which are perpendicular to every element of V. Show that W is a subspace of $\mathbf{R}^n$. This subspace W is often denoted by $V^\perp$, and is called V **perp**, or also the **orthogonal complement of** V.

III, §2. Linear Combinations

Let V be a vector space, and let $v_1, \dots, v_n$ be elements of V. We shall say that $v_1, \dots, v_n$ **generate** V if given an element $v \in V$ there exist numbers $x_1, \dots, x_n$ such that

$$v = x_1 v_1 + \cdots + x_n v_n.$$

Example 1. Let $E_1, \dots, E_n$ be the standard unit vectors in $\mathbf{R}^n$, so E_i has component 1 in the i-th place, and component 0 in all other places.

Then $E_1,\dots,E_n$ generate $\mathbf{R}^n$. Proof: given $X = (x_1,\dots,x_n) \in \mathbf{R}^n$. Then

$$X = \sum_{i=1}^{n} x_i E_i,$$

so there exist numbers satisfying the condition of the definition.

Let V be an arbitrary vector space, and let $v_1,\dots,v_n$ be elements of V. Let $x_1,\dots,x_n$ be numbers. An expression of type

$$x_1 v_1 + \cdots + x_n v_n$$

is called a **linear combination** of $v_1,\dots,v_n$. The numbers $x_1,\dots,x_n$ are then called the **coefficients** of the linear combination.

The set of all linear combinations of $v_1,\dots,v_n$ is a subspace of V.

Proof. Let W be the set of all such linear combinations. Let $y_1,\dots,y_n$ be numbers. Then

$$(x_1 v_1 + \cdots + x_n v_n) + (y_1 v_1 + \cdots + y_n v_n) = (x_1 + y_1)v_1 + \cdots + (x_n + y_n)v_n.$$

Thus the sum of two elements of W is again an element of W, i.e. a linear combination of $v_1,\dots,v_n$. Furthermore, if c is a number, then

$$c(x_1 v_1 + \cdots + x_n v_n) = c x_1 v_1 + \cdots + c x_n v_n$$

is a linear combination of $v_1,\dots,v_n$, and hence is an element of W. Finally,

$$O = 0v_1 + \cdots + 0v_n$$

is an element of W. This proves that W is a subspace of V.

The subspace W consisting of all linear combinations of $v_1,\dots,v_n$ is called the subspace **generated** by $v_1,\dots,v_n$.

Example 2. Let v_1 be a non-zero element of a vector space V, and let w be any element of V. The set of elements

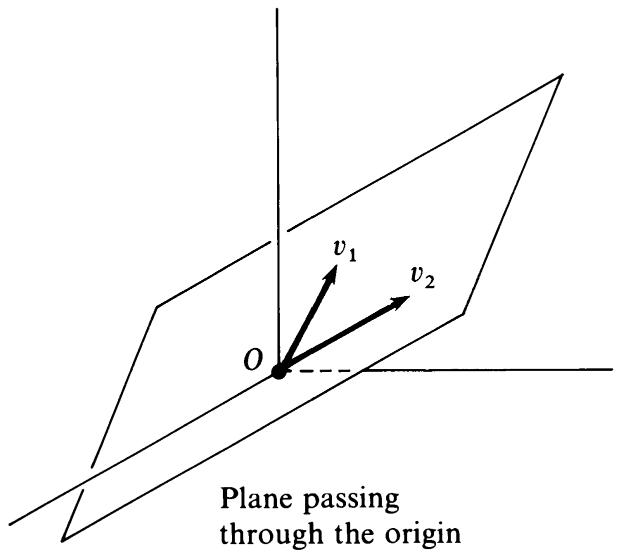
$$w + t v_1 \qquad \text{with} \qquad t \in \mathbf{R}$$

is called the **line passing through** w **in the direction of** v_1. We have already met such lines in Chapter I, §5. If $w = O$, then the line consisting of all scalar multiples tv_1 with $t \in \mathbf{R}$ is a subspace, generated by v_1.

Let v_1, v_2 be elements of a vector space V, and assume that neither is a scalar multiple of the other. The subspace generated by v_1, v_2 is called the **plane** generated by v_1, v_2. It consists of all linear combinations

$$t_1v_1 + t_2v_2 \qquad \text{with} \qquad t_1,\ t_2 \text{ arbitrary numbers.}$$

This plane passes through the origin, as one sees by putting $t_1 = t_2 = 0$.



Plane passing through the origin

Figure 2

We obtain the most general notion of a plane by the following operation. Let S be an arbitrary subset of V. Let P be an element of V. If we add P to all elements of S, then we obtain what is called the **translation** of S by P. It consists of all elements $P + v$ with v in S.

Example 3. Let v_1, v_2 be elements of a vector space V such that neither is a scalar multiple of the other. Let P be an element of V. We define the **plane passing through** P, **parallel to** v_1, v_2 to be the set of all elements

$$P + t_1v_1 + t_2v_2$$

where t_1, t_2 are arbitrary numbers. This notion of plane is the analogue, with two elements v_1, v_2, of the notion of parametrized line considered in Chapter I.

Warning. Usually such a plane does not pass through the origin, as shown on Fig. 3. Thus such a plane is **not** a subspace of V. If we take $P = O$, however, then the plane is a subspace.

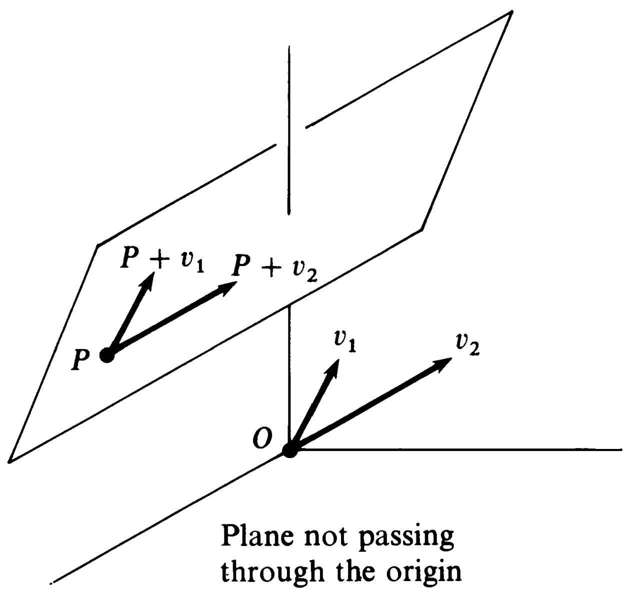


Figure 3

Sometimes it is interesting to restrict the coefficients of a linear combination. We give a number of examples below.

Example 4. Let V be a vector space and let v, u be elements of V. We define the **line segment** between v and $v + u$ to be the set of all points

$$v + tu, \qquad 0 \leqq t \leqq 1.$$

This line segment is illustrated in the following picture.

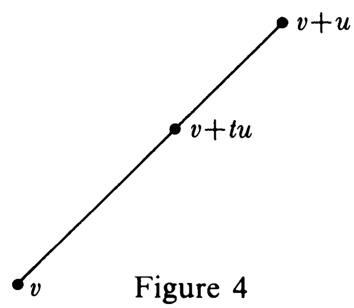


Figure 4

For instance, if $t = \frac{1}{2}$, then $v + \frac{1}{2}u$ is the point midway between v and $v + u$. Similarly, if $t = \frac{1}{3}$, then $v + \frac{1}{3}u$ is the point one third of the way between v and $v + u$ (Fig. 5).

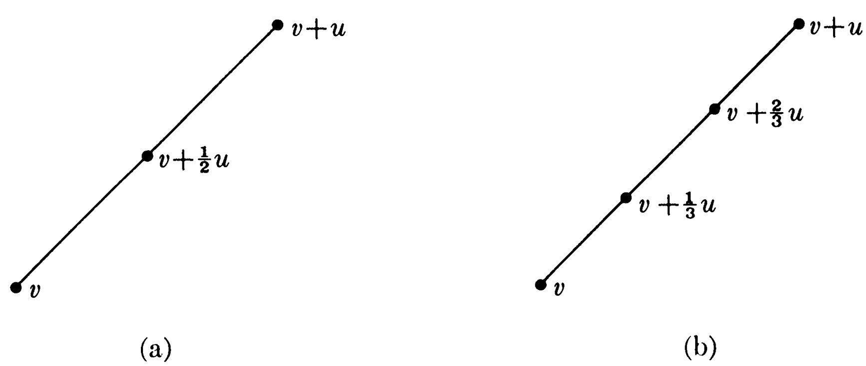


Figure 5

If v, w are elements of V, let $u = w - v$. Then the **line segment between** v and w is the set of all points $v + tu$, or

$$v + t(w - v), \qquad 0 \leqq t \leqq 1.$$

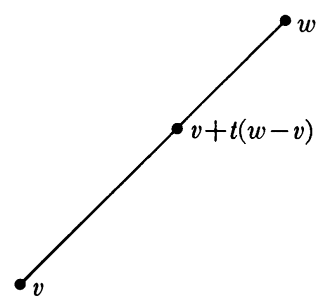


Figure 6

Observe that we can rewrite the expression for these points in the form

$$(1) \qquad (1 - t)v + tw, \qquad 0 \leqq t \leqq 1,$$

and letting $s = 1 - t$, $t = 1 - s$, we can also write it as

$$sv + (1 - s)w, \qquad 0 \leqq s \leqq 1.$$

Finally, we can write the points of our line segment in the form

$$(2) \qquad t_1 v + t_2 w \quad \text{with} \quad t_1,\ t_2 \geqq 0 \quad \text{and} \quad t_1 + t_2 = 1.$$

Indeed, letting $t = t_2$, we see that every point which can be written in the form (2) satisfies (1). Conversely, we let $t_1 = 1 - t$ and $t_2 = t$ and see that every point of the form (1) can be written in the form (2).

Example 5. Let v, w be elements of a vector space V. Assume that neither is a scalar multiple of the other. We define the **parallelogram spanned** by v, w to be the set of all points

$$t_1 v + t_2 w, \qquad 0 \leqq t_i \leqq 1 \quad \text{for} \quad i = 1, 2.$$

This definition is clearly justified since $t_1 v$ is a point of the segment between O and v (Fig. 7), and $t_2 w$ is a point of the segment between O

and w. For all values of t_1, t_2 ranging independently between 0 and 1, we see geometrically that $t_1 v + t_2 w$ describes all points of the parallelogram.

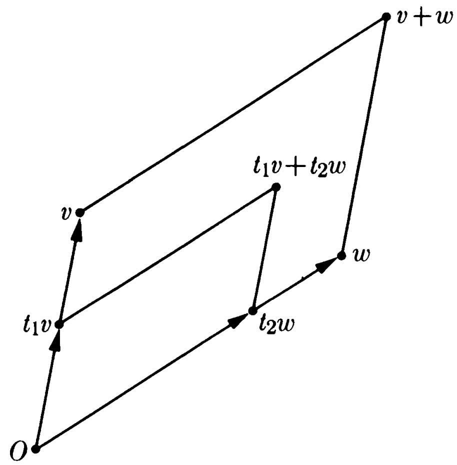


Figure 7

We obtain the most general parallelogram (Fig. 8) by taking the translation of the parallelogram just described. Thus if u is an element of V, the translation by u of the parallelogram spanned by v and w consists of all points

$$u + t_1 v + t_2 w, \qquad 0 \leqq t_i \leqq 1 \quad \text{for} \quad i = 1, 2.$$

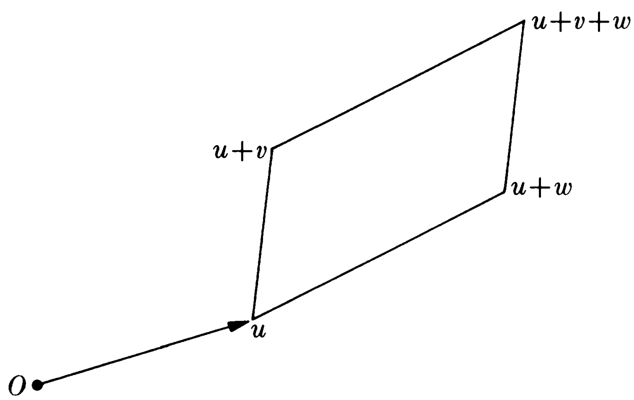


Figure 8

Similarly, in higher dimensions, let v_1, v_2, v_3 be elements of a vector space V. We define the **box spanned by these elements** to be the set of linear combinations

$$t_1 v_1 + t_2 v_2 + t_3 v_3 \qquad \text{with} \qquad 0 \leqq t_i \leqq 1.$$

We draw the picture when v_1, v_2, v_3 are in general position:

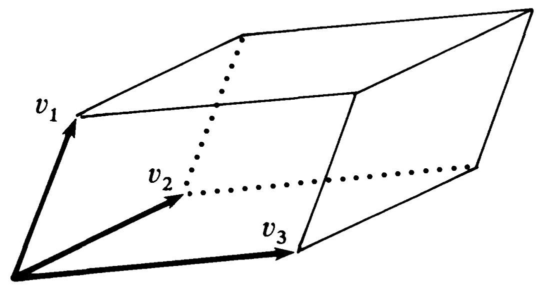


Figure 9

There may be degenerate cases, which will lead us into the notion of linear dependence a little later.

Exercises III, §2

1. Let $A_1, \ldots, A_r$ be generators of a subspace V of $\mathbf{R}^n$. Let W be the set of all elements of $\mathbf{R}^n$ which are perpendicular to $A_1, \ldots, A_r$. Show that the vectors of W are perpendicular to every element of V.
2. Draw the parallelogram spanned by the vectors $(1, 2)$ and $(-1, 1)$ in $\mathbf{R}^2$.
3. Draw the parallelogram spanned by the vectors $(2, -1)$ and $(1, 3)$ in $\mathbf{R}^2$.

III, §3. Convex Sets

Let S be a subset of a vector space V. We shall say that S is **convex** if given points P, Q in S then the line segment between P and Q is contained in S. In Fig. 10, the set on the left is convex. The set on the right is not convex since the line segment between P and Q is not entirely contained in S.

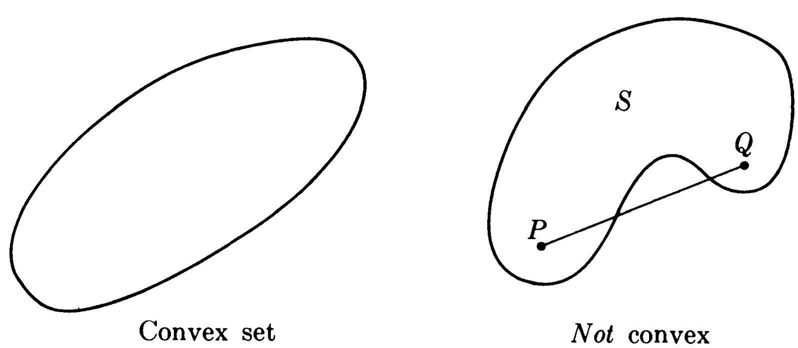


Figure 10

We recall that the line segment between P and Q consists of all points

$$(1 - t)P + tQ \qquad \text{with} \qquad 0 \leqq t \leqq 1.$$

This gives us a simple test to determine whether a set is convex or not.

Example 1. Let S be the parallelogram spanned by two vectors v_1, v_2, so S is the set of linear combinations

$$t_1 v_1 + t_2 v_2 \quad \text{with} \quad 0 \leqq t_i \leqq 1.$$

We wish to prove that S is convex. Let

$$P = t_1 v_1 + t_2 v_2 \quad \text{and} \quad Q = s_1 v_1 + s_2 v_2$$

be points in S. Then

$$\begin{aligned}(1-t)P + tQ &= (1-t)(t_1 v_1 + t_2 v_2) + t(s_1 v_1 + s_2 v_2)\\ &= (1-t)t_1 v_1 + (1-t)t_2 v_2 + ts_1 v_1 + ts_2 v_2\\ &= r_1 v_1 + r_2 v_2,\end{aligned}$$

where

$$r_1 = (1-t)t_1 + ts_1 \quad \text{and} \quad r_2 = (1-t)t_2 + ts_2.$$

But we have

$$0 \leqq (1-t)t_1 + ts_1 \leqq (1-t) + t = 1$$

and

$$0 \leqq (1-t)t_2 + ts_2 \leqq (1-t) + t = 1.$$

Hence

$$(1-t)P + tQ = r_1 v_1 + r_2 v_2 \quad \text{with} \quad 0 \leqq r_i \leqq 1.$$

This proves that $(1-t)P + tQ$ is in the parallelogram, which is therefore convex.

Example 2. Half planes. Consider a linear equation like

$$2x - 3y = 6.$$

This is the equation of a line as shown on Fig. 11.

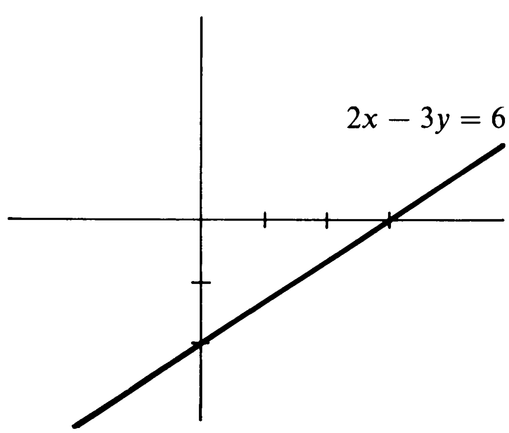


Figure 11

The inequalities

$$2x - 3y \leqq 6 \qquad \text{and} \qquad 2x - 3y \geqq 6$$

determine two half planes; one of them lies below the line and the other lies above the line, as shown on Fig. 12.

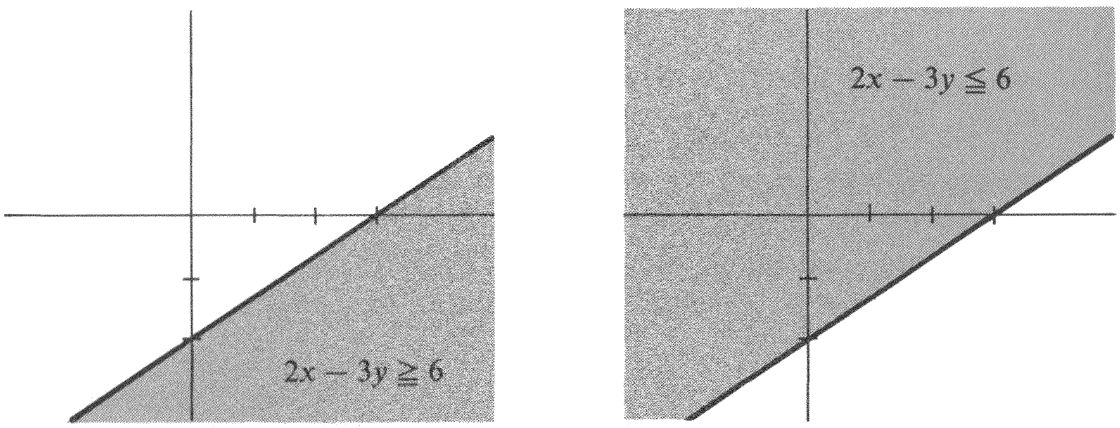


Figure 12

Let $A = (2, -3)$. We can, and should write the linear inequalities in the form

$$A \cdot X \geqq 6 \qquad \text{and} \qquad A \cdot X \leqq 6,$$

where $X = (x, y)$. Prove as Exercise 2 that each half plane is convex. This is clear intuitively from the picture, at least in $\mathbf{R}^2$, but your proof should be valid for the analogous situation in $\mathbf{R}^n$.

Theorem 3.1. *Let $P_1, \ldots, P_n$ be points of a vector space V. Let S be the set of all linear combinations*

$$t_1 P_1 + \cdots + t_n P_n$$

with $0 \leqq t_i$ and $t_1 + \cdots + t_n = 1$. Then S is convex.

Proof. Let

$$P = t_1 P_1 + \cdots + t_n P_n$$

and

$$Q = s_1 P_1 + \cdots + s_n P_n$$

with $0 \leqq t_i$, $0 \leqq s_i$, and

$$t_1 + \cdots + t_n = 1,$$

$$s_1 + \cdots + s_n = 1.$$

Let $0 \leqq t \leqq 1$. Then:

$$\begin{aligned}(1-t)P + tQ &= (1-t)t_1P_1 + \cdots + (1-t)t_nP_n \\ &\quad + ts_1P_1 + \cdots + ts_nP_n \\ &= [(1-t)t_1 + ts_1]P_1 + \cdots + [(1-t)t_n + ts_n]P_n.\end{aligned}$$

We have $0 \leqq (1-t)t_i + ts_i$ for all i, and

$$\begin{aligned}(1-t)t_1 &+ ts_1 + \cdots + (1-t)t_n + ts_n \\ &= (1-t)(t_1 + \cdots + t_n) + t(s_1 + \cdots + s_n) \\ &= (1-t) + t \\ &= 1.\end{aligned}$$

This proves our theorem.

In the next theorem, we shall prove that the set of all linear combinations

$$t_1P_1 + \cdots + t_nP_n \qquad \text{with} \qquad 0 \leqq t_i \quad \text{and} \quad t_1 + \cdots + t_n = 1$$

is the smallest convex set containing $P_1, \ldots, P_n$. For example, suppose that P_1, P_2, P_3 are three points in the plane not on a line. Then it is geometrically clear that the smallest convex set containing these three points is the triangle having these points as vertices.

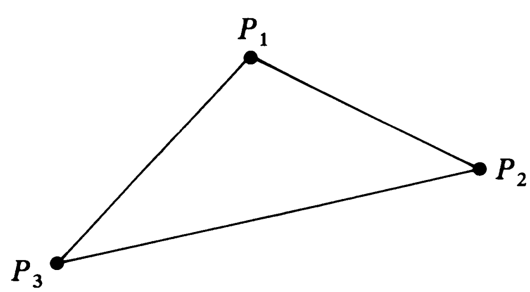


Figure 13

Thus it is natural to take as definition of a triangle the following property, valid in any vector space.

Let P_1, P_2, P_3 be three points in a vector space V, not lying on a line. Then the **triangle spanned** by these points is the set of all combinations

$$t_1P_1 + t_2P_2 + t_3P_3 \qquad \text{with} \quad 0 \leqq t_i \quad \text{and} \quad t_1 + t_2 + t_3 = 1.$$

When we deal with more than three points, then the set of linear combinations as in Theorem 3.1 looks as in the following figure.

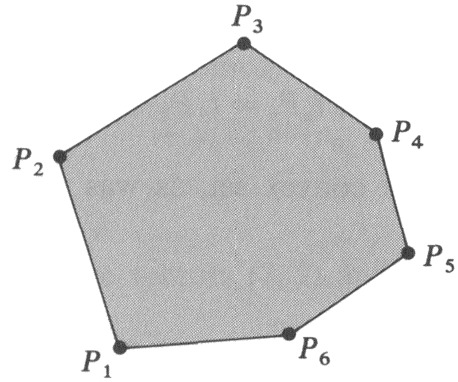


Figure 14

We shall call the convex set of Theorem 3.1 the convex set **spanned** by $P_1, \ldots, P_n$. Although we shall not need the next result, it shows that this convex set is the smallest convex set containing all the points $P_1, \ldots, P_n$. Omit the proof if you can't handle the argument by induction.

Theorem 3.2. *Let $P_1, \ldots, P_n$ be points of a vector space V. Any convex set which contains $P_1, \ldots, P_n$ also contains all linear combinations*

$$t_1 P_1 + \cdots + t_n P_n$$

with $0 \leqq t_i$ for all i and $t_1 + \cdots + t_n = 1$.

Proof. We prove this by induction. If $n = 1$, then $t_1 = 1$, and our assertion is obvious. Assume the theorem proved for some integer $n - 1 \geqq 1$. We shall prove it for n. Let $t_1, \ldots, t_n$ be numbers satisfying the conditions of the theorem. Let S' be a convex set containing $P_1, \ldots, P_n$. We must show that S' contains all linear combinations

$$t_1 P_1 + \cdots + t_n P_n.$$

If $t_n = 1$, then our assertion is trivial because $t_1 = \cdots = t_{n-1} = 0$. Suppose that $t_n \neq 1$. Then the linear combination $t_1 P_1 + \cdots + t_n P_n$ is equal to

$$(1 - t_n)\left(\frac{t_1}{1 - t_n} P_1 + \cdots + \frac{t_{n-1}}{1 - t_n} P_{n-1}\right) + t_n P_n.$$

Let

$$s_i = \frac{t_i}{1 - t_i} \qquad \text{for} \qquad i = 1, \ldots, n - 1.$$

Then $s_i \geqq 0$ and $s_1 + \cdots + s_{n-1} = 1$ so that by induction, we conclude that the point

$$Q = s_1 P_1 + \cdots + s_{n-1} P_{n-1}$$

lies in S'. But then

$$(1 - t_n)Q + t_nP_n = t_1P_1 + \cdots + t_nP_n$$

lies in S' by definition of a convex set, as was to be shown.

Exercises III, §3

1. Let S be the parallelogram consisting of all linear combinations $t_1v_1 + t_2v_2$ with $0 \leqq t_1 \leqq 1$ and $0 \leqq t_2 \leqq 1$. Prove that S is convex.
2. Let A be a non-zero vector in $\mathbf{R}^n$ and let c be a fixed number. Show that the set of all elements X in $\mathbf{R}^n$ such that $A \cdot X \geqq c$ is convex.
3. Let S be a convex set in a vector space. If c is a number, denote by cS the set of all elements cv with v in S. Show that cS is convex.
4. Let S_1 and S_2 be convex sets. Show that the intersection $S_1 \cap S_2$ is convex.
5. Let S be a convex set in a vector space V. Let w be an arbitrary element of V. Let $w + S$ be the set of all elements $w + v$ with v in S. Show that $w + S$ is convex.

III, §4. Linear Independence

Let V be a vector space, and let $v_1, \ldots, v_n$ be elements of V. We shall say that $v_1, \ldots, v_n$ are **linearly dependent** if there exist numbers $a_1, \ldots, a_n$ not all equal to 0 such that

$$a_1v_1 + \cdots + a_nv_n = O.$$

If there do not exist such numbers, then we say that $v_1, \ldots, v_n$ are **linearly independent.** In other words, vectors $v_1, \ldots, v_n$ are linearly independent if and only if the following condition is satisfied:

Let $a_1, \ldots, a_n$ be numbers such that

$$a_1v_1 + \cdots + a_nv_n = O;$$

then $a_i = 0$ for all $i = 1, \ldots, n$.

Example 1. Let $V = \mathbf{R}^n$ and consider the vectors

$$\begin{aligned} E_1 &= (1, 0, \ldots, 0) \\ &\vdots \\ E_n &= (0, 0, \ldots, 1). \end{aligned}$$

Then $E_1, \dots, E_n$ are linearly independent. Indeed, let $a_1, \dots, a_n$ be numbers such that $a_1 E_1 + \cdots + a_n E_n = O$. Since

$$a_1 E_1 + \cdots + a_n E_n = (a_1, \dots, a_n),$$

it follows that all $a_i = 0$.

Example 2. Show that the vectors $(1, 1)$ and $(-3, 2)$ are linearly independent.

Let a, b be two numbers such that

$$a(1, 1) + b(-3, 2) = O.$$

Writing this equation in terms of components, we find

$$a - 3b = 0, \qquad a + 2b = 0.$$

This is a system of two equations which we solve for a and b. Subtracting the second from the first, we get $-5b = 0$, whence $b = 0$. Substituting in either equation, we find $a = 0$. Hence, a, b are both 0, and our vectors are linearly independent.

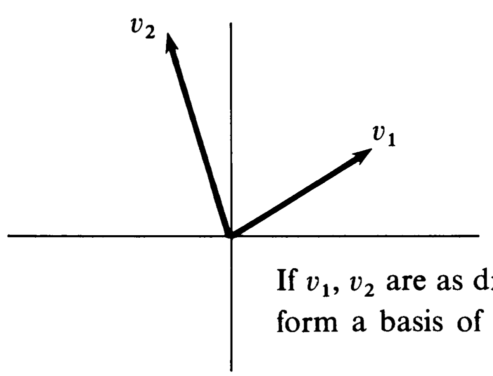
If elements $v_1, \dots, v_n$ of V generate V and in addition are linearly independent, then $\{v_1, \dots, v_n\}$ is called a **basis** of V. We shall also say that the elements $v_1, \dots, v_n$ **constitute** or **form** a basis of V.

Example 3. The vectors $E_1, \dots, E_n$ of Example 1 form a basis of $\mathbf{R}^n$. To prove this we have to prove that they are linearly independent, which was already done in Example 1; and that they generate $\mathbf{R}^n$. Given an element $A = (a_1, \dots, a_n)$ of $\mathbf{R}^n$ we can write A as a linear combination

$$A = a_1 E_1 + \cdots + a_n E_n,$$

so by definition, $E_1, \dots, E_n$ generate $\mathbf{R}^n$. Hence they form a basis.

However, there are many other bases. Let us look at $n = 2$. We shall find out that any two vectors which are not parallel form a basis of $\mathbf{R}^2$. Let us first consider an example.



If v_1, v_2 are as drawn, they form a basis of $\mathbf{R}^2$.

Figure 15

Example 4. Show that the vectors $(1, 1)$ and $(-1, 2)$ form a basis of $\mathbf{R}^2$.

We have to show that they are linearly independent and that they generate $\mathbf{R}^2$. To prove linear independence, suppose that a, b are numbers such that

$$a(1,\ 1) + b(-1,\ 2) = (0,\ 0)$$

Then

$$a - b = 0, \qquad a + 2b = 0.$$

Subtracting the first equation from the second yields $3b = 0$, so that $b = 0$. But then from the first equation, $a = 0$, thus proving that our vectors are linearly independent.

Next, we must show that $(1, 1)$ and $(-1, 2)$ generate $\mathbf{R}^2$. Let (s, t) be an arbitrary element of $\mathbf{R}^2$. We have to show that there exist numbers x, y such that

$$x(1, 1) + y(-1, 2) = (s,\ t).$$

In other words, we must solve the system of equations

$$\begin{aligned} x - y &= s, \\ x + 2y &= t. \end{aligned}$$

Again subtract the first equation from the second. We find

$$3y = t - s,$$

whence

$$y = \frac{t - s}{3},$$

and finally

$$x = y + s = \frac{t - s}{3} + s.$$

This proves that $(1, 1)$ and $(-1, 2)$ generate $\mathbf{R}^2$, and concludes the proof that they form a basis of $\mathbf{R}^2$.

The general story for $\mathbf{R}^2$ is expressed in the following theorem.

Theorem 4.1. *Let (a, b) and (c, d) be two vectors in $\mathbf{R}^2$.*

(i) *They are linearly dependent if and only if $ad - bc = 0$.*
(ii) *If they are linearly independent, then they form a basis of $\mathbf{R}^2$.*

Proof. First work it out as an exercise (see Exercise 4). If you can't do it, you will find the proof in the answer section. It parallels closely the procedure of Example 4.

Let V be a vector space, and let $\{v_1,\dots,v_n)$ be a basis of V. The elements of V can be represented by n-tuples relative to this basis, as follows. If an element v of V is written as a linear combination

$$v = x_1v_1 + \cdots + x_nv_n$$

of the basis elements, then we call $(x_1,\dots,x_n)$ the **coordinates** of v with respect to our basis, and we call x_i the i-th coordinate. The coordinates with respect to the usual basis $E_1,\dots,E_n$ of $\mathbf{R}^n$ are simply the coordinates as defined in Chapter I, §1.

The following theorem shows that there can only be one set of coordinates for a given vector.

Theorem 4.2. *Let V be a vector space. Let $v_1,\dots,v_n$ be linearly independent elements of V. Let $x_1,\dots,x_n$ and $y_1,\dots,y_n$ be numbers such that*

$$x_1v_1 + \cdots + x_nv_n = y_1v_1 + \cdots + y_nv_n.$$

Then we must have $x_i = y_i$ for all $i = 1,\dots,n$.

Proof. Subtract the right-hand side from the left-hand side. We get

$$x_1v_1 - y_1v_1 + \cdots + x_nv_n - y_nv_n = O.$$

We can write this relation also in the form

$$(x_1 - y_1)v_1 + \cdots + (x_n - y_n)v_n = O.$$

By definition, we must have $x_i - y_i = 0$ for all $i = 1,\dots,n$, thereby proving our assertion.

The theorem expresses the fact that when an element is written as a linear combination of $v_1,\dots,v_n$, then its coefficients $x_1,\dots,x_n$ are uniquely determined. This is true only when $v_1,\dots,v_n$ are linearly independent.

Example 5. Find the coordinates of $(1, 0)$ with respect to the two vectors $(1, 1)$ and $(-1, 2)$.

We must find numbers a, b such that

$$a(1, 1) + b(-1, 2) = (1, 0).$$

Writing this equation in terms of coordinates, we find

$$a - b = 1, \qquad a + 2b = 0.$$

Solving for a and b in the usual manner yields $b = -\frac{1}{3}$ and $a = \frac{2}{3}$. Hence the coordinates of $(1, 0)$ with respect to $(1, 1)$ and $(-1, 2)$ are $(\frac{2}{3}, -\frac{1}{3})$.

Example 6. The two functions e^t, e^{2t} are linearly independent. To prove this, suppose that there are numbers a, b such that

$$ae^t + be^{2t} = 0$$

(for all values of t). Differentiate this relation. We obtain

$$ae^t + 2be^{2t} = 0.$$

Subtract the first from the second relation. We obtain $be^{2t} = 0$, and hence $b = 0$. From the first relation, it follows that $ae^t = 0$, and hence $a = 0$. Hence e^t, e^{2t} are linearly independent.

Example 7. Let V be the vector space of all functions of a variable t. Let $f_1, \dots, f_n$ be n functions. To say that they are linearly dependent is to say that there exist n numbers $a_1, \dots, a_n$ not all equal to 0 such that

$$a_1 f_1(t) + \cdots + a_n f_n(t) = 0$$

for *all* values of t.

Warning. We emphasize that linear dependence for functions means that the above relation holds for all values of t. For instance, consider the relation

$$a \sin t + b \cos t = 0,$$

where a, b are two fixed numbers not both zero. There may be *some* values of t for which the above equation is satisfied. For instance, if $a \neq 0$ we then can solve

$$\frac{\sin t}{\cos t} = \frac{b}{a},$$

or in other words, $\tan t = b/a$ to get at least one solution. However, the above relation cannot hold for *all* values of t, and consequently $\sin t$, $\cos t$ are linearly independent, as functions.

Example 8. Let V be the vector space of functions generated by the two functions e^t, e^{2t}. Then the coordinates of the function

$$3e^t + 5e^{2t}$$

with respect to the basis $\{e^t, e^{2t}\}$ are $(3, 5)$.

When dealing with two vectors v, w there is another convenient way of expressing linear independence.

Theorem 4.3. *Let v, w be elements of a vector space V. They are linearly dependent if and only if one of them is a scalar multiple of the other, i.e. there is a number $c \neq 0$ such that we have $v = cw$ or $w = cv$.*

Proof. Left as an exercise, cf. Exercise 5.

In the light of this theorem, the condition imposed in various examples in the preceding section could be formulated in terms of two vectors being linearly independent.

Exercises III, §4

1. Show that the following vectors are linearly independent.
   (a) $(1, 1, 1)$ and $(0, 1, -2)$
   (b) $(1, 0)$ and $(1, 1)$
   (c) $(-1, 1, 0)$ and $(0, 1, 2)$
   (d) $(2, -1)$ and $(1, 0)$
   (e) $(\pi, 0)$ and $(0, 1)$
   (f) $(1, 2)$ and $(1, 3)$
   (g) $(1, 1, 0)$, $(1, 1, 1)$, and $(0, 1, -1)$
   (h) $(0, 1, 1)$, $(0, 2, 1)$, and $(1, 5, 3)$
2. Express the given vector X as a linear combination of the given vectors A, B, and find the coordinates of X with respect to A, B.
   (a) $X = (1, 0)$, $A = (1, 1)$, $B = (0, 1)$
   (b) $X = (2, 1)$, $A = (1, -1)$, $B = (1, 1)$
   (c) $X = (1, 1)$, $A = (2, 1)$, $B = (-1, 0)$
   (d) $X = (4, 3)$, $A = (2, 1)$, $B = (-1, 0)$
3. Find the coordinates of the vector X with respect to the vectors A, B, C.
   (a) $X = (1, 0, 0)$, $A = (1, 1, 1)$, $B = (-1, 1, 0)$, $C = (1, 0, -1)$
   (b) $X = (1, 1, 1)$, $A = (0, 1, -1)$, $B = (1, 1, 0)$, $C = (1, 0, 2)$
   (c) $X = (0, 0, 1)$, $A = (1, 1, 1)$, $B = (-1, 1, 0)$, $C = (1, 0, -1)$
4. Let (a, b) and (c, d) be two vectors in $\mathbf{R}^2$.
   (i) If $ad - bc \neq 0$, show that they are linearly independent.
   (ii) If they are linearly independent, show that $ad - bc \neq 0$.
   (iii) If $ad - bc \neq 0$ show that they form a basis of $\mathbf{R}^2$.
5. (a) Let v, w be elements of a vector space. If v, w are linearly dependent, show that there is a number c such that $w = cv$, or $v = cw$.
   (b) Conversely, let v, w be elements of a vector space, and assume that there exists a number c such that $w = cv$. Show that v, w are linearly dependent.

6. Let $A_1,\dots,A_r$ be vectors in $\mathbf{R}^n$, and assume that they are mutually perpendicular, in other words $A_i \perp A_j$ if $i \neq j$. Also assume that none of them is O. Prove that they are linearly independent.

7. Consider the vector space of all functions of a variable t. Show that the following pairs of functions are linearly independent.
(a) $1, t$ (b) t, t^2 (c) t, t^4 (d) e^t, t (e) te^t, e^{2t} (f) $\sin t, \cos t$
(g) $t, \sin t$ (h) $\sin t$, $\sin 2t$ (i) $\cos t$, $\cos 3t$

8. Consider the vector space of functions defined for $t > 0$. Show that the following pairs of functions are linearly independent.
(a) t, $1/t$ (b) e^t, $\log t$

9. What are the coordinates of the function $3\sin t + 5\cos t = f(t)$ with respect to the basis $\{\sin t, \cos t\}$?

10. Let D be the derivative d/dt. Let $f(t)$ be as in Exercise 9. What are the coordinates of the function $Df(t)$ with respect to the basis of Exercise 9?

In each of the following cases, exhibit a basis for the given space, and prove that it is a basis.

11. The space of 2×2 matrices.
12. The space of $m \times n$ matrices.
13. The space of $n \times n$ matrices all of whose components are 0 except possibly the diagonal components.
14. The **upper triangular matrices**, i.e. matrices of the following type:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

15. (a) The space of symmetric 2×2 matrices.
(b) The space of symmetric 3×3 matrices.
16. The space of symmetric $n \times n$ matrices.

III, §5. Dimension

We ask the question: Can we find three linearly independent elements in $\mathbf{R}^2$? For instance, are the elements

$$A = (1, 2), \qquad B = (-5, 7), \qquad C = (10, 4)$$

linearly independent? If you write down the linear equations expressing the relation

$$xA + yB + zC = O,$$

you will find that you can solve them for x, y, z not equal to 0. Namely, these equations are:

$$x - 5y + 10z = 0,$$

$$2x + 7y + 4z = 0.$$

This is a system of two homogeneous equations in three unknowns, and we know by Theorem 2.1 of Chapter II that we can find a non-trivial solution (x, y, z) not all equal to zero. Hence A, B, C are linearly dependent.

We shall see in a moment that this is a general phenomenon. In $\mathbf{R}^n$, we cannot find more than n linearly independent vectors. Furthermore, we shall see that any n linearly independent elements of $\mathbf{R}^n$ must generate $\mathbf{R}^n$, and hence form a basis. Finally, we shall also see that if one basis of a vector space has n elements, and another basis has m elements, then $m = n$. In short, two bases must have the same number of elements. This property will allow us to define the **dimension** of a vector space as the number of elements in any basis. We now develop these ideas systematically.

Theorem 5.1. *Let V be a vector space, and let $\{v_1, \ldots, v_m\}$ generate V. Let $w_1, \ldots, w_n$ be elements of V and assume that $n > m$. Then $w_1, \ldots, w_n$ are linearly dependent.*

Proof. Since $\{v_1, \ldots, v_m\}$ generate V, there exist numbers (a_{ij}) such that we can write

$$\begin{aligned} w_1 &= a_{11}v_1 + \cdots + a_{m1}v_m \\ &\ \,\vdots \\ w_n &= a_{1n}v_1 + \cdots + a_{mn}v_m. \end{aligned}$$

If $x_1, \ldots, x_n$ are numbers, then

$$\begin{aligned} &x_1w_1 + \cdots + x_nw_n \\ &\qquad = (x_1a_{11} + \cdots + x_na_{1n})v_1 + \cdots + (x_1a_{m1} + \cdots + x_na_{mn})v_m \end{aligned}$$

(just add up the coefficients of $v_1, \ldots, v_m$ vertically downward). According to Theorem 2.1 of Chapter II, the system of equations

$$\begin{aligned} x_1a_{11} + \cdots + x_na_{1n} &= 0 \\ &\ \,\vdots \\ x_1a_{m1} + \cdots + x_na_{mn} &= 0 \end{aligned}$$

has a non-trivial solution, because $n > m$. In view of the preceding remark, such a solution $(x_1, \ldots, x_n)$ is such that

$$x_1w_1 + \cdots + x_nw_n = 0.$$

as desired.

Theorem 5.2. *Let V be a vector space and suppose that one basis has n elements, and another basis has m elements. Then $m = n$.*

Proof. We apply Theorem 5.1 to the two bases. Theorem 5.1 implies that both alternatives $n > m$ and $m > n$ are impossible, and hence $m = n$.

Let V be a vector space having a basis consisting of n elements. We shall say that n is the **dimension** of V. If V consists of O alone, then V does not have a basis, and we shall say that V has dimension 0.

We may now reformulate the definitions of a line and a plane in an arbitrary vector space V. A **line passing through the origin** is simply a one-dimensional subspace. A **plane passing through the origin** is simply a two-dimensional subspace.

An arbitrary **line** is obtained as the translation of a one-dimensional subspace. An arbitrary **plane** is obtained as the translation of a two-dimensional subspace. When a basis $\{v_1\}$ has been selected for a one-dimensional space, then the points on a line are expressed in the usual form

$$P + t_1 v_1 \text{ with all possible numbers } t_1.$$

When a basis $\{v_1, v_2\}$ has been selected for a two-dimensional space, then the points on a plane are expressed in the form

$$P + t_1 v_1 + t_2 v_2 \text{ with possible numbers } t_1,\ t_2.$$

Let $\{v_1, \dots, v_n\}$ be a set of elements of a vector space V. Let r be a positive integer $\leqq n$. We shall say that $\{v_1, \dots, v_r\}$ is a **maximal** subset of linearly independent elements if $v_1, \dots, v_r$ are linearly independent, and if in addition, given any v_i with $i > r$, the elements $v_1, \dots, v_r$, v_i are linearly dependent.

The next theorem gives us a useful criterion to determine when a set of elements of a vector space is a basis.

Theorem 5.3. *Let $\{v_1, \dots, v_n\}$ be a set of generators of a vector space V. Let $\{v_1, \dots, v_r\}$ be a maximal subset of linearly independent elements. Then $\{v_1, \dots, v_r\}$ is a basis of V.*

Proof. We must prove that $v_1, \dots, v_r$ generate V. We shall first prove that each v_i (for $i > r$) is a linear combination of $v_1, \dots, v_r$. By hypothesis, given v_i, there exists numbers $x_1, \dots, x_r$, y not all 0 such that

$$x_1 v_1 + \cdots + x_r v_r + y v_i = O.$$

Furthermore, $y \neq 0$, because otherwise, we would have a relation of linear dependence for $v_1, \dots, v_r$. Hence we can solve for v_i, namely

$$v_i = \frac{x_1}{-y} v_1 + \cdots + \frac{x_r}{-y} v_r,$$

thereby showing that v_i is a linear combination of $v_1, \dots, v_r$.

Next, let v be any element of V. There exist numbers $c_1, \dots, c_n$ such that

$$v = c_1 v_1 + \cdots + c_n v_n.$$

In this relation, we can replace each v_i $(i > r)$ by a linear combination of $v_1, \dots, v_r$. If we do this, and then collect terms, we find that we have expressed v as a linear combination of $v_1, \dots, v_r$. This proves that $v_1, \dots, v_r$ generate V, and hence form a basis of V.

We shall now give criteria which allow us to tell when elements of a vector space constitute a basis.

Let $v_1, \dots, v_n$ be linearly independent elements of a vector space V. We shall say that they form a **maximal set of linearly independent elements of** V if given any element w of V, the elements $w, v_1, \dots, v_n$ are linearly dependent.

Theorem 5.4. *Let V be a vector space, and $\{v_1, \dots, v_n\}$ a maximal set of linearly independent elements of V. Then $\{v_1, \dots, v_n\}$ is a basis of V.*

Proof. We must now show that $v_1, \dots, v_n$ generate V, i.e. that every element of V can be expressed as a linear combination of $v_1, \dots, v_n$. Let w be an element of V. The elements $w, v_1, \dots, v_n$ of V must be linearly dependent by hypothesis, and hence there exist numbers $x_0, x_1, \dots, x_n$ not all 0 such that

$$x_0 w + x_1 v_1 + \cdots + x_n v_n = O.$$

We cannot have $x_0 = 0$, because if that were the case, we would obtain a relation of linear dependence among $v_1, \dots, v_n$. Therefore we can solve for w in terms of $v_1, \dots, v_n$, namely

$$w = -\frac{x_1}{x_0} v_1 - \cdots - \frac{x_n}{x_0} v_n.$$

This proves that w is a linear combination of $v_1, \dots, v_n$, and hence that $\{v_1, \dots, v_n\}$ is a basis.

Theorem 5.5. *Let V be a vector space of dimension n, and let $v_1, \dots, v_n$ be linearly independent elements of V. Then $v_1, \dots, v_n$ constitute a basis of V.*

Proof. According to Theorem 5.1., $\{v_1, \dots, v_n\}$ is a maximal set of linearly independent elements of V. Hence it is a basis by Theorem 5.4.

Theorem 5.6. *Let V be a vector space of dimension n and let W be a subspace, also of dimension n. Then $W = V$.*

Proof. A basis for W must also be a basis for V.

Theorem 5.7. *Let V be a vector space of dimension n. Let r be a positive integer with $r < n$, and let $v_1, \dots, v_r$ be linearly independent elements of V. Then one can find elements $v_{r+1}, \dots, v_n$ such that*

$$\{v_1, \dots, v_n\}$$

is a basis of V.

Proof. Since $r < n$ we know that $\{v_1, \dots, v_r\}$ cannot form a basis of V, and thus cannot be a maximal set of linearly independent elements of V. In particular, we can find v_{r+1} in V such that

$$v_1, \dots, v_{r+1}$$

are linearly independent. If $r + 1 < n$, we can repeat the argument. We can thus proceed stepwise (by induction) until we obtain n linearly independent elements $\{v_1, \dots, v_n\}$. These must be a basis by Theorem 5.4, and our corollary is proved.

Theorem 5.8. *Let V be a vector space having a basis consisting of n elements. Let W be a subspace which does not consist of O alone. Then W has a basis, and the dimension of W is $\leqq n$.*

Proof. Let w_1 be a non-zero element of W. If $\{w_1\}$ is not a maximal set of linearly independent elements of W, we can find an element w_2 of W such that w_1, w_2 are linearly independent. Proceeding in this manner, one element at a time, there must be an integer $m \leqq n$ such that we can find linearly independent elements $w_1, w_2, \dots, w_m$, and such that

$$\{w_1, \dots, w_m\}$$

is a maximal set of linearly independent elements of W (by Theorem 5.1 we cannot go on indefinitely finding linearly independent elements, and the number of such elements is at most n). If we now use Theorem 5.4, we conclude that $\{w_1, \dots, w_m\}$ is a basis for W.

Exercises III, §5

1. What is the dimension of the following spaces (refer to Exercises 11 through 16 of the preceding section):
   (a) 2×2 matrices (b) $m \times n$ matrices
   (c) $n \times n$ matrices all of whose components are 0 expect possibly on the diagonal.
   (d) Upper triangular $n \times n$ matrices.
   (e) Symmetric 2×2 matrices.
   (f) Symmetric 3×3 matrices.
   (g) Symmetric $n \times n$ matrices.

2. Let V be a subspace of $\mathbf{R}^2$. What are the possible dimensions for V? Show that if $V \neq \mathbf{R}^2$, then either $V = \{O\}$, or V is a straight line passing through the origin.

3. Let V be a subspace of $\mathbf{R}^3$. What are the possible dimensions for V? Show that if $V \neq \mathbf{R}^3$, then either $V = \{O\}$, or V is a straight line passing through the origin, or V is a plane passing through the origin.

III, §6. The Rank of a Matrix

Let

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

be an $m \times n$ matrix. The columns of A generate a vector space, which is a subspace of $\mathbf{R}^m$. The dimension of that subspace is called the **column rank** of A. In light of Theorem 5.4, the column rank is equal to the maximum number of linearly independent columns. Similarly, the rows of A generate a subspace of $\mathbf{R}^n$, and the dimension of this subspace is called **the row rank**. Again by Theorem 5.4, the row rank is equal to the maximum number of linearly independent rows. We shall prove below that these two ranks are equal to each other. We shall give two proofs. The first in this section depends on certain operations on the rows and columns of a matrix. Later we shall give a more geometric proof using the notion of perpendicularity.

We define the **row space** of A to be the subspace generated by the rows of A. We define the **column space** of A to be the subspace generated by the columns.

Consider the following operations on the rows of a matrix.

Row 1. Adding a scalar multiple of one row to another.

Row 2. Interchanging rows.

Row 3. Multiplying one row by a non-zero scalar.

These are called the row operations (sometimes, the elementary row operations). We have similar operations for columns, which will be denoted by **Col 1**, **Col 2**, **Col 3** respectively. We shall study the effect of these operations on the ranks.

First observe that each one of the above operations has an inverse operation in the sense that by performing similar operations we can revert to the original matrix. For instance, let us change a matrix A by adding c times the second row to the first. We obtain a new matrix B whose rows are

$$B_1 = A_1 + cA_2,\ A_2, \dots, A_m.$$

If we now add $-cA_2$ to the first row of B, we get back A_1. A similar argument can be applied to any two rows.

If we interchange two rows, then interchange them again, we revert to the original matrix.

If we multiply a row by a number $c \neq 0$, then multiplying again by c^{-1} yields the original row.

Theorem 6.1. *Row and column operations do not change the row rank of a matrix, nor do they change the column rank.*

Proof. First we note that interchanging rows of a matrix does not affect the row rank since the subspace generated by the rows is the same, no matter in what order we take the rows.

Next, suppose we add a scalar multiple of one row to another. We keep the notation before the theorem, so the new rows are

$$B_1 = A_1 + cA_2,\ A_2, \dots, A_m.$$

Any linear combination of the rows of B, namely any linear combination of

$$B_1,\ A_2, \dots, A_m$$

is also a linear combination of $A_1, A_2, \dots, A_m$. Consequently the row space of B is contained in the row space of A. Hence by Theorem 5.6, we have

$$\text{row rank of } B \leqq \text{row rank of } A.$$

Since A is also obtained from B by a similar operation, we get the reverse inequality

$$\text{row rank of } A \leqq \text{row rank of } B.$$

Hence these two row ranks are equal.

Third, if we multiply a row A_i by $c \neq 0$, we get the new row cA_i. But $A_i = c^{-1}(cA_i)$, so the row spaces of the matrix A and the new matrix obtained by multiplying the row by c are the same. Hence the third operation also does not change the row rank.

We could have given the above argument with any pair of rows A_i, A_j $(i \neq j)$, so we have seen that row operations do not change the row rank.

We now prove that they do not change the column rank.

Again consider the matrix obtained by adding a scalar multiple of the second row to the first:

$$B = \begin{pmatrix} a_{11} + ca_{21} & a_{12} + ca_{22} & \cdots & a_{1n} + ca_{2n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Let $B^1,\dots,B^n$ be the columns of this new matrix B. We shall see that the relation of linear dependence between the columns of B are precisely the same as the relations of linear dependence between the columns of A. In other words:

A vector $X = (x_1,\dots,x_n)$ gives a relation of linear dependence

$$x_1B^1 + \cdots + x_nB^n = O$$

between the columns of B if and only if X gives a relation of linear dependence

$$x_1A^1 + \cdots + x_nA^n = O$$

between the columns of A.

Proof. We know from Chapter II, §2 that a relation of linear dependence among the columns can be written in terms of the dot product with the rows of the matrix. So suppose we have a relation

$$x_1B^1 + \cdots + x_nB^n = O.$$

This is equivalent with the fact that

$$X \cdot B_i = 0 \qquad \text{for} \quad i = 1,\dots,m.$$

Therefore

$$X \cdot (A_1 + cA_2) = 0, \quad X \cdot A_2 = 0, \quad \dots, \quad X \cdot A_m = 0.$$

The first equation can be written

$$X \cdot A_1 + cX \cdot A_2 = 0.$$

Since $X \cdot A_2 = 0$ we conclude that $X \cdot A_1 = 0$. Hence X is perpendicular to the rows of A. Hence X gives a linear relation among the columns of A. The converse is proved similarly.

The above statement proves that if r among the columns of B are linearly independent, then r among the columns of A are also linearly independent, and conversely. Therefore A and B have the same column rank.

We leave the verification that the other row operations do not change the column ranks to the reader.

Similarly, one proves that the column operations do not change the row rank. The situation is symmetric between rows and columns. This concludes the proof of the theorem.

Theorem 6.2. *Let A be a matrix of row rank r. By a succession of row and column operations, the matrix can be transformed to the matrix having components equal to 1 on the diagonal of the first r rows and columns, and 0 everywhere else.*

$$
r\left\{\vphantom{\begin{matrix}1\\0\\\vdots\\0\end{matrix}}\right.
\begin{pmatrix}
\overbrace{\begin{matrix}1 & 0 & \cdots & 0\end{matrix}}^{r} & 0 & \cdots & 0 \\
\begin{matrix}0 & 1 & \cdots & 0\end{matrix} & 0 & \cdots & 0 \\
\begin{matrix}\vdots & \vdots & \ddots & \vdots\end{matrix} & \vdots & & \vdots \\
\begin{matrix}0 & 0 & \cdots & 1\end{matrix} & 0 & \cdots & 0 \\
\begin{matrix}0 & 0 & \cdots & 0\end{matrix} & 0 & \cdots & 0 \\
\begin{matrix}\vdots & \vdots & & \vdots\end{matrix} & \vdots & \ddots & \vdots \\
\begin{matrix}0 & 0 & \cdots & 0\end{matrix} & 0 & \cdots & 0
\end{pmatrix}.
$$

In particular, the row rank is equal to the column rank.

Proof. Suppose $r \neq 0$ so the matrix is not the zero matrix. Some component is not zero. After interchanging rows and columns, we may assume that this component is in the upper left-hand corner, that is this component is equal to $a_{11} \neq 0$. Now we go down the first column. We multiply the first row by a_{21}/a_{11} and subtract it from the second row. We then obtain a new matrix with 0 in the first place of the second row. Next we multiply the first row by a_{31}/a_{11} and subtract it from the third row. Then our new matrix has first component equal to 0 in the third row. Proceeding in the same way, we can transform the matrix so that it is of the form

$$
\begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
0 & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & & \vdots \\
0 & a_{m2} & \cdots & a_{mn}
\end{pmatrix}.
$$

Next, we subtract appropriate multiples of the first column from the second, third, ..., n-th column to get zeros in the first row. This transforms the matrix to a matrix of type

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Now we have an $(m-1) \times (n-1)$ matrix in the lower right. If we perform row and column operations on all but the first row and column, then first we do not disturb the first component a_{11}; and second we can repeat the argument, in order to obtain a matrix of the form

$$\begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & a_{m3} & \cdots & a_{mn} \end{pmatrix}.$$

Proceeding stepwise by induction we reach a matrix of the form

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 & 0 \\ 0 & a_{22} & \cdots & 0 & \vdots \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & a_{ss} & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

with diagonal elements $a_{11}, \ldots, a_{ss}$ which are $\neq 0$. We divide the first row by a_{11}, the second row by a_{22}, etc. We then obtain a matrix

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

Thus we have the unit $s \times s$ matrix in the upper left-hand corner, and zeros everywhere else. Since row and column operations do not change the row or column rank, it follows that $r = s$, and also that the row rank is equal to the column rank. This proves the theorem.

Since we have proved that the row rank is equal to the column rank, we can now omit "row" or "column" and just speak of the **rank** of a matrix. Thus by definition the **rank** of a matrix is equal to the dimension of the space generated by the rows.

Remark. Although the systematic procedure provides an effective method to find the rank, in practice one can usually take shortcuts to get as many zeros as possible by making row and column operations, so that at some point it becomes obvious what the rank of the matrix is.

Of course, one can also use the simple mechanism of linear equations to find the rank.

Example. Find the rank of the matrix

$$\begin{pmatrix} 2 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix}.$$

There are only two rows, so the rank is at most 2. On the other hand, the two columns

$$\begin{pmatrix} 2 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

are linearly independent, for if a, b are numbers such that

$$a\begin{pmatrix} 2 \\ 0 \end{pmatrix} + b\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

then

$$\begin{aligned} 2a + b &= 0, \\ b &= 0, \end{aligned}$$

so that $a = 0$. Therefore the two columns are linearly independent, and the rank is equal to 2.

Later we shall also see that determinants give a computation way of determining when vectors are linearly independent, and thus can be used to determine the rank.

Example. Find the rank of the matrix.

$$\begin{pmatrix} 1 & 2 & -3 \\ 2 & 1 & 0 \\ -2 & -1 & 3 \\ -1 & 4 & -2 \end{pmatrix}.$$

We subtract twice the first column from the second and add 3 times the first column to the third. This gives

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & -3 & 6 \\ -2 & 3 & -3 \\ -1 & 6 & -5 \end{pmatrix}.$$

We add 2 times the second column to the third. This gives

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & -3 & 0 \\ -2 & 3 & 3 \\ -1 & 6 & 7 \end{pmatrix}.$$

This matrix is in column echelon form, and it is immediate that the first three rows or columns are linearly independent. Since there are only three columns, it follows that the rank is 3.

Exercises III, §6

1. Find the rank of the following matrices.

(a) $\begin{pmatrix} 2 & 1 & 3 \\ 7 & 2 & 0 \end{pmatrix}$

(b) $\begin{pmatrix} -1 & 2 & -2 \\ 3 & 4 & -5 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & 2 & 7 \\ 2 & 4 & -1 \end{pmatrix}$

(d) $\begin{pmatrix} 1 & 2 & -3 \\ -1 & -2 & 3 \\ 4 & 8 & -12 \\ 0 & 0 & 0 \end{pmatrix}$

(e) $\begin{pmatrix} 2 & 0 \\ 0 & -5 \end{pmatrix}$

(f) $\begin{pmatrix} -1 & 0 & 1 \\ 0 & 2 & 3 \\ 0 & 0 & 7 \end{pmatrix}$

(g) $\begin{pmatrix} 2 & 0 & 0 \\ -5 & 1 & 2 \\ 3 & 8 & -7 \end{pmatrix}$

(h) $\begin{pmatrix} 1 & 2 & -3 \\ -1 & -2 & 3 \\ 4 & 8 & -12 \\ 1 & -1 & 5 \end{pmatrix}$

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

2. Let A be a triangular matrix

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

Assume that none of the diagonal elements is equal to 0. What is the rank of A?

3. Let A be an $m \times n$ matrix and let B be an $n \times r$ matrix, so we can form the product AB.
   (a) Show that the columns of AB are linear combinations of the columns of A. Thus prove that

$$\operatorname{rank} AB \leqq \operatorname{rank} A.$$

   (b) Prove that $\operatorname{rank} AB \leqq \operatorname{rank} B$. [*Hint*: Use the fact that

$$\operatorname{rank} AB = \operatorname{rank} {}^t(AB)$$

   and

$$\operatorname{rank} B = \operatorname{rank} {}^tB.]$$