Basic Linear algebra

Andrew Baker

[21/12/2008]

Department of Mathematics, University of Glasgow.
E-mail address: a.baker@maths.gla.ac.uk
URL: http://www.maths.gla.ac.uk/∼ajb

Linear Algebra is one of the most important basic areas in Mathematics, having at least as great an impact as Calculus, and indeed it provides a significant part of the machinery required to generalise Calculus to vector-valued functions of many variables. Unlike many algebraic systems studied in Mathematics or applied within or outwith it, many of the problems studied in Linear Algebra are amenable to systematic and even algorithmic solutions, and this makes them implementable on computers – this explains why so much calculational use of computers involves this kind of algebra and why it is so widely used. Many geometric topics are studied making use of concepts from Linear Algebra, and the idea of a linear transformation is an algebraic version of geometric transformation. Finally, much of modern abstract algebra builds on Linear Algebra and often provides concrete examples of general ideas.

These notes were originally written for a course at the University of Glasgow in the years 2006–7. They cover basic ideas and techniques of Linear Algebra that are applicable in many subjects including the physical and chemical sciences, statistics as well as other parts of mathematics. Two central topics are: the basic theory of *vector spaces* and the concept of a *linear transformation*, with emphasis on the use of matrices to represent linear maps. Using these, a geometric notion of *dimension* can be made mathematically rigorous leading its widespread appearance in physics, geometry, and many parts of mathematics.

The notes end by discussing *eigenvalues* and *eigenvectors* which play a rôle in the theory of *diagonalisation* of square matrices, as well as many applications of linear algebra such as in geometry, differential equations and physics.

There are some assumptions that the reader will already have met vectors in 2 and 3-dimensional contexts, and has familiarity with their algebraic and geometric aspects. Basic algebraic theory of matrices is also assumed, as well as the solution of systems of linear equations using Gaussian elimination and row reduction of matrices. Thus the notes are suitable for a secondary course on the subject, building on existing foundations.

There are very many books on Linear Algebra. The Bibliography lists some at a similar level to these notes. University libraries contain many other books that may be useful and there are some helpful Internet sites discussing aspects of the subject.

Contents

CHAPTER 1

Vector spaces and subspaces

1.1. Fields of scalars

Before discussing vectors, first we explain what is meant by *scalars*. These are 'numbers' of various types together with algebraic operations for combining them. The main examples we will consider are the *rational numbers* $\mathbb{Q}$, the *real numbers* $\mathbb{R}$ and the *complex numbers* $\mathbb{C}$. But mathematicians routinely work with other fields such as the *finite fields* (also known as *Galois fields*) $\mathbb{F}_{p^n}$ which are important in coding theory, cryptography and other modern applications.

DEFINITION 1.1. A *field of scalars* (or just a *field*) consists of a set F whose elements are called *scalars*, together with two algebraic operations, *addition* $+$ and *multiplication* $\times$, for combining every pair of scalars $x, y \in F$ to give new scalars $x + y \in F$ and $x \times y \in F$. These operations are required to satisfy the following rules which are sometimes known as the *field axioms*.

Associativity: For $x, y, z \in F$,

$$(x + y) + z = x + (y + z),$$
$$(x \times y) \times z = x \times (y \times z).$$

Zero and unity: There are unique and distinct elements $0, 1 \in F$ such that for $x \in F$,

$$x + 0 = x = 0 + x,$$
$$x \times 1 = x = 1 \times x.$$

Distributivity: For $x, y, z \in F$,

$$(x + y) \times z = x \times z + y \times z,$$
$$z \times (x + y) = z \times x + z \times y.$$

Commutativity: For $x, y \in F$,

$$x + y = y + x,$$
$$x \times y = y \times x.$$

Additive and multiplicative inverses: For $x \in F$ there is a unique element $-x \in F$ (the *additive inverse* of x) for which

$$x + (-x) = 0 = (-x) + x.$$

For each non-zero $y \in F$ there is a unique element $y^{-1} \in F$ (the *multiplicative inverse* of y) for which

$$y \times (y^{-1}) = 1 = (y^{-1}) \times y.$$

REMARK 1.2.

- Usually we just write xy instead of $x \times y$, and then we always have $xy = yx$.

- Because of commutativity, some of the above rules are redundant in the sense that they are consequences of others.
- When working with vectors we will always have a specific field of scalars in mind and will make use of all of these rules. It is possible to remove commutativity or multiplicative inverses and still obtain mathematically interesting structures but in this course we definitely always assume the full strength of these rules.
- The most important examples are $\mathbb{R}$ and $\mathbb{C}$ and it is worthwhile noting that the above rules are obeyed by these as well as $\mathbb{Q}$. However, other examples of number systems such as $\mathbb{N}$ and $\mathbb{Z}$ do not obey all of these rules.

PROPOSITION 1.3. *Let F be a field of scalars. For any $x \in F$,*

$$\text{(a)} \quad 0x = 0, \qquad \text{(b)} \quad -x = (-1)x.$$

PROOF. Consider the following calculations which use many of the rules in Definition 1.1. For $x \in F$,

$$0x = (0+0)x = 0x + 0x,$$

hence

$$0 = -(0x) + 0x = -(0x) + (0x + 0x) = (-(0x) + 0x) + 0x = 0 + 0x = 0x.$$

This means that $0x = 0$ as required for (a).

Using (a) we also have

$$x + (-1)x = 1x + (-1)x = (1 + (-1))x = 0x = 0,$$

thus establishing (b). □

EXAMPLE 1.4. Let F be a field. Let $a, b \in F$ and assume that $a \neq 0$. Show that the equation

$$ax = b$$

has a unique solution for $x \in F$.

Challenge: Now suppose that $a_{ij}, b_1, b_2 \in F$ for $i, j = 1, 2$. With the aid of the 'usual' method of solving a pair of simultaneous linear equations show that the system

$$\left\{\begin{array}{ccccc} a_{11}x_1 & + & a_{12}x_2 & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & = & b_2 \end{array}\right\}$$

has a unique solution for $x_1, x_2 \in F$ if $a_{11}a_{22} - a_{12}a_{21} \neq 0$. What can be said about solutions when $a_{11}a_{22} - a_{12}a_{21} = 0$?

SOLUTION. As $a \neq 0$ there is an inverse a^{-1}, hence the equation implies that

$$x = 1x = (a^{-1}a)x = a^{-1}(ax) = a^{-1}b,$$

so if x *is* a solution then it must equal $a^{-1}b$. But it is also clear that

$$a(a^{-1}b) = (aa^{-1})b = 1b = b,$$

so this scalar does satisfy the equation. Notice that if $a = 0$, then the equation

$$0x = b$$

can only have a solution if $b = 0$ and in that case any $x \in F$ will work so the solution is not unique.

Challenge: For this you will need to recall things about 2×2 linear systems. The upshot is that the system can have either no or infinitely many solutions. □

1.2. Vector spaces and subspaces

We now come to the key idea of a *vector space*. This involves an abstraction of properties already met with in special cases when dealing with vectors and systems of linear equations. Before meeting the general definition, here are some examples to have in mind as motivation.

The plane and 3-dimensional space are usually modelled with coordinates and their points correspond to elements of $\mathbb{R}^2$ and $\mathbb{R}^3$ which we write in the form (x, y) or (x, y, z). These are usually called vectors and they are added by adding corresponding coordinates. Scalar multiplication is also defined by coordinate-wise multiplication with scalars. These operations have geometric interpretations in terms of the parallelogram rule and dilation of vectors.

It is usual to identify the set of all real column vectors of length n with $\mathbb{R}^n$, where $(x_1, \ldots, x_n)$ corresponds to the column vector $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. Note that the use of the different types of brackets is very important here and will be used in this way throughout the course. Matrix addition and scalar multiplication correspond to coordinate-wise addition and scalar multiplication in $\mathbb{R}^n$. More generally, the set of all $m \times n$ real matrices has an addition and scalar multiplication.

In the above, $\mathbb{R}$ can be replaced by $\mathbb{C}$ and all the algebraic properties are still available.

Now we give the very important definition of a vector space.

Definition 1.5. A *vector space over a field of scalars F* consists of a set V whose elements are called *vectors* together with two algebraic operations, $+$ (*addition of vectors*) and $\cdot$ (*multiplication by scalars*). Vectors will usually be denoted with boldface symbols such as $\mathbf{v}$ which is hand written as $\underset{\sim}{v}$. The operations $+$ and $\cdot$ are required to satisfy the following rules, which are sometimes known as the *vector space axioms*.

Associativity: For $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $s, t \in F$,

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}),$$
$$(st) \cdot \mathbf{v} = s \cdot (t \cdot \mathbf{v}).$$

Zero and unity: There is a unique element $\mathbf{0} \in V$ such that for $\mathbf{v} \in V$,

$$\mathbf{v} + \mathbf{0} = \mathbf{v} = \mathbf{0} + \mathbf{v},$$

and multiplication by $1 \in F$ satisfies

$$1 \cdot \mathbf{v} = \mathbf{v}.$$

Distributivity: For $s, t \in F$ and $\mathbf{u}, \mathbf{v} \in V$,

$$(s + t) \cdot \mathbf{v} = s \cdot \mathbf{v} + t \cdot \mathbf{v},$$
$$s \cdot (\mathbf{u} + \mathbf{v}) = s \cdot \mathbf{u} + s \cdot \mathbf{u}.$$

Commutativity: For $\mathbf{u}, \mathbf{v} \in V$,

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}.$$

Additive inverses: For $\mathbf{v} \in V$ there is a unique element $-\mathbf{v} \in V$ for which

$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0} = (-\mathbf{v}) + \mathbf{v}.$$

Again it is normal to write $s\mathbf{v}$ in place of $s \cdot \mathbf{v}$ when the meaning is clear. Care may need to be exercised in correctly interpreting expressions such as

$$(st)\mathbf{v} = (s \times t) \cdot \mathbf{v}$$

for $s, t \in F$ and $\mathbf{v} \in V$; here the product $(st) = (s \times t)$ is calculated in F, while $(st)\mathbf{v} = (st) \cdot \mathbf{v}$ is calculated in V. Note that we do not usually *multiply* vectors together, although there is a special situation with the *vector* (or *cross*) product defined on $\mathbb{R}^3$, but we will not often consider that in this course.

From now on, let V be a vector space over a field of scalars F. If $F = \mathbb{R}$ we refer to V as a *real vector space*, while if $F = \mathbb{C}$ we refer to it as a *complex vector space*.

PROPOSITION 1.6. *For $s \in F$ and $\mathbf{v} \in V$, the following identities are valid:*

(a) $$0\mathbf{v} = \mathbf{0},$$

(b) $$s\mathbf{0} = \mathbf{0},$$

(c) $$(-s)\mathbf{v} = -(s\mathbf{v}) = s(-\mathbf{v}),$$

and in particular, taking $s = 1$ we have

(d) $$-\mathbf{v} = (-1)\mathbf{v}.$$

PROOF. The proofs are similar to those of Proposition 1.3, but need care in distinguishing multiplication of scalars and scalar multiplication of vectors. □

Here are some important examples of vector spaces which will be met throughout the course.

EXAMPLE 1.7. Let F be any field of scalars. For $n = 1, 2, 3, \ldots$, let F^n denote the set of all n-tuples $\mathbf{t} = (t_1, \ldots, t_n)$ of elements of F. For $s \in F$ and $\mathbf{u}, \mathbf{v} \in F^n$, we define

$$\begin{aligned}(u_1, \ldots, u_n) + (v_1, \ldots, v_n) =& (u_1 + v_1, \ldots, u_n + v_n),\\ s \cdot (v_1, \ldots, v_n) =& s(v_1, \ldots, v_n) = (sv_1, \ldots, sv_n).\end{aligned}$$

The zero vector is $\mathbf{0} = (0, \ldots, 0)$. The cases $F = \mathbb{R}$ and $F = \mathbb{C}$ will be the most important in these notes.

EXAMPLE 1.8. Let F be a field. The set $\mathrm{M}_{m\times n}(F)$ of all $m \times n$ matrices with entries in F forms a vector space over F with addition and multiplication by scalars defined in the usual way. We identify the set $\mathrm{M}_{n\times 1}(F)$ with F^n, to obtain the vector space of Example 1.7. We also set $\mathrm{M}_n(F) = \mathrm{M}_{n\times n}(F)$. In all cases, the zero vector of $\mathrm{M}_{m\times n}(F)$ is the zero matrix $O_{m\times n}$.

EXAMPLE 1.9. The complex numbers $\mathbb{C}$ can be viewed as a vector space over $\mathbb{R}$ with the usual addition and scalar multiplication being the usual multiplication.

More generally, if V is a complex vector space then we can view it as a *real* vector space by only allowing real numbers as scalars. For example, $\mathrm{M}_{m\times n}(\mathbb{C})$ becomes a real vector space in this way as does $\mathbb{C}^n$. We then refer to the real vector space V as the *underlying real vector space* of the complex vector space V.

EXAMPLE 1.10. Let F be a field. Consider the set $F[X]$ consisting of all polynomials in X with coefficients in F. We define addition and multiplication by scalars in F by

$$\begin{aligned}&(a_0 + a_1X + \cdots + a_rX^r) + (b_0 + b_1X + \cdots + b_rX^r)\\ &\qquad\qquad = (a_0 + b_0) + (a_1 + b_1)X + \cdots + (a_r + b_r)X^r,\\ &s \cdot (a_0 + a_1X + \cdots + a_rX^r) = (sa_0) + (sa_1)X + \cdots + (sa_r)X^r.\end{aligned}$$

The zero vector is the zero polynomial

$$\mathbf{0} = 0 + 0X + \cdots + 0X^r = 0.$$

EXAMPLE 1.11. Take $\mathbb{R}$ to be the field of scalars and consider the set $\mathcal{D}(\mathbb{R})$ consisting of all infinitely differentiable functions $\mathbb{R} \longrightarrow \mathbb{R}$, *i.e.*, functions $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ for which all possible derivatives $f^{(n)}(x)$ exist for $n \geqslant 1$ and $x \in \mathbb{R}$. We define addition and multiplication by scalars as follows. Let $t \in \mathbb{R}$ and $f, g \in \mathcal{D}(\mathbb{R})$, then $f + g$ and $s \cdot f$ are the functions given by the following rules: for $x \in \mathbb{R}$,

$$(f+g)(x) = f(x) + g(x),$$
$$(s \cdot f)(x) = sf(x).$$

Of course we really ought check that all the derivatives of $f + g$ and $s \cdot f$ exist so that these function are in $\mathcal{D}(\mathbb{R})$, but this is an exercise in Calculus. The zero vector here is the constant function which sends every real number to 0, and this is usually written 0 in Calculus, although this might seem confusing in our context! More generally, for a given real number a, it is also standard to write a for the constant function which sends every real number to a.

Note that Example 1.10 can also be viewed as a vector space consisting of functions since every polynomial over a field can be thought of as the function obtained by evaluation of the variable at values in F. When $F = \mathbb{R}$, $\mathbb{R}[X] \subseteq \mathcal{D}(\mathbb{R})$ and we will see that this is an example of a *vector subspace* which will be defined soon after some motivating examples.

EXAMPLE 1.12. Let V be a vector space over a field of scalars F. Suppose that $W \subseteq V$ contains $\mathbf{0}$ and is *closed* under addition and multiplication by scalars, *i.e.*, for $s \in F$ and $\mathbf{u}, \mathbf{v} \in W$, we have

$$\mathbf{u} + \mathbf{v} \in W, \qquad s\mathbf{u} \in W.$$

Then W is also a vector space over F.

EXAMPLE 1.13. Consider the case where $F = \mathbb{R}$ and $V = \mathbb{R}^2$. Then the line L with equation

$$2x - y = 0$$

consists of all vectors of the form $(t, 2t)$ with $t \in \mathbb{R}$, *i.e.*,

$$L = \{(t, 2t) \in \mathbb{R}^2 : t \in \mathbb{R}\}.$$

Clearly this passes through the origin so it contains $\mathbf{0}$. It is also closed under addition and scalar multiplication: if $(t, 2t), (t_1, 2t_1), (t_2, 2t_2) \in L$ and $s \in \mathbb{R}$ then

$$\begin{aligned}(t_1, 2t_1) + (t_2, 2t_2) &= (t_1 + t_2, 2t_1 + 2t_2) \\ &= (t_1 + t_2, 2(t_1 + t_2)) \in L, \\ s(t, 2t) &= (st, 2(st)) \in L.\end{aligned}$$

More generally, any line with equation of form

$$ax + by = 0$$

with $(a, b) \neq \mathbf{0} = (0, 0)$ is similarly closed under addition and multiplication by scalars.

EXAMPLE 1.14. Consider the case where $F = \mathbb{R}$ and $V = \mathbb{R}^3$. If $(a, b, c) \in \mathbb{R}^3$ is non-zero, then consider the plane P with equation

$$ax + by + cz = 0,$$

thus as a set P is given by

$$P = \{(x, y, z) \in \mathbb{R}^3 : ax + by + cz = 0\}.$$

Then P contains $\mathbf{0}$ and is closed under addition and scalar multiplication since whenever $(x, y, z), (x_1, y_1, z_1), (x_2, y_2, z_2) \in P$ and $s \in \mathbb{R}$, the sum $(x_1, y_1, z_1) + (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2)$ satisfies

$$a(x_1 + x_2) + b(y_1 + y_2) + c(z_1 + z_2) = (ax_1 + by_1 + cz_1) + (ax_2 + by_2 + cz_2) = 0,$$

while the scalar multiple $s(x, y, z) = (sx, sy, sz)$ satisfies

$$a(sx) + b(sy) + c(sz) = s(ax + by + cz) = 0,$$

showing that

$$(x_1, y_1, z_1) + (x_2, y_2, z_2), s(x, y, z) \in P.$$

A line in $\mathbb{R}^3$ through the origin consists of all vectors of the form $t\mathbf{u}$ for $t \in \mathbb{R}$, where $\mathbf{u}$ is some non-zero direction vector. This is also closed under addition and scalar multiplication.

Definition 1.15. Let V be a vector space over a field of scalars F. Suppose that the subset $W \subseteq V$ is non-empty and is closed under addition and multiplication by scalars, thus it forms a vector space over F. Then W is called a (*vector*) *subspace* of V.

Remark 1.16.

- In order to verify that W is non-empty it is usually easiest to show that it contains the zero vector $\mathbf{0}$.
- The conditions for W to be closed under addition and scalar multiplication are equivalent to the single condition that for all $s, t \in F$ and $\mathbf{u}, \mathbf{v} \in W$,
$$s\mathbf{u} + t\mathbf{v} \in W.$$
- If $W \subseteq V$ is a subspace, then for any $\mathbf{w} \in W$, all of its scalar multiples (including $0\mathbf{w} = \mathbf{0}$ and $(-1)\mathbf{w} = -\mathbf{w}$) are also in W.

Example 1.17. Let V be a vector space over a field F.

- The sets $\{\mathbf{0}\}$ and V are both subspaces of V, usually thought of as its uninteresting subspaces.
- If $\mathbf{v} \in V$ is a non-zero vector, then $\{\mathbf{v}\}$ is never a subspace of V since (for example) $0\mathbf{v} = \mathbf{0}$ is not an element of $\{\mathbf{v}\}$.
- The smallest subspace of V containing a given non-zero vector $\mathbf{v}$ is the *line spanned by* $\mathbf{v}$ which is the set of all scalar multiplies of $\mathbf{v}$,
$$\{t\mathbf{v} : t \in F\}.$$

Before giving more examples, here are some *non-examples.*

Example 1.18. Consider the real vector space $\mathbb{R}^2$ and the following subsets with the same addition and scalar multiplication.

(a) $V_1 = \{(x, y) \in \mathbb{R}^2 : x \geqslant 0\}$ is not a real vector space: for example, given any vector $(x, y) \in V_1$ with $x > 0$, there is no additive inverse $-(x, y) \in V_1$ since this would have to be $(-x, -y)$ which has $-x < 0$. So V_1 is not closed under addition.

(b) $V_2 = \{(x, y) \in \mathbb{R}^2 : y = x^2\}$ is not a real vector space: for example, $(1, 1) \in V_2$ but $2(1, 1) = (2, 2)$ does not satisfy $y = x^2$ since $x^2 = 4$ and $y = 2$. So V_2 is not closed under scalar multiplication.

(c) $V_3 = \{(x, y) \in \mathbb{R}^2 : x = 2\}$ is not a real vector space: for example, $(2, 0), (2, 1) \in V_3$ but $(2, 0) + (2, 1) = (4, 1) \notin V_3$, so V_3 is not closed under addition.

Here are some further examples of subspaces.

EXAMPLE 1.19. Suppose that F is a field and that we have a system of homogenous linear equations

$$\left\{\begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & 0 \\ \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}\right\} \tag{1.1}$$

for scalars $a_{ij} \in F$. Then the set W of solutions $(x_1, \ldots, x_n) = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ in F^n of the system (1.1) is a subspace of F^n. This is most easily checked by recalling that the system is equivalent to a single matrix equation

$$A\mathbf{x} = \mathbf{0}, \tag{1.2}$$

where $A = [a_{ij}]$ and then verifying that given solutions $\mathbf{x}, \mathbf{y} \in W$ and $s \in F$,

$$A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

and

$$A(s\mathbf{x}) = s(A\mathbf{x}) = s(A\mathbf{0}) = s\mathbf{0} = \mathbf{0}.$$

This example indicates an extremely important relationship between vector space ideas and linear equations, and this is one of the main themes in the subject of Linear Algebra.

EXAMPLE 1.20. Recall the vector space $F[X]$ of Example 1.10. Consider the following subsets of $F[X]$:

$$\begin{aligned} W_0 &= F[X], \\ W_1 &= \{f(X) \in F[X] : f(0) = 0\}, \\ W_2 &= \{f(X) \in F[X] : f(0) = f'(0) = 0\}, \end{aligned}$$

and in general, for $n \geqslant 1$,

$$W_n = \{f(X) \in F[X] : \text{for } k = 0, \ldots, n-1,\ f^{(k)}(0) = 0\}.$$

Show that each W_n is a subspace of $F[X]$ and that W_n is a subspace of W_{n-1}.

SOLUTION. Let $n \geqslant 1$. Suppose that $s, t \in F$ and $f(X), g(X) \in W_n$. Then for each $k = 0, \ldots, n-1$ we have

$$f^{(k)}(0) = 0 = g^{(k)}(0),$$

hence

$$\frac{d^k}{dX^k}(sf(X) + tg(X)) = sf^{(k)}(X) + tg^{(k)}(X),$$

and evaluating at $X = 0$ gives

$$\frac{d^k}{dX^k}(sf + tg)(0) = sf^{(k)}(0) + tg^{(k)}(0) = 0.$$

This shows that W_n is closed under addition and scalar multiplication, therefore it is a subspace of $F[X]$. Clearly $W_n \subseteq W_{n-1}$ and as it is closed under addition and scalar multiplication, it is a subspace of W_{n-1}. $\square$

We can combine subspaces to form new subspaces. Given subsets $Y_1, \ldots, Y_n \subseteq X$ of a set X, their *intersection* is the subset

$$Y_1 \cap Y_2 \cap \cdots \cap Y_n = \{x \in X : \text{for } k = 1, \ldots, n,\ x \in Y_k\} \subseteq X,$$

while their *union* is the subset

$$Y_1 \cup Y_2 \cup \cdots \cup Y_n = \{x \in X : \text{there is a } k = 1, \ldots, n \text{ such that } x \in Y_k\} \subseteq X,$$

PROPOSITION 1.21. *Let V be a vector space. If $W_1, \ldots, W_n \subseteq V$ are subspaces of V, then $W_1 \cap \cdots \cap W_n$ is a subspace of V.*

PROOF. Let $s, t \in F$ and $\mathbf{u}, \mathbf{v} \in W_1 \cap \cdots \cap W_n$. Then for each $k = 1, \ldots, n$ we have $\mathbf{u}, \mathbf{v} \in W_k$, and since W_k is a subspace,

$$s\mathbf{u} + t\mathbf{v} \in W_k.$$

But this means that

$$s\mathbf{u} + t\mathbf{v} \in W_1 \cap \cdots \cap W_n,$$

i.e., $W_1 \cap \cdots \cap W_n$ is a subspace of V. □

On the other hand, the *union* of subspace does not always behave so well.

EXAMPLE 1.22. Consider the real vector space $\mathbb{R}^2$ and the subspaces

$$W' = \{(s, 0) : s \in \mathbb{R}\}, \quad W'' = \{(0, t) : t \in \mathbb{R}\}.$$

Show that $W' \cup W''$ is not a subspace of $\mathbb{R}^2$ and determine $W' \cap W''$.

SOLUTION. The union of these subspaces is

$$W' \cup W'' = \{\mathbf{v} \in \mathbb{R}^2 : \mathbf{v} = (s, 0) \text{ or } \mathbf{v} = (0, s) \text{ for some } s \in \mathbb{R}\}.$$

Then $(1, 0) \in W'$ and $(0, 1) \in W''$, but

$$(1, 1) = (1, 0) + (0, 1)$$

is visibly not in $W' \cup W''$, so this is not closed under addition. (However, it is closed under scalar multiplication, so care is required when checking this sort of example.)

If $\mathbf{x} \in W' \cap W''$ then $\mathbf{x} = (s, 0)$ and $\mathbf{x} = (0, t)$ for some $s, t \in \mathbb{R}$, which is only possible if $s = t = 0$ and so $\mathbf{x} = (0, 0)$. Therefore $W' \cap W'' = \{\mathbf{0}\}$. □

Here is another way to think about Example 1.19.

EXAMPLE 1.23. Suppose that F is a field and that we have a system of homogenous linear equations

$$\left\{\begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & 0 \\ \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}\right\}$$

with the $a_{ij} \in F$.

For each $r = 1, \ldots, m$, let

$$W_r = \{(x_1, \ldots, x_n) : a_{r1}x_1 + \cdots + a_{rn}x_n = 0\} \subseteq F^n,$$

which is a subspace of F^n.

Then the set W of solutions $(x_1, \dots, x_n)$ of the above linear system is given by the intersection

$$W = W_1 \cap \cdots \cap W_m,$$

therefore W is a subspace of V.

EXAMPLE 1.24. In the real vector space $\mathbb{R}^4$, determine the intersection $U \cap V$ of the subspaces

$$U = \{(w, x, y, z) \in \mathbb{R}^4 : x + 2y = 0 = w + 2x - z\}, \quad V = \{(w, x, y, z) \in \mathbb{R}^4 : w + x + 2y = 0\}.$$

SOLUTION. Before doing this we first use elementary row operations to find the general solutions of the systems of equations used to define U and V.

For U we consider the system

$$\left\{\begin{array}{ccccccccc} & & x & + & 2y & & & = & 0 \\ w & + & 2x & & & - & z & = & 0 \end{array}\right\}$$

with augmented matrix

$$\begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 1 & 2 & 0 & -1 & 0 \end{bmatrix} \underset{R_1 \leftrightarrow R_2}{\sim} \begin{bmatrix} 1 & 2 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 & 0 \end{bmatrix} \underset{R_1 \rightarrow R_1 - 2R_2}{\sim} \begin{bmatrix} 1 & 0 & -4 & -1 & 0 \\ 0 & 1 & 2 & 0 & 0 \end{bmatrix},$$

where the last matrix is a reduced echelon matrix. So the general solution vector of this system is $(4s + t, -2s, s, t)$ where $s, t \in \mathbb{R}$. Thus we have

$$U = \{(4s + t, -2s, s, t) \in \mathbb{R}^4 : s, t \in \mathbb{R}\}.$$

Similarly, the general solution vector of the equation

$$w + x + 2y = 0$$

is $(-r - 2s, r, s, t)$ where $r, s, t \in \mathbb{R}$ and we have

$$V = \{(-r - 2s, r, s, t) \in \mathbb{R}^4 : r, s, t \in \mathbb{R}\}.$$

Now to find $U \cap V$ we have to solve the system

$$\left\{\begin{array}{ccccccccc} & & x & + & 2y & & & = & 0 \\ w & + & 2x & & & - & z & = & 0 \\ w & + & x & + & 2y & & & = & 0 \end{array}\right\}$$

with augmented matrix

$$\begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 1 & 2 & 0 & -1 & 0 \\ 1 & 1 & 2 & 0 & 0 \end{bmatrix} \underset{R_1 \leftrightarrow R_2}{\sim} \begin{bmatrix} 1 & 2 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 1 & 1 & 2 & 0 & 0 \end{bmatrix} \underset{R_3 \rightarrow R_3 - R_1}{\sim} \begin{bmatrix} 1 & 2 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & -1 & 2 & 1 & 0 \end{bmatrix}$$

$$\underset{\substack{R_1 \rightarrow R_1 - 2R_2 \\ R_3 \rightarrow R_3 + R_1}}{\sim} \begin{bmatrix} 1 & 0 & -4 & -1 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 4 & 1 & 0 \end{bmatrix} \underset{R_3 \rightarrow (1/4)R_3}{\sim} \begin{bmatrix} 1 & 0 & -4 & -1 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1/4 & 0 \end{bmatrix}$$

$$\underset{\substack{R_1 \rightarrow R_1 + 4R_3 \\ R_2 \rightarrow R_2 - 2R_3}}{\sim} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1/2 & 0 \\ 0 & 0 & 1 & 1/4 & 0 \end{bmatrix},$$

where the last matrix is reduced echelon. The general solution vector is $(0, t/2, -t/4, t)$ where $t \in \mathbb{R}$, thus we have

$$U \cap V = \{(0, t/2, -t/4, t) : t \in \mathbb{R}\} = \{(0, 2s, -s, 4s) : s \in \mathbb{R}\}. \qquad \square$$

CHAPTER 2

Spanning sequences, linear independence and bases

When working with $\mathbb{R}^2$ or $\mathbb{R}^3$ we often use the fact that every vector $\mathbf{v}$ can be expressed in terms of the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2$ or $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ *i.e.*,

$$\mathbf{v} = x\mathbf{e}_1 + y\mathbf{e}_2 \quad \text{or} \quad \mathbf{v} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3,$$

where $\mathbf{v} = (x, y)$ or $\mathbf{v} = (x, y, z)$. Something similar is true for every vector space and involves the concepts of *spanning sequence* and *basis*.

2.1. Linear combinations and spanning sequences

We will always assume that V is a vector space over a field F.

The first notion we will require is that of a *linear combination*.

DEFINITION 2.1. If $t_1, \dots, t_r \in F$ and $\mathbf{v}_1, \dots, \mathbf{v}_r$ is a sequence of vectors in V (some of which may be $\mathbf{0}$ or may be repeated), then the vector

$$t_1\mathbf{v}_1 + \cdots + t_r\mathbf{v}_r \in V$$

is said to be a *linear combination* of the sequence $\mathbf{v}_1, \dots, \mathbf{v}_r$ (or of the vectors $\mathbf{v}_i$). Sometimes it is useful to allow $r = 0$ and view the $\mathbf{0}$ as a linear combination of the elements of the empty sequence $\varnothing$.

A sequence is often denoted $(\mathbf{v}_1, \dots, \mathbf{v}_r)$, but we will not use that notation to avoid confusion with the notation for elements of F^n.

We will sometimes refer to a sequence $\mathbf{v}_1, \dots, \mathbf{v}_r$ of vectors in V as a *sequence in* V. There are various operations which can be performing on sequences such as deleting terms, reordering terms, multiplying terms by non-zero scalars, combining two sequences into a new one by concatinating them in either order.

Notice that $\mathbf{0}$ can be expressed as a linear combination in other ways. For example, when $F = \mathbb{R}$ and $V = \mathbb{R}^2$,

$$3(1, -1) + (-3)(2, 0) + 3(1, 1) = \mathbf{0}, \quad 1(0, 0) = \mathbf{0}, \quad 1(5, 0) + 0(0, 0) + (-5)(1, 0) = \mathbf{0},$$

so $\mathbf{0}$ is a linear combination of each of the sequences

$$(1, -1), (2, 0), (1, 1), \quad \mathbf{0}, \quad (5, 0), \mathbf{0}, (1, 0).$$

For any sequence $\mathbf{v}_1, \dots, \mathbf{v}_r$ we can always write

$$\mathbf{0} = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_r,$$

so $\mathbf{0}$ is a linear combination of any sequence (even the empty sequence of length 0).

Sometimes it is useful to allow infinite sequences as in the next example.

EXAMPLE 2.2. Let F be a field and consider the vector space of polynomials $F[X]$. Then every $f(X) \in F[X]$ has a unique expression of the form

$$f(X) = a_0 + a_1X + \cdots + a_rX^r + \cdots ,$$

where $a_i \in F$ and for large i, $a_i = 0$. Thus every polynomial in $F[X]$ can be uniquely written as a linear combination of the sequence $1, X, X^2, \ldots, X^r, \ldots$.

Such a special set of vectors is called a *basis* and soon we will develop the theory of bases. Bases are extremely important and essential for working in vector spaces. In applications such as to Fourier series, there are bases consisting of trigonometric functions, while in other context orthogonal polynomials play this rôle.

DEFINITION 2.3. Let $S\colon \mathbf{w}_1, \ldots$ be a sequence (possibly infinite) of elements of V. Then S is a *spanning sequence* for V (or is said to *span* V) if for every element $\mathbf{v} \in V$ there is an expression of the form

$$\mathbf{v} = s_1\mathbf{w}_1 + \cdots + s_k\mathbf{w}_k,$$

where $s_1, \ldots, s_k \in F$, *i.e.*, every vector in V is a linear combination of the sequence S.

Notice that in this definition we do not require any sort of uniqueness of the expansion.

DEFINITION 2.4. Let $S\colon \mathbf{w}_1, \ldots$ be a sequence of elements of V. Then the *linear span* of S, $\mathrm{Span}(S) \subseteq V$, is the subset consisting of all linear combinations of S, *i.e.*,

$$\mathrm{Span}(S) = \{t_1\mathbf{w}_1 + \cdots + t_r\mathbf{w}_r : r = 0, 1, 2, \ldots,\ t_i \in F\}.$$

REMARK 2.5.

- By definition, S is a spanning sequence for $\mathrm{Span}(S)$.
- We always have $\mathbf{0} \in \mathrm{Span}(S)$.
- If $S = \varnothing$ is the empty sequence, then we make the definition $\mathrm{Span}(S) = \{\mathbf{0}\}$.

LEMMA 2.6. *Let $S\colon \mathbf{w}_1, \ldots$ be a sequence of elements of V. Then the following hold.*
(a) $\mathrm{Span}(S) \subseteq V$ *is a subspace and S is a spanning sequence of* $\mathrm{Span}(S)$.
(b) *Every subspace of V containing the elements of S contains* $\mathrm{Span}(S)$ *as a subspace, therefore* $\mathrm{Span}(S)$ *is the smallest such subspace.*
(c) *If S' is a subsequence of S, then* $\mathrm{Span}(S') \subseteq \mathrm{Span}(S)$.

In (c), we use the idea that S' *is a subsequence of* S if it is obtained from S by removing some terms and renumbering. For example, if $S\colon \mathbf{u}, \mathbf{v}, \mathbf{w}$ is a sequence then the following are subsequences:

$$S'\colon \mathbf{u}, \mathbf{v}, \quad S''\colon \mathbf{u}, \mathbf{w}, \quad S'''\colon \mathbf{v}.$$

PROOF. (a) It is easy to see that any sum of linear combinations of elements for S is also such a linear combination, similarly for scalar multiples.
(b) If a subspace $W \subseteq V$ contains all the elements of S, then all linear combinations of S are in W, therefore $\mathrm{Span}(S) \subseteq W$.
(c) This follows from (b) since the elements of S' are all in $\mathrm{Span}(S)$. □

DEFINITION 2.7. The vector space V is called *finite dimensional* if it has a spanning sequence $S\colon \mathbf{w}_1, \ldots, \mathbf{w}_\ell$ of finite length ℓ.

Much of the theory we will develop will be about finite dimensional vector spaces although many important examples are *infinite dimensional*.

Here are some further useful properties of sequences and the subspaces they span.

PROPOSITION 2.8. *Let $S\colon \mathbf{w}_1, \ldots$ be a sequence in V. Then the following hold.*
(a) *If $S'\colon \mathbf{w}'_1, \ldots$ is a sequence obtained by permuting* (i.e., *reordering*) *the terms of S, then* $\mathrm{Span}(S') = \mathrm{Span}(S)$*. Hence the span of a sequence is unchanged by permuting its terms.*
(b) *If $r \geqslant 1$ and $\mathbf{w}_r \in \mathrm{Span}(\mathbf{w}_1, \ldots, \mathbf{w}_{r-1})$, then for the sequence*

$$S''\colon \mathbf{w}_1, \ldots, \mathbf{w}_{r-1}, \mathbf{w}_{r+1}, \ldots,$$

we have $\mathrm{Span}(S'') = \mathrm{Span}(S)$*. In particular, we can remove $\mathbf{0}$'s and any repetitions of vectors without changing the span of a sequence.*
(c) *If $s \geqslant 1$ and S''' is obtained from S by replacing $\mathbf{w}_s$ by a vector of the form*

$$\mathbf{w}'''_s = (x_1\mathbf{w}_1 + \cdots + x_{s-1}\mathbf{w}_{s-1}) + \mathbf{w}_s,$$

where $x_1, \ldots, x_{s-1} \in F$, then $\mathrm{Span}(S''') = \mathrm{Span}(S)$.
(d) *If $t_1, \ldots$ is sequence of non-zero scalars, then for the sequence*

$$T\colon t_1\mathbf{w}_1, t_2\mathbf{w}_2, \ldots$$

we have $\mathrm{Span}(T) = \mathrm{Span}(S)$.

PROOF. (a),(b),(d) are simple consequences of the commutativity, associativity and distributivity of addition and scalar multiplication of vectors. (c) follows from the obvious fact that

$$\mathrm{Span}(\mathbf{w}_1, \ldots, \mathbf{w}_{s-1}, \mathbf{w}'''_s) = \mathrm{Span}(\mathbf{w}_1, \ldots, \mathbf{w}_{s-1}, \mathbf{w}_s). \qquad \square$$

2.2. Linear independence and bases

Now we introduce a notion which is related to the uniqueness issue.

DEFINITION 2.9. Let $S\colon \mathbf{w}_1, \ldots$ be a sequence in V.

- S is *linearly dependent* if for some $r \geqslant 1$ there are scalars $s_1, \ldots, s_r \in F$ not all zero and for which
$$s_1\mathbf{w}_1 + \cdots + s_r\mathbf{w}_r = \mathbf{0}.$$
- S is *linearly independent* if it is not linearly dependent. This means that whenever $s_1, \ldots, s_r \in F$ satisfy
$$s_1\mathbf{w}_1 + \cdots + s_r\mathbf{w}_r = \mathbf{0},$$
then $s_1 = \cdots = s_r = 0$.

EXAMPLE 2.10.

- Any sequence S containing $\mathbf{0}$ is linearly dependent since $1 \cdot \mathbf{0} = \mathbf{0}$.
- If a vector $\mathbf{v}$ occurs twice in S then $1\mathbf{v} + (-1)\mathbf{v} = \mathbf{0}$, so again S is linearly dependent.
- If a vector in S can be expressed as a linear combination of the list obtained by deleting it, then S is linearly dependent.
- In the real vector space $\mathbb{R}^2$, the sequence $(1, 2), (3, -1), (0, 7)$ is linearly dependent since
$$3(1, 2) + (-1)(3, -1) + (-1)(0, 7) = (0, 0),$$
while its subsequence $(1, 2), (3, -1)$ is linearly independent.

LEMMA 2.11. *Suppose that $S\colon \mathbf{w}_1, \ldots$ is a linearly dependent sequence in V. Then there is some $k \geqslant 1$ and scalars $t_1, \ldots, t_{k-1} \in F$ for which*

$$\mathbf{w}_k = t_1\mathbf{w}_1 + \cdots + t_{k-1}\mathbf{w}_{k-1}.$$

Hence if S' is the sequence obtained by deleting $\mathbf{w}_k$ then $\operatorname{Span}(S') = \operatorname{Span}(S)$.

PROOF. Since S is linearly dependence, there must be an $r \geqslant 1$ and scalars $s_1, \ldots, s_r \in F$ where not all of the s_i are zero and they satisfy

$$s_1\mathbf{w}_1 + \cdots + s_r\mathbf{w}_r = \mathbf{0}.$$

Let k be the largest value of i for which $s_i \neq 0$. Then

$$s_1\mathbf{w}_1 + \cdots + s_k\mathbf{w}_k = \mathbf{0}.$$

with $s_k \neq 0$. Multiplying by s_k^{-1} gives

$$s_k^{-1}(s_1\mathbf{w}_1 + \cdots + s_k\mathbf{w}_k) = s_k^{-1}\mathbf{0} = \mathbf{0},$$

whence

$$s_k^{-1}s_1\mathbf{w}_1 + s_k^{-1}s_2\mathbf{w}_2 + \cdots + s_k^{-1}s_k\mathbf{w}_k = s_k^{-1}s_1\mathbf{w}_1 + s_k^{-1}s_2\mathbf{w}_2 + \cdots + s_k^{-1}s_{k-1}\mathbf{w}_{k-1} + \mathbf{w}_k = \mathbf{0}.$$

Thus we have

$$\mathbf{w}_k = (-s_k^{-1}s_1)\mathbf{w}_1 + \cdots + (-s_k^{-1}s_{k-1})\mathbf{w}_{k-1}.$$

The equality of the two spans is clear. □

The next result is analogous to Proposition 2.8 for spanning sequences.

PROPOSITION 2.12. *Let $S\colon \mathbf{w}_1, \ldots$ be a sequence in V. Then the following hold.*
(a) *If $S'\colon \mathbf{w}'_1, \ldots$ is a sequence obtained by permuting the terms of S, then S' is linearly dependent (respectively linearly independent) if S is.*
(b) *If $s \geqslant 1$ and S'' is obtained from S by replacing $\mathbf{w}_s$ by a vector of the form*

$$\mathbf{w}''_s = (x_1\mathbf{w}_1 + \cdots + x_{s-1}\mathbf{w}_{s-1}) + \mathbf{w}_s,$$

where $x_1, \ldots, x_{s-1} \in F$, then S'' is linearly dependent (respectively linearly independent) if S is.
(c) *If $t_1, \ldots$ is sequence of non-zero scalars, then for the sequence*

$$T\colon t_1\mathbf{w}_1, t_2\mathbf{w}_2, \ldots,$$

we have T linearly dependent (resp. linearly independent) if S is.

PROOF. Again these are simple consequences of properties of addition and scalar multiplication. □

PROPOSITION 2.13. *Suppose that $S\colon \mathbf{w}_1, \ldots$ be a linearly independent sequence in V. If for some $k \geqslant 1$, the scalars $s_1, \ldots, s_k, t_1, \ldots, t_k \in F$ satisfy*

$$s_1\mathbf{w}_1 + \cdots + s_k\mathbf{w}_k = t_1\mathbf{w}_1 + \cdots + t_k\mathbf{w}_k,$$

then $s_1 = t_1, \ldots, s_k = t_k$. Hence every vector has at most one expression as a linear combination of S.

Proof. The equation can be rewritten as

$$(s_1 - t_1)\mathbf{w}_1 + \cdots + (s_k - t_k)\mathbf{w}_k = \mathbf{0}.$$

Now linear independence shows that each coefficient satisfies $(s_i - t_i) = 0$. $\square$

Proposition 2.14. *Let $S\colon \mathbf{w}_1, \ldots$ be a spanning sequence for V. Then there is a subsequence T of S which is a spanning sequence for V and is also linearly independent.*

Proof. We indicate a proof for the case when S is *finite*, the general case is slightly more involved. Notice that $\mathrm{Span}(S) = V$ and for some $n \geqslant 1$, the sequence is $S\colon \mathbf{w}_1, \ldots, \mathbf{w}_n$.

If S is already linearly independent then we can take $T = S$. Otherwise, by Lemma 2.11, one of the $\mathbf{w}_i$ can be expressed as a linear combination of the others; let k_1 be the largest such i. Then $\mathbf{w}_{k_1}$ can be expressed as a linear combination of the list

$$S_1\colon \mathbf{w}_1, \ldots \mathbf{w}_{k_1-1}, \mathbf{w}_{k_1+1}, \ldots, \mathbf{w}_n$$

and we have $\mathrm{Span}(S_1) = \mathrm{Span}(S) = V$, so S_1 is a spanning sequence for V. The length of S_1 is $n - 1 < n$, so this sequence is shorter than S.

Clearly we can repeat this obtaining new spanning sequences $S_1, S_2, \ldots$ for V, where the lengths of S_r is $n - r$. Eventually we must get to a spanning sequence S_k which cannot be reduced further, but this must be linearly independent. $\square$

Example 2.15. Taking $F = \mathbb{R}$ and $V = \mathbb{R}^2$, show that the sequence

$$S\colon (1,2), (1,1), (-1,0), (1,3)$$

spans $\mathbb{R}^2$ and find a linearly independent spanning subsequence of S.

Solution. Following the idea in the proof of Proposition 2.14, first observe that

$$(1,3) = 3(1,1) + 2(-1,0),$$

so we can replace S by the sequence $S'\colon (1,2), (1,1), (-1,0)$ for which $\mathrm{Span}(S') = \mathrm{Span}(S)$. Next notice that

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

so we can replace S' by the sequence $S''\colon (1,2), (1,1)$ for which $\mathrm{Span}(S'') = \mathrm{Span}(S') = \mathrm{Span}(S)$. For every $(a,b) \in \mathbb{R}^2$, the equation

$$x(1,2) + y(1,1) = (a,b)$$

corresponds to the matrix equation

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix},$$

and since $\det \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = -1 \neq 0$, this has the unique solution

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}.$$

Thus S'' spans $\mathbb{R}^2$ and is linearly independent. $\square$

The next definition is very important.

DEFINITION 2.16. A linearly independent spanning sequence of a vector space V is called a *basis* of V.

The next result shows why bases are useful and leads to the notion of *coordinates* with respect to a basis which will be introduced in Section 2.3.

PROPOSITION 2.17. *Let V be a vector space over F. A sequence S in V is a basis of V if and only if every vector of V has a unique expression as a linear combination of S.*

PROOF. This follows from Proposition 2.13 together with the fact that a basis is a spanning sequence. □

THEOREM 2.18. *If S is a spanning sequence of V, then there is a subsequence of S which is a basis. In particular, if V is finite dimensional then it has a finite basis.*

PROOF. The first part follows from Proposition 2.14. For the second part, start with a finite spanning sequence then find a subsequence which is a basis, and note that this sequence is finite. □

EXAMPLE 2.19. Consider the vector space $V = \mathbb{R}^3$ over $\mathbb{R}$. Show that the sequence

$$S\colon (1,0,1), (1,1,1), (1,2,1), (1,-1,2)$$

is a spanning sequence of V and find a subsequence which is a basis.

SOLUTION. The following method for the first part uses the general approach to solving systems of linear equations. Given any numbers $h_1, h_2, h_3 \in \mathbb{R}$, consider the equation

$$x_1(1,0,1) + x_2(1,1,1) + x_3(1,2,1) + x_4(1,-1,2) = (h_1, h_2, h_3) = \mathbf{h}.$$

This is equivalent to the system

$$\left\{\begin{array}{ccccccccc} x_1 & + & x_2 & + & x_3 & + & x_4 & = & h_1 \\ & & x_2 & + & 2x_3 & - & x_4 & = & h_2 \\ x_1 & + & x_2 & + & x_3 & + & 2x_4 & = & h_3 \end{array}\right\}$$

which has augmented matrix

$$[A \mid \mathbf{h}] = \begin{bmatrix} 1 & 1 & 1 & 1 & h_1 \\ 0 & 1 & 2 & -1 & h_2 \\ 1 & 1 & 1 & 2 & h_3 \end{bmatrix}.$$

Performing elementary row operations we find that

$$[A \mid \mathbf{h}] \underset{R_3 \to R_3 - R_1}{\sim} \begin{bmatrix} 1 & 1 & 1 & 1 & h_1 \\ 0 & 1 & 2 & -1 & h_2 \\ 0 & 0 & 0 & 1 & h_3 - h_1 \end{bmatrix} \underset{R_1 \to R_1 - R_2}{\sim} \begin{bmatrix} 1 & 0 & -1 & 2 & h_1 - h_2 \\ 0 & 1 & 2 & -1 & h_2 \\ 0 & 0 & 0 & 1 & h_3 - h_1 \end{bmatrix}$$
$$\underset{\substack{R_1 \to R_1 - 2R_3 \\ R_2 \to R_2 + R_3}}{\sim} \begin{bmatrix} 1 & 0 & -1 & 0 & 3h_1 - h_2 - 2h_3 \\ 0 & 1 & 2 & 0 & -h_1 + h_2 + h_3 \\ 0 & 0 & 0 & 1 & -h_1 + h_3 \end{bmatrix},$$

where the last matrix is reduced echelon. So the system is consistent with general solution

$$x_1 = s + 3h_1 - h_2 - 2h_3, \quad x_2 = -2s - h_1 + h_2 + h_3, \quad x_3 = s, \quad x_4 = -h_1 + h_3 \qquad (s \in \mathbb{R}).$$

So every element of $\mathbb{R}^3$ can indeed be expressed as a linear combination of S.

Notice that when $\mathbf{h} = \mathbf{0}$ we have the solution

$$x_1 = s, \quad x_2 = -2s, \quad x_3 = s, \quad x_4 = 0 \qquad (s \in \mathbb{R}).$$

Thus shows that S is linearly dependent. Taking for example $s = 1$, we see that

$$(1, 0, 1) - 2(1, 1, 1) + (1, 2, 1) + 0(1, -1, 2) = \mathbf{0},$$

hence

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

so the sequence $T\colon (1, 0, 1), (1, 1, 1), (1, -1, 2)$ also spans $\mathbb{R}^3$. Notice that there is exactly one solution of the above system of equations for $\mathbf{h} = \mathbf{0}$ with $s = 0$, so T is linearly independent and therefore it is a basis for $\mathbb{R}^3$. $\square$

The approach of the last example leads to a general method. We state it over any field F since the method works in general. For this we need to recall the idea of a reduced echelon matrix and the way this is used to determine the general solution of a system of linear equations.

THEOREM 2.20. *Suppose that* $\mathbf{a}_1, \ldots, \mathbf{a}_n$ *is a sequence of vectors in* F^m. *Let* $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ *be the* $m \times n$ *matrix with* $\mathbf{a}_j$ *as its* j*-th column and let* A' *be its reduced echelon matrix.*

(a) $\mathbf{a}_1, \ldots, \mathbf{a}_n$ *is a spanning sequence of* F^m *if and only if the number of non-zero rows of* A' *is* m.

(b) *If the* j*-th column of* A' *does not contain a pivot, then the corresponding vector* $\mathbf{a}_j$ *is expressible as a linear combination of those associated with columns which do contain a pivot. Such a linear combination can found by setting the free variable for the* j*-th column equal to* 1 *and all the other free variables equal to* 0.

(c) *The vectors* $\mathbf{a}_i$ *for those* i *where the* i*-th column of* A' *contains a pivot, form a basis for* $\mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$.

Here is another kind of situation.

EXAMPLE 2.21. Consider the vector space $\mathbb{R}^3$ over $\mathbb{R}$. Find a basis for the plane W with equation

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

SOLUTION. This time we need to start with a system of one equation and its associated matrix $\begin{bmatrix} 1 & -2 & 3 \end{bmatrix}$ which is reduced echelon, hence we have the general solution

$$x = 2s - 3t, \quad y = s, \quad z = t \qquad (s, t \in \mathbb{R}).$$

So every vector (x, y, z) in W is a linear combination

$$(x, y, z) = s(2, 1, 0) + t(-3, 0, 1)$$

of the sequence $(2, 1, 0), (-3, 0, 1)$ obtained by setting $s = 1, t = 0$ and $s = 0, t = 1$ respectively. In fact this expression is unique since these two vectors are linearly independent. $\square$

EXAMPLE 2.22. Consider the vector space $\mathcal{D}(\mathbb{R})$ over $\mathbb{R}$ discussed in Example 1.11. Let $W \subseteq \mathcal{D}(\mathbb{R})$ be the set of all solutions of the differential equation

$$\frac{dy}{dx} + 3y = 0.$$

Show that W is a subspace of $\mathcal{D}(\mathbb{R})$ and find a basis for it.

Solution. It is easy to see that W is a subspace. Notice that if y_1, y_2 are two solutions then

$$\begin{aligned}\frac{d(y_1+y_2)}{dx} + 3(y_1+y_2) &= \frac{dy_1}{dx} + \frac{dy_2}{dx} + 3y_1 + 3y_2 \\ &= \left(\frac{dy_1}{dx} + 3y_1\right) + \left(\frac{dy_2}{dx} + 3y_2\right) = 0,\end{aligned}$$

hence $y_1 + y_2$ is a solution. Also, if y is a solution and $s \in \mathbb{R}$, then

$$\begin{aligned}\frac{d(sy)}{dx} + 3(sy) &= s\frac{dy}{dx} + s(3y) \\ &= s\left(\frac{dy}{dx} + 3y\right) = 0,\end{aligned}$$

so sy is a solution. Clearly the constant function 0 is a solution.

The general solution is

$$y = ae^{-3x} \qquad (a \in \mathbb{R}),$$

so the function e^{-3x} spans W. Furthermore, if for some $t \in \mathbb{R}$,

$$se^{-3x} = 0,$$

where the right hand side means the constant function, this has to hold for every $x \in \mathbb{R}$. Taking $x = 0$ we find that

$$t = te^0 = 0,$$

so $t = 0$. Hence the sequence e^{-3x} is a basis of W. □

A good exercise is to try to generalise this, for example by considering the solutions of the differential equation

$$\frac{d^2y}{dx^2} - 6y = 0.$$

There is a brief review of the solution of ordinary differential equations in Appendix A. Here is another example.

Example 2.23. Let $W \subseteq \mathcal{D}(\mathbb{R})$ be the set of all solutions of the differential equation

$$\frac{d^2y}{dx^2} + 4y = 0.$$

Show that W is a subspace of $\mathcal{D}(\mathbb{R})$ and find a basis for it.

Solution. Checking W is a subspace is similar to Example 2.22 The general solution has the form

$$y = a\cos 2x + b\sin 2x \qquad (a, b \in \mathbb{R}),$$

so the functions $\cos 2x, b\sin 2x$ span W. Suppose that for some $s, t \in \mathbb{R}$,

$$s\cos 2x + t\sin 2x = 0,$$

where the right hand side means the constant function. This has to hold whatever the value of $x \in \mathbb{R}$, so try taking values $x = 0, \pi/4$. These give the equations

$$s\cos 0 + t\sin 0 = 0, \quad s\cos \pi/2 + t\sin \pi/2 = 0,$$

i.e., $s = 0$, $t = 0$. So $\cos 2x, \sin 2x$ are linearly independent and hence $\cos 2x, \sin 2x$ is a basis. Notice that this is a finite dimensional vector space. □

EXAMPLE 2.24. Let $\mathbb{R}^\infty$ denote the set consisting of all sequences (a_n) of real numbers which satisfy $a_n = 0$ for all large enough n, *i.e.*,

$$(a_n) = (a_1, a_2, \dots, a_k, 0, 0, 0, \dots).$$

This can be viewed as a vector space over $\mathbb{R}$ with addition and scalar multiplication given by

$$(a_n) + (b_n) = (a_n + b_n),$$
$$t \cdot (c_n) = (tc_n).$$

Then the sequences

$$e_r = (0, \dots, 0, 1, 0, \dots)$$

with a single 1 in the r-th place, form a basis for $\mathbb{R}^\infty$.

SOLUTION. First note that

$$(x_n) = (x_1, x_2, \dots, x_k, 0, 0, 0, \dots)$$

can be expressed as

$$(x_n) = x_1 e_1 + \cdots + x_k e_k,$$

so the e_i span $\mathbb{R}^\infty$. If

$$t_1 e_1 + \cdots + t_k e_k = (0),$$

then

$$(t_n) = (t_1, \dots, t_k, 0, 0, 0, \dots) = (0, \dots, 0, \dots),$$

and so $t_k = 0$ for all k. Thus the e_i are also linearly independent. □

Here is another important method for constructing bases.

THEOREM 2.25. *Let V be a finite dimensional vector space and let S be a linearly independent sequence in V. Then there is a basis S' of V in which S is a subsequence. In particular, S and S' are both finite.*

PROOF. We will indicate a proof for the case where S is a finite sequence. Choose a finite spanning sequence, say $T\colon \mathbf{w}_1, \dots, \mathbf{w}_k$. By Proposition 2.14 we might as well assume that T is linearly independent.

If S does not span V then certainly S, T (the sequence obtained by joining S and T together) spans (since T does). Now choose the largest value of $\ell = 1, \dots, k$ for which the sequence

$$S_\ell = S, \mathbf{w}_1, \dots, \mathbf{w}_\ell$$

is linearly independent. Then S, T is linearly dependent and if $\ell < k$, we can write $\mathbf{w}_k$ as a linear combination of $S, \mathbf{w}_1, \dots, \mathbf{w}_{k-1}$, so this is still a spanning set for V. Similarly we can keep discarding the vectors $\mathbf{w}_{k-j}$ for $j = 0, \dots, (k - \ell - 1)$ and at each stage we have a spanning sequence

$$S, \mathbf{w}_1, \dots, \mathbf{w}_{k-j-1}.$$

Eventually we are left with just S_ℓ which is still a spanning set for V as well as being linearly independent, and this is the required basis. □

THEOREM 2.26. *Let V be a finite dimensional vector space and suppose that it has a basis with n elements. Then any other basis is also finite and has n elements.*

This result allows us to make the following important definition.

DEFINITION 2.27. Let V be a finite dimensional vector space. Then the number of elements in any basis of V is finite and is called the *dimension* of the vector space V and is denoted $\dim_F V$ or just $\dim V$ if the field is clear.

EXAMPLE 2.28. Let F be any field. Then the vector space F^n over F has dimension n, *i.e.*, $\dim_F F^n = n$.

PROOF. For each $k = 1, 2, \dots, n$ let $\mathbf{e}_k = (0, \dots, 0, 1, 0, \dots, 0)$ (with the 1 in the k-th place). Then the sequence $\mathbf{e}_1, \dots, \mathbf{e}_n$ spans F^n since for any vector $(x_1, \dots, x_n) \in F^n$,

$$(x_1, \dots, x_n) = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n.$$

Also, if the right hand side is $\mathbf{0}$, then $(x_1, \dots, x_n) = (0, \dots, 0)$ and so $x_1 = \cdots = x_n = 0$, showing that this sequence is linearly independent. Therefore, $\mathbf{e}_1, \dots, \mathbf{e}_n$ is a basis for F^n and $\dim_F F^n = n$.

The basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ of F^n is often called the *standard basis* of F^n. □

PROOF OF THEOREM 2.26. Let $B\colon \mathbf{b}_1, \dots, \mathbf{b}_n$ be a basis and suppose that $S\colon \mathbf{w}_1, \dots$ is another basis for V. We will assume that S is finite, say $S\colon \mathbf{w}_1, \dots, \mathbf{w}_m$, the more general case is similar.

The sequence $\mathbf{w}_1, \mathbf{b}_1, \dots, \mathbf{b}_n$ must be linearly dependent since B spans V. As B is linearly independent, there is a number $k_1 = 1, \dots, n$ for which

$$\mathbf{b}_{k_1} = s_0\mathbf{w}_1 + s_1\mathbf{b}_1 + \cdots + s_{k_1-1}\mathbf{b}_{k_1-1}$$

for some scalars $s_0, s_1 \dots, s_{k_1-1}$. By reordering B, we can assume that $k_1 = n$, so

$$\mathbf{b}_n = s_0\mathbf{w}_1 + s_1\mathbf{b}_1 + \cdots + s_{n-1}\mathbf{b}_{n-1},$$

where $s_0 \neq 0$ since B is linearly independent. Thus the sequence $B_1\colon \mathbf{w}_1, \mathbf{b}_1, \dots, \mathbf{b}_{n-1}$ spans V. If it were linearly dependent then there would be an equation of the form

$$t_0\mathbf{w}_1 + t_1\mathbf{b}_1 + \cdots + t_{n-1}\mathbf{b}_{n-1} = \mathbf{0}$$

in which not all the t_i are zero; in fact $t_0 \neq 0$ since B is linearly independent. From this we obtain a non-trivial linear combination of the form

$$t_1'\mathbf{b}_1 + \cdots + t_n'\mathbf{b}_n = \mathbf{0},$$

which is impossible. Thus B_1 is a basis.

Now we repeat this argument with $\mathbf{w}_2$, $\mathbf{w}_3$, and so on, at each stage forming (possibly after reordering and renumbering the $\mathbf{b}_i$'s) a sequence $B_k\colon \mathbf{w}_1, \dots, \mathbf{w}_k, \mathbf{b}_1, \dots, \mathbf{b}_{n-k}$ which is linearly independent and a spanning sequence for V, *i.e.*, it is a basis.

If the length of S were finite (say equal to ℓ) and *less* than the length of B (*i.e.*, $\ell < n$), then the sequence $B_\ell\colon \mathbf{w}_1, \dots, \mathbf{w}_\ell, \mathbf{b}_1, \dots, \mathbf{b}_{n-\ell}$ would be a basis. But then $B_\ell = S$ and so we could not have any $\mathbf{b}_i$ at the right hand end since S is a spanning sequence, so B_ℓ would be linearly dependent. Thus we must be able to find at least n terms of S, and so eventually we obtain the sequence $B_n\colon \mathbf{w}_1, \dots, \mathbf{w}_n$ which is a basis. This shows that S must have finite length which equals the length of B, *i.e.*, any two bases have finite and equal length . □

The next result is obtained from our earlier results.

PROPOSITION 2.29. *Let V be a finite dimensional vector space of dimension n. Suppose that $R\colon \mathbf{v}_1, \dots, \mathbf{v}_r$ is a linearly independent sequence in V and that $S\colon \mathbf{w}_1, \dots, \mathbf{w}_s$ is a spanning sequence in V. Then the following inequalities are satisfied:*

$$r \leqslant n \leqslant s.$$

Furthermore, R is a basis if and only if $r = n$, while S is a basis if and only if $s = n$.

THEOREM 2.30. *Let V be a finite dimensional vector space and suppose that $W \subseteq V$ is a vector subspace. Then W is finite dimensional and*

$$\dim W \leqslant \dim V,$$

with equality precisely when $W = V$.

PROOF. Choose a basis S for W. Then by Theorem 2.25, this extends to a basis S' for and these are both finite sets with $\dim W$ and $\dim V$ elements respectively. As S is subsequence of S', we have

$$\dim W \leqslant \dim V,$$

with equality exactly when $S = S'$, which can only happen when $W = V$. □

A similar inequality holds when the vector space V is infinite dimensional.

EXAMPLE 2.31. Recall from Example 2.23 the infinite dimensional real vector space $\mathcal{D}(\mathbb{R})$ and the subspace $W \subseteq \mathcal{D}(\mathbb{R})$. Then $\dim W = 2$ since the vectors $\cos 2x, \sin 2x$ were shown to form a basis.

The next result gives are some other characterisations of bases in a vector space V. Let X and Y be sets; then X is a *proper subset* of Y (indicated by $X \subsetneq Y$) if $X \subseteq Y$ and $X \neq Y$.

PROPOSITION 2.32. *Let S be a sequence in V.*

- *Let S be a spanning sequence of V. Then S is a basis if and only if it is a minimal spanning sequence,* i.e., *no proper subsequence of S is a spanning sequence.*
- *Let S be a linearly independent sequence. Then is a basis if and only if it is a maximal linearly independent set,* i.e., *no sequence in V containing S as a proper subsequence is linearly independent.*

PROOF. Again this follows from things we already know. □

EXAMPLE 2.33. Recall Example 1.8: given a field of scalars F, for each pair $m, n \geqslant 1$ we have the vector space over F consisting of all $m \times n$ matrices with entries in F, namely $\mathrm{M}_{m\times n}(F)$.

For $r = 1, \dots, m$ and $s = 1, \dots, n$ let E^{rs} be the $m \times n$ matrix all of whose entries is 0 except for the (r, s) entry which is 1.

Now any matrix $[a_{ij}] \in \mathrm{M}_{m\times n}(F)$ can be expressed in the form

$$\sum_{r=1}^{m} \sum_{s=1}^{n} a_{rs} E^{rs},$$

so the matrices E^{rs} span $\mathrm{M}_{m\times n}(F)$. But the only way that the expression on the right hand side can be the zero matrix is if $[a_{ij}] = O_{m\times n}$, *i.e.*, if $a_{ij} = 0$ for every entry a_{ij}. Thus the E^{rs} form a basis and so

$$\dim_F \mathrm{M}_{m\times n}(F) = mn.$$

In particular,

$$\dim_F \mathrm{M}_n(F) = n^2$$

and

$$\dim_F \mathrm{M}_{n\times 1}(F) = n = \dim_F \mathrm{M}_{1\times n}(F).$$

The next result provides another useful way to find a basis for a subspace of F^n.

Proposition 2.34. *Suppose that $S\colon \mathbf{w}_1, \ldots, \mathbf{w}_m$ is a sequence of vectors in F^n and that $W = \mathrm{Span}(S)$. Arrange the coordinates of each $\mathbf{w}_r$ as the r-th row of an $m \times n$ matrix A. Let A' be the reduced echelon form of A, and use the coordinates of the r-th row of A' to form the vector $\mathbf{w}'_r$. If A' has k non-zero rows, then $S'\colon \mathbf{w}'_1, \ldots, \mathbf{w}'_k$ is a basis for W.*

Proof. To see this, notice that we can recover A from A' by reversing the elementary row operations required to obtain A' from A, hence the rows of A' are all in $\mathrm{Span}(A)$ and vice versa. This shows that $\mathrm{Span}(A') = \mathrm{Span}(A) = W$. Furthermore, it is easy to see that the non-zero rows of A' are linearly independent. □

Example 2.35. Consider the real vector space $\mathbb{R}^4$ and let W be the subspace spanned by the vectors $(1, 2, 3, 0), (2, 0, 1, -2), (0, 1, 0, 4), (2, 1, 1, 2)$. Find a basis for W.

Solution. Let

$$A = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & 0 & 1 & -2 \\ 0 & 1 & 0 & 4 \\ 2 & 1 & 1 & 2 \end{bmatrix}.$$

Then

$$A \underset{\substack{R_2 \to R_2 - 2R_1 \\ R_4 \to R_4 - 2R_1}}{\sim} \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & -4 & -5 & -2 \\ 0 & 1 & 0 & 4 \\ 0 & -3 & -5 & 2 \end{bmatrix} \underset{R_2 \leftrightarrow R_3}{\sim} \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 0 & 4 \\ 0 & -4 & -5 & -2 \\ 0 & -3 & -5 & 2 \end{bmatrix}$$

$$\underset{\substack{R_1 \to R_1 - 2R_2 \\ R_3 \to R_3 + 4R_2 \\ R_4 \to R_4 + 3R_2}}{\sim} \begin{bmatrix} 1 & 0 & 3 & -8 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & -5 & 14 \\ 0 & 0 & -5 & 14 \end{bmatrix} \underset{R_4 \to R_4 - R_3}{\sim} \begin{bmatrix} 1 & 0 & 3 & -8 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & -5 & 14 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\underset{R_3 \to (-1/5)R_4}{\sim} \begin{bmatrix} 1 & 0 & 3 & -8 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -14/5 \\ 0 & 0 & 0 & 0 \end{bmatrix} \underset{R_1 \to R_1 - 3R_3}{\sim} \begin{bmatrix} 1 & 0 & 0 & 2/5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -14/5 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since the last matrix is reduced echelon, the sequence $(1, 0, 0, 2/5), (0, 1, 0, 0), (0, 0, 1, -14/5)$ is a basis of W. An alternative basis is $(5, 0, 0, 2), (0, 1, 0, 0), (0, 0, 5, -14)$. □

2.3. Coordinates with respect to bases

Let V be a finite dimensional vector space over a field F and let $n = \dim V$. Suppose that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis of V. Then by Proposition 2.17, every vector $\mathbf{x} \in V$ has a unique expression of the form

$$\mathbf{x} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n,$$

where $x_1, \ldots, x_n \in F$ are the *coordinates* of $\mathbf{x}$ with respect to this basis. Of course, given the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, each vector $\mathbf{x}$ is determined by the *coordinate vector* $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ with respect to this basis. If $\mathbf{v}'_1, \ldots, \mathbf{v}'_n$ is another basis and if the coordinates of $\mathbf{x}$ with respect it are $x'_1, \ldots, x'_n \in F$, then

$$\mathbf{x} = x'_1\mathbf{v}'_1 + \cdots + x'_n\mathbf{v}'_n,$$

If we write

$$\mathbf{v}'_k = a_{1k}\mathbf{v}_1 + \cdots + a_{nk}\mathbf{v}_n$$

for $a_{rs} \in F$, we obtain

$$\begin{aligned} \mathbf{x} &= x'_1 \sum_{r=1}^{n} a_{r1}\mathbf{v}_r + \cdots + x'_n \sum_{r=1}^{n} a_{rn}\mathbf{v}_r, \\ &= (\sum_{s=1}^{n} a_{1s}x'_s)\mathbf{v}_1 + \cdots + (\sum_{s=1}^{n} a_{ns}x'_s)\mathbf{v}_n. \end{aligned}$$

Thus we obtain the equation

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = (\sum_{s=1}^{n} a_{1s}x'_s)\mathbf{v}_1 + \cdots + (\sum_{s=1}^{n} a_{ns}x'_s)\mathbf{v}_n.$$

and so for each $r = 1, \ldots, n$ we have

$$x_r = \sum_{s=1}^{n} a_{rs}x'_s. \tag{2.1a}$$

This is more efficiently expressed in the matrix form

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix}. \tag{2.1b}$$

Clearly we could also express the coordinates x'_r in terms of the x_s. Thus implies that the matrix $[a_{ij}]$ is invertible and in fact we have

$$\begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}. \tag{2.1c}$$

2.4. Sums of subspaces

Let V be a vector space over a field F. In this section we will study another way to combine subspaces of a vector space by taking a kind of sum.

Definition 2.36. Let V', V'' be two vector subspaces of V. Then their *sum* is the subset

$$V' + V'' = \{\mathbf{v}' + \mathbf{v}'' : \mathbf{v}' \in V', \mathbf{v}'' \in V''\}.$$

More generally, for subspaces $V_1, \ldots, V_k$ of V,

$$V_1 + \cdots + V_k = \{\mathbf{v}_1 + \cdots + \mathbf{v}_k : \text{for } j = 1, \ldots, k, \ \mathbf{v}_j \in V_j\}.$$

Notice that $V_1 + \cdots + V_k$ contains each V_j as a subset.

PROPOSITION 2.37. *For subspaces $V_1, \ldots, V_k$ of V, the sum $V_1 + \cdots + V_k$ is also a subspace of V which is the smallest subspace of V containing each of the V_j as a subset.*

PROOF. Suppose that $\mathbf{x}, \mathbf{y} \in V_1 + \cdots + V_k$. Then for $j = 1, \ldots, k$ there are vectors $\mathbf{x}_j, \mathbf{y}_j \in V_j$ for which

$$\mathbf{x} = \mathbf{x}_1 + \cdots + \mathbf{x}_k, \quad \mathbf{y} = \mathbf{y}_1 + \cdots + \mathbf{y}_k.$$

We have

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (\mathbf{x}_1 + \cdots + \mathbf{x}_k) + (\mathbf{y}_1 + \cdots + \mathbf{y}_k) \\ &= (\mathbf{x}_1 + \mathbf{y}_1) + \cdots + (\mathbf{x}_k + \mathbf{y}_k). \end{aligned}$$

Since each V_j is a subspace it is closed under addition, hence $\mathbf{x}_j + \mathbf{y}_j \in V_j$, therefore $\mathbf{x} + \mathbf{y} \in V_1 + \cdots + V_k$. Also, for $s \in F$,

$$s\mathbf{x} = s(\mathbf{x}_1 + \cdots + \mathbf{x}_k) = s\mathbf{x}_1 + \cdots + s\mathbf{x}_k,$$

and each V_j is closed under multiplication by scalars, hence $s\mathbf{x} \in V_1 + \cdots + V_k$. Therefore $V_1 + \cdots + V_k$ is a subspace of V.

Any subspace $W \subseteq V$ which contains each of the V_j as a subset has to also contain every element of the form

$$\mathbf{v}_1 + \cdots + \mathbf{v}_k \quad (\mathbf{v}_j \in V_j)$$

and hence $V_1 + \cdots + V_k \subseteq W$. Therefore $V_1 + \cdots + V_k$ is the smallest such subspace. □

DEFINITION 2.38. Let V', V'' be two vector subspaces of V. Then $V' + V''$ is a *direct sum* if

$$V' \cap V'' = \{\mathbf{0}\}.$$

More generally, for subspaces $V_1, \ldots, V_k$ of V, $V_1 + \cdots + V_k$ is a *direct sum* if for each $r = 1, \ldots, k$,

$$V_r \cap (V_1 + \cdots + V_{r-1} + V_{r+1} + \cdots + V_k) = \{\mathbf{0}\}.$$

Such direct sums are sometimes denoted $V' \oplus V''$ and $V_1 \oplus \cdots \oplus V_k$.

THEOREM 2.39. *Let V', V'' be subspaces of V for which $V' + V''$ is a direct sum $V' \oplus V''$. Then every vector $\mathbf{v} \in V' \oplus V''$ has a unique expression*

$$\mathbf{v} = \mathbf{v}' + \mathbf{v}'' \quad (\mathbf{v}' \in V', \ \mathbf{v}'' \in V'').$$

More generally, if $V_1, \ldots, V_k$ are subspaces of V for which $V_1 + \cdots + V_k$ is a direct sum $V_1 \oplus \cdots \oplus V_k$, then every vector $\mathbf{v} \in V_1 \oplus \cdots \oplus V_k$ has a unique expression

$$\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_k \quad (\mathbf{v}_j \in V_j, \ j = 1, \ldots, k).$$

PROOF. We will give the proof for the first case, the general case is more involved but similar.

Let $\mathbf{v} \in V' \oplus V''$. Then certainly there are vectors $\mathbf{v}' \in V'$, $\mathbf{v}'' \in V''$ for which $\mathbf{v} = \mathbf{v}' + \mathbf{v}''$. Suppose there is a second pair of vectors $\mathbf{w}' \in V'$, $\mathbf{w}'' \in V''$ for which $\mathbf{v} = \mathbf{w}' + \mathbf{w}''$. Then

$$(\mathbf{w}' - \mathbf{v}') + (\mathbf{w}'' - \mathbf{v}'') = \mathbf{0},$$

and $\mathbf{w}' - \mathbf{v}' \in V'$, $\mathbf{w}'' - \mathbf{v}'' \in V''$. But this means that

$$\mathbf{w}' - \mathbf{v}' = \mathbf{v}'' - \mathbf{w}''$$

where the left hand side is in V' and the right hand side is in V''. But the equality here means that each of these terms is in $V' \cap V'' = \{\mathbf{0}\}$, hence $\mathbf{w}' = \mathbf{v}'$ and $\mathbf{w}'' = \mathbf{v}''$. Thus such an expression for $\mathbf{v}$ must be unique. □

Proposition 2.40. *Let V', V'' be subspaces of V. If these are finite dimensional, then so is $V' + V''$ and*

$$\dim(V' + V'') = \dim V' + \dim V'' - \dim(V' \cap V'').$$

Proof. Notice that $V' \cap V''$ is a subspace of each of V' and V'' and of $V' + V''$. Choose a basis for $V' \cap V''$, say $\mathbf{v}_1, \dots, \mathbf{v}_a$. This extends to bases for V' and V'', say $\mathbf{v}_1, \dots, \mathbf{v}_a, \mathbf{v}'_1, \dots, \mathbf{v}'_b$ and $\mathbf{v}_1, \dots, \mathbf{v}_a, \mathbf{v}''_1, \dots, \mathbf{v}''_c$. Every element of $V' + V''$ is a linear combination of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_a, \mathbf{v}'_1, \dots, \mathbf{v}'_b, \mathbf{v}''_1, \dots, \mathbf{v}''_c$, we will show that they are linearly independent.

Suppose that for scalars $s_1, \dots, s_a, s'_1, \dots, s'_b, s''_1, \dots, s''_c$, the equation

$$s_1\mathbf{v}_1 + \cdots + s_a\mathbf{v}_a + s'_1\mathbf{v}'_1 + \cdots + s'_b\mathbf{v}'_b + s''_1\mathbf{v}''_1 + \cdots + s''_c\mathbf{v}''_c = \mathbf{0}$$

holds. Then we have

$$s''_1\mathbf{v}''_1 + \cdots + s''_c\mathbf{v}''_c = -(s_1\mathbf{v}_1 + \cdots + s_a\mathbf{v}_a + s'_1\mathbf{v}'_1 + \cdots + s'_b\mathbf{v}'_b) \in V',$$

but the left hand side is also in V'', hence both sides are in $V' \cap V''$. By our choice of basis for $V' \cap V''$, we find that $s''_1\mathbf{v}''_1 + \cdots + s''_c\mathbf{v}''_c$ is a linear combination of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_a$ and since $\mathbf{v}_1, \dots, \mathbf{v}_a, \mathbf{v}''_1, \dots, \mathbf{v}''_c$ is linearly independent the only way that can happen is if all the coefficients are zero, hence $s''_1 = \cdots = s''_c = 0$. A similar argument also shows that $s'_1 = \cdots = s'_b = 0$. We are left with the equation

$$s_1\mathbf{v}_1 + \cdots + s_a\mathbf{v}_a = \mathbf{0}$$

which also implies that $s_1 = \cdots = s_a = 0$ since the $\mathbf{v}_i$ form a basis for $V' \cap V''$.

Now we have

$$\dim(V' \cap V'') = a, \quad \dim V' = a + b, \quad \dim V'' = a + c,$$

and

$$\dim(V' + V'') = a + b + c = \dim V' + \dim V'' - \dim(V' \cap V''). \qquad \square$$

Definition 2.41. Let V be a vector space and $W \subseteq V$ be a subspace. A subspace $W' \subseteq V$ for which $W + W' = V$ and $W \cap W' = \{\mathbf{0}\}$ is called a *linear complement* of W in V. For such a linear complement we have $V = W \oplus W'$.

We remark that although linear complements always exist they are by no means unique! The next result summarises the situation.

Proposition 2.42. *Let $W \subseteq V$ be subspace.*

(a) *There is a linear complement $W' \subseteq V$ of W in V.*

(b) *If V is finite dimensional, then any subspace $W'' \subseteq V$ satisfying the conditions*

- $W \cap W'' = \{\mathbf{0}\}$,
- $\dim W'' = \dim V - \dim W$

is a linear complement of W in V.

(c) *Any subspace $W''' \subseteq V$ satisfying the conditions*

- $W + W''' = V$,
- $\dim W''' = \dim V - \dim W$

is a linear complement of W in V.

PROOF.
(a) We show how to find a linear complement in the case where V is finite dimensional with $n = \dim V$. By Theorem 2.30, we know that $\dim W \leqslant \dim V$. So let W have a basis $\mathbf{w}_1, \ldots, \mathbf{w}_r$, where $r = \dim W$. Now by Theorem 2.25 there is an extension of this to a basis of V, say

$$\mathbf{w}_1, \ldots, \mathbf{w}_r, \mathbf{w}'_1, \ldots, \mathbf{w}'_{n-r}.$$

Now let $W' = \operatorname{Span}(\mathbf{w}'_1, \ldots, \mathbf{w}'_{n-r})$. Then by Lemma 2.6, W' is a subspace of V and it has $\mathbf{w}'_1, \ldots, \mathbf{w}'_{n-r}$ as a basis, so $\dim W' = n - r$.

Now it is clear that $W + W' = V$. Suppose that $\mathbf{v} \in W \cap W'$. Then there are expressions

$$\mathbf{v} = s_1\mathbf{w}_1 + \cdots + s_r\mathbf{w}_r,$$
$$\mathbf{v} = s'_1\mathbf{w}'_1 + \cdots + s'_{n-r}\mathbf{w}'_{n-r},$$

hence there is an equation

$$s_1\mathbf{w}_1 + \cdots + s_r\mathbf{w}_r + (-s'_1)\mathbf{w}'_1 + \cdots + (-s'_{n-r})\mathbf{w}'_{n-r} = \mathbf{0}.$$

But the vectors $\mathbf{w}_i, \mathbf{w}'_j$ form a basis so are linearly independent. Therefore

$$s_1 = \cdots = s_r = s'_1 = \cdots = s'_{n-r} = 0.$$

This means that $\mathbf{v} = \mathbf{0}$, hence $W \cap W' = \{\mathbf{0}\}$ and the sum $W + W'$ is direct. So we have shown that W' is a linear complement of W in V.

(b) and (c) can be verified in similar fashion. $\square$

EXAMPLE 2.43. Consider the vector space $\mathbb{R}^2$ over $\mathbb{R}$. If $V' \subseteq \mathbb{R}^2$ and $V'' \subseteq \mathbb{R}^2$ are two distinct lines through the origin, show that $V' + V''$ is a direct sum and that $V' \oplus V'' = \mathbb{R}^2$.

SOLUTION. As these lines are distinct and $V' \cap V''$ is subspace of each of V' and V'', we have

$$\dim(V' \cap V'') < \dim V' = \dim V'' = 1,$$

hence $\dim(V' \cap V'') = 0$. This means that $V' \cap V'' = \{\mathbf{0}\}$ which is geometrically obvious. Now notice that

$$\dim(V' + V'') = \dim V' + \dim V'' - 0 = 1 + 1 = 2,$$

hence $\dim(V' + V'') = \dim \mathbb{R}^2$ and so $V' + V'' = \mathbb{R}^2$. $\square$

EXAMPLE 2.44. Consider the vector space $\mathbb{R}^3$ over $\mathbb{R}$. If $V' \subseteq \mathbb{R}^3$ is a line through the origin and $V'' \subseteq \mathbb{R}^2$ is a plane through the origin which does not contain V', show that $V' + V''$ is a direct sum and that $V' \oplus V'' = \mathbb{R}^3$.

SOLUTION. $V' \cap V''$ is subspace of each of V' and V'', so we have

$$\dim(V' \cap V'') \leqslant \dim V' = 1 < 2 = \dim V''.$$

In fact the inequality must be strict since otherwise $V' \subseteq V''$, hence $\dim(V' \cap V'') = 0$. This means that $V' \cap V'' = \{\mathbf{0}\}$ which is also geometrically obvious. Now notice that

$$\dim(V' + V'') = \dim V' + \dim V'' - 0 = 1 + 2 = 3,$$

therefore $\dim(V' + V'') = \dim \mathbb{R}^3$ and so $V' + V'' = \mathbb{R}^3$. $\square$

EXAMPLE 2.45. Consider the vector space $\mathbb{R}^4$ over $\mathbb{R}$ and the subspaces

$$V' = \{(s+t, s, -s, -t) : s, t \in \mathbb{R}\}, \quad V'' = \{(u, 2v, -v, -u) : u, v \in \mathbb{R}\}.$$

Show that $\dim(V' + V'') = 3$.

SOLUTION. V' is spanned by the vectors $(1, 1, -1, 0), (1, 0, 0, -1)$, which are linearly independent and hence form a basis. V'' is spanned by $(1, 0, 0, -1), (0, 2, -1, 0)$ which are linearly independent and hence form a basis. Thus we find that

$$\dim V' = 2 = \dim V''.$$

Now suppose that $(w, x, y, z) \in V' \cap V''$. Then there are $s, t, u, v \in \mathbb{R}$ for which

$$(w, x, y, z) = (s+t, s, -s, -t) = (u, 2v, -v, -u).$$

By comparing coefficients, this leads to the system of four linear equations

$$\left\{\begin{array}{rrrrrrrrr} & s & + & t & - & u & & & = 0 \\ & s & & & & & - & 2v & = 0 \\ - & s & & & & & + & v & = 0 \\ & & - & t & + & u & & & = 0 \end{array}\right\} \tag{2.2}$$

which has associated augmented matrix

$$\begin{bmatrix} 1 & 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & -2 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix} \underset{\substack{R_2 \to R_2 - R_1 \\ R_3 \to R_3 + R_1}}{\sim} \begin{bmatrix} 1 & 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & -2 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix} \underset{R_2 \leftrightarrow R_3}{\sim} \begin{bmatrix} 1 & 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & -1 & 1 & -2 & 0 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix}$$

$$\underset{\substack{R_1 \to R_1 - R_2 \\ R_3 \to R_3 + R_2 \\ R_4 \to R_4 + R_2}}{\sim} \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \underset{R_3 \leftrightarrow R_4}{\sim} \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 \end{bmatrix}$$

$$\underset{\substack{R_1 \to R_1 + R_3 \\ R_2 \to R_2 - R_3 \\ R_4 \to R_4 + R_3}}{\sim} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

where the last matrix is reduced echelon. The general solution of the system (2.2) is

$$s = 0, \quad t = u, \quad u \in \mathbb{R}, \quad v = 0.$$

hence we have

$$V' \cap V'' = \{(u, 0, 0, -u) : u \in \mathbb{R}\},$$

and this has basis $(1, 0, 0, -1)$, hence $\dim(V' \cap V'') = 1$. This means that these two planes intersect in a line in $\mathbb{R}^4$.

Now we see that

$$\dim(V' + V'') = \dim V' + \dim V'' - \dim(V' \cap V'') = 2 + 2 - 1 = 3. \qquad \square$$

CHAPTER 3

Linear transformations

3.1. Functions

Before introducing *linear transformations*, we will review some basic ideas about *functions*, including properties of *composition* and *inverse functions* which will be required later.

Let X, Y and Z be sets.

DEFINITION 3.1. A *function* (or *mapping*) $f\colon X \longrightarrow Y$ from X to Y involves a rule which associates to each $x \in X$ a unique element $f(x) \in Y$. X is called the *domain* of f, denoted $\operatorname{dom} f$, while Y is called the *codomain* of f, denoted $\operatorname{codom} f$. Sometimes the rule is indicated by writing $x \mapsto f(x)$.

Given functions $f\colon X \longrightarrow Y$ and $g\colon Y \longrightarrow Z$, we can form their *composition* $g \circ f\colon X \longrightarrow Z$ which has the rule

$$g \circ f(x) = g(f(x)).$$

Note the order of composition here!

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[rr, bend left, "g\circ f"] & Y \arrow[r, "g"] & Z \end{tikzcd}$$

We often write gf for $g \circ f$ when no confusion will arise. If X, Y, Z are different it may not be possible to define $f \circ g$ since $\operatorname{dom} f = X$ may not be the same as $\operatorname{codom} g = Z$.

EXAMPLE 3.2. For any set X, the identity function $\mathrm{Id}_X \longrightarrow X$ has rule

$$x \mapsto \mathrm{Id}_X(x) = x.$$

EXAMPLE 3.3. Let $X = Y = \mathbb{R}$. Then the following rules define functions $\mathbb{R} \longrightarrow \mathbb{R}$:

$$x \mapsto x + 1, \quad x \mapsto x^2, \quad x \mapsto \frac{1}{x^2+1}, \quad x \mapsto \sin x, x \mapsto e^{5x^3}.$$

EXAMPLE 3.4. Let $X = Y = \mathbb{R}^+$, the set of all positive real numbers. Then the following rules define two functions $\mathbb{R}^+ \longrightarrow \mathbb{R}$:

$$x \mapsto \sqrt{x}, \quad x \mapsto -\sqrt{x}.$$

EXAMPLE 3.5. Let X and Y be sets and suppose that $w \in Y$ is an element. The *constant function $c_w\colon X \longrightarrow Y$ taking value w* has the rule

$$x \mapsto c_w(x) = w.$$

For example, if $X = Y = \mathbb{R}$ then c_0 is the function which returns the value $c_0(x) = 0$ for every real number $x \in \mathbb{R}$.

DEFINITION 3.6. A function $f\colon X \longrightarrow Y$ is

- *injective* (or an *injection* or *one-to-one*) if for $x_1, x_2 \in X$,

$$f(x_1) = f(x_2) \text{ implies } x_1 = x_2,$$

- *surjective* (or a *surjection* or *onto*) if for every $y \in Y$ there is an $x \in X$ for which $y = f(x)$,
- *bijective* (or *injective* or a *one-to-one correspondence*) if it is both injective and surjective.

We will use the following basic fact.

PROPOSITION 3.7. *The function $f \colon X \longrightarrow Y$ is a bijection if and only if there is an inverse function $Y \longrightarrow X$ which is the unique function $h \colon Y \longrightarrow X$ satisfying*

$$h \circ f = \mathrm{Id}_X, \quad f \circ h = \mathrm{Id}_Y .$$

If such an inverse exists it is usual to denote it by $f^{-1} \colon Y \longrightarrow X$, and then we have

$$f^{-1} \circ f = \mathrm{Id}_X, \quad f \circ f^{-1} = \mathrm{Id}_Y .$$

Later we will see examples of all these notions in the context of vector spaces.

3.2. Linear transformations

In this section, let F be a field of scalars.

DEFINITION 3.8. Let V and W be two vector spaces over F and let $f \colon V \longrightarrow W$ be a function. Then f is called a *linear transformation* or *linear mapping* if it satisfies the two conditions

- for all vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$,

$$f(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1) + f(\mathbf{v}_2),$$

- for all vectors $\mathbf{v} \in V$ and scalars $t \in F$,

$$f(t\mathbf{v}) = tf(\mathbf{v}).$$

These two conditions together are equivalent to the single condition

- for all vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$, and scalars $t_1, t_2 \in F$,

$$f(t_1\mathbf{v}_1 + t_2\mathbf{v}_1) = t_1 f(\mathbf{v}_1) + t_2 f(\mathbf{v}_1).$$

REMARK 3.9. For a linear transformation $f \colon V \longrightarrow W$, we always have $f(\mathbf{0}) = \mathbf{0}$, where the left hand $\mathbf{0}$ means the zero vector in V and the right hand $\mathbf{0}$ means the zero vector in W.

The next example introduces an important kind of linear transformation associated with a matrix.

EXAMPLE 3.10. Let $A \in \mathrm{M}_{m\times n}$ be an $m \times n$ matrix. Define $f_A \colon F^n \longrightarrow F^m$ by the rule

$$f_A(\mathbf{x}) = A\mathbf{x}.$$

Then for $s, t \in F$ and $\mathbf{u}, \mathbf{v} \in F^n$,

$$f_A(s\mathbf{u} + t\mathbf{v}) = A(s\mathbf{u} + t\mathbf{v}) = sA\mathbf{u} + tA\mathbf{v} = sf_A(\mathbf{u}) + tf_A(\mathbf{v}).$$

So f_A is a linear transformation.

Observe that the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ (viewed as column vectors) satisfy

$$f_A(\mathbf{e}_j) = j\text{-th column of } A,$$

so

$$A = \begin{bmatrix} f_A(\mathbf{e}_1) & \cdots & f_A(\mathbf{e}_n) \end{bmatrix} = \begin{bmatrix} A\mathbf{e}_1 & \cdots & A\mathbf{e}_n \end{bmatrix}.$$

Now it easily follows that $f_A(\mathbf{e}_1), \ldots, f_A(\mathbf{e}_n)$ spans $\operatorname{Im} f_A$ since every element of the latter is a linear combination of this sequence.

THEOREM 3.11. *Let $f \colon V \longrightarrow W$ and $g \colon U \longrightarrow V$ be two linear transformations between vector spaces over F. Then the composition $f \circ g \colon U \longrightarrow W$ is also a linear transformation.*

PROOF. For $s_1, s_2 \in F$ and $\mathbf{u}_1, \mathbf{u}_2 \in V$, we have to show that

$$f \circ g(s_1\mathbf{u}_1 + s_2\mathbf{u}_2) = s_1 f \circ g(\mathbf{u}_1) + s_2 f \circ g(\mathbf{u}_2).$$

Using the fact that both f and g are linear transformations we have

$$\begin{aligned} f \circ g(s_1\mathbf{v}_1 + s_2\mathbf{v}_2) &= f\left(s_1 g(\mathbf{u}_1) + s_2 g(\mathbf{u}_2)\right) \\ &= s_1 f\left(g(\mathbf{u}_1)\right) + s_2 f\left(g(\mathbf{u}_2)\right) \\ &= s_1 f \circ g(\mathbf{u}_1) + s_2 f \circ g(\mathbf{u}_2), \end{aligned}$$

as required. □

THEOREM 3.12. *Let V and W be vector spaces over F. Suppose that $f, g \colon V \longrightarrow W$ are linear transformations and $s, t \in F$. Then the function*

$$sf + tg \colon V \longrightarrow W; \quad (sf + tg)(\mathbf{v}) = sf(\mathbf{v}) + tg(\mathbf{v})$$

is a linear transformation.

DEFINITION 3.13. Let $f \colon V \longrightarrow W$ be a linear transformation.

- The *kernel* (or *nullspace*) of f is the following subset of V:

$$\operatorname{Ker} f = \{\mathbf{v} \in V : f(\mathbf{v}) = \mathbf{0}\} \subseteq V.$$

- The *image* (or *range*) of f is the following subset of W:

$$\operatorname{Im} f = \{\mathbf{w} \in W : \text{there is an } \mathbf{v} \in V \text{ s.t. } \mathbf{w} = f(\mathbf{v})\} \subseteq W.$$

THEOREM 3.14. *Let $f \colon V \longrightarrow W$ be a linear transformation. Then*
(a) $\operatorname{Ker} f$ *is a subspace of* V,
(b) $\operatorname{Im} f$ *is a subspace of* W.

PROOF.
(a) Let $s_1, s_2 \in F$ and $\mathbf{v}_1, \mathbf{v}_2 \in \operatorname{Ker} f$. We must show that $s_1\mathbf{v}_1 + s_2\mathbf{v}_2 \in \operatorname{Ker} f$, *i.e.*, that $f(s_1\mathbf{v}_1 + s_2\mathbf{v}_2) = \mathbf{0}$. Since f is a linear transformation and $f(\mathbf{v}_1) = \mathbf{0} = f(\mathbf{v}_2)$, we have

$$\begin{aligned} f(s_1\mathbf{v}_1 + s_2\mathbf{v}_2) &= s_1 f(\mathbf{v}_1) + s_2 f(\mathbf{v}_2) \\ &= s_1\mathbf{0} + s_2\mathbf{0} = \mathbf{0}, \end{aligned}$$

hence $f(s_1\mathbf{v}_1 + s_2\mathbf{v}_2) \in \operatorname{Ker} f$. Therefore $\operatorname{Ker} f$ is a subspace of V.

Now suppose that $t_1, t_2 \in F$ and $\mathbf{w}_1, \mathbf{w}_2 \in \operatorname{Im} f$. We must show that $t_1\mathbf{w}_1 + t_2\mathbf{w}_2 \in \operatorname{Im} f$, *i.e.*, that $t_1\mathbf{w}_1 + t_2\mathbf{w}_2 = f(\mathbf{v})$ for some $\mathbf{v} \in V$. Since $\mathbf{w}_1, \mathbf{w}_2 \in \operatorname{Im} f$, there are vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ for which

$$f(\mathbf{v}_1) = \mathbf{w}_1, \quad f(\mathbf{v}_2) = \mathbf{w}_2.$$

As f is a linear transformation,

$$\begin{aligned} t_1\mathbf{w}_1 + t_2\mathbf{w}_2 &= t_1 f(\mathbf{v}_1) + t_2 f(\mathbf{v}_2) \\ &= f(t_1\mathbf{v}_1 + t_2\mathbf{v}_2), \end{aligned}$$

so we can take $\mathbf{v} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2$ to get

$$t_1\mathbf{w}_1 + t_2\mathbf{w}_2 = f(\mathbf{v}).$$

Hence $\operatorname{Im} f$ is a subspace of W. □

For linear transformations the following result holds, where we make use of notions introduced in Section 3.1.

THEOREM 3.15. *Let $f\colon V \longrightarrow W$ be a linear transformation.*
(a) *f is injective if and only if* $\operatorname{Ker} f = \{\mathbf{0}\}$.
(b) *f is surjective if and only if* $\operatorname{Im} f = W$.
(c) *f is bijective if and only if* $\operatorname{Ker} f = \{\mathbf{0}\}$ *and* $\operatorname{Im} f = W$.

PROOF.
(a) Notice that $f(\mathbf{0}) = \mathbf{0}$. Hence if f is injective then $f(\mathbf{v}) = \mathbf{0}$ implies $\mathbf{v} = \mathbf{0}$, *i.e.*, $\operatorname{Ker} f = \{\mathbf{0}\}$. We need to show the converse, *i.e.*, if $\operatorname{Ker} f = \{\mathbf{0}\}$ then f is injective.

So suppose that $\operatorname{Ker} f = \{\mathbf{0}\}$, $\mathbf{v}_1, \mathbf{v}_2 \in V$, and $f(\mathbf{v}_1) = f(\mathbf{v}_2)$. Since f is a linear transformation,

$$\begin{aligned} f(\mathbf{v}_1 - \mathbf{v}_2) &= f(\mathbf{v}_1 + (-1)\mathbf{v}_2) \\ &= f(\mathbf{v}_1) + (-1)f(\mathbf{v}_2) \\ &= f(\mathbf{v}_1) - f(\mathbf{v}_2) \\ &= \mathbf{0}, \end{aligned}$$

hence $f(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}$ and therefore $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$, *i.e.*, $\mathbf{v}_1 = \mathbf{v}_2$. This shows that f is injective.
(b) This is immediate from the definition of surjectivity.
(c) This comes by combining (a) and (b). □

THEOREM 3.16. *Let $f\colon V \longrightarrow W$ be a bijective linear transformation. Then the inverse function $f^{-1}\colon W \longrightarrow V$ is a linear transformation.*

PROOF. Let $t_1, t_2 \in F$ and $\mathbf{w}_1, \mathbf{w}_2 \in W$. We must show that

$$f^{-1}(t_1\mathbf{w}_1 + t_2\mathbf{w}_2) = t_1 f^{-1}(\mathbf{w}_1) + t_2 f^{-1}(\mathbf{w}_2).$$

The following calculation does this:

$$\begin{aligned} f\left(f^{-1}(t_1\mathbf{w}_1+t_2\mathbf{w}_2)-t_1f^{-1}(\mathbf{w}_1)+t_2f^{-1}(\mathbf{w}_2)\right) =& f\circ f^{-1}(t_1\mathbf{w}_1+t_2\mathbf{w}_2) \\ &\quad -f(t_1f^{-1}(\mathbf{w}_1)+t_2f^{-1}(\mathbf{w}_2)) \\ =& \operatorname{Id}_W(t_1\mathbf{w}_1+t_2\mathbf{w}_2) \\ &\quad -f(t_1f^{-1}(\mathbf{w}_1)+t_2f^{-1}(\mathbf{w}_2)) \\ =&(t_1\mathbf{w}_1+t_2\mathbf{w}_2) \\ &\quad -(t_1f\circ f^{-1}(\mathbf{w}_1)+t_2f\circ f^{-1}(\mathbf{w}_2)) \\ =&(t_1\mathbf{w}_1+t_2\mathbf{w}_2) \\ &\quad -(t_1\operatorname{Id}_W(\mathbf{w}_1)+t_2\operatorname{Id}_W(\mathbf{w}_2)) \\ =&(t_1\mathbf{w}_1+t_2\mathbf{w}_2)-(t_1\mathbf{w}_1+t_2\mathbf{w}_2)=\mathbf{0}. \end{aligned}$$

Since f is injective this means that

$$f^{-1}(t_1\mathbf{w}_1+t_2\mathbf{w}_2)-(t_1f^{-1}(\mathbf{w}_1)+t_2f^{-1}(\mathbf{w}_2))=\mathbf{0},$$

giving the desired formula. □

Definition 3.17. A linear transformation $f\colon V\longrightarrow W$ which is a bijective function is called an *isomorphism* from V to W, and these are said to be *isomorphic vector spaces*.

Notice that for an isomorphism $f\colon V\longrightarrow W$, the inverse $f^{-1}\colon W\longrightarrow V$ is also an isomorphism.

Definition 3.18. Let $f\colon V\longrightarrow W$ be a linear transformation. Then the *rank* of f is

$$\operatorname{rank} f=\dim\operatorname{Im} f,$$

and the *nullity* of f is

$$\operatorname{null} f=\dim\operatorname{Ker} f,$$

whenever these are finite.

Theorem 3.19 (The rank-nullity theorem). *Let $f\colon V\longrightarrow W$ be a linear transformation with V finite dimensional. Then*

$$\dim V=\operatorname{rank} f+\operatorname{null} f.$$

Proof. Start by choosing a basis for $\operatorname{Ker} f$, say $\mathbf{u}_1,\ldots,\mathbf{u}_k$. Notice that $k=\operatorname{null} f$. By Theorem 2.25, this can be extended to a basis of V, say $\mathbf{u}_1,\ldots,\mathbf{u}_k,\mathbf{u}_{k+1},\ldots,\mathbf{u}_n$, where $n=\dim V$. Now for $s_1,\ldots,s_n\in F$,

$$\begin{aligned} f(s_1\mathbf{u}_1+\cdots+s_n\mathbf{u}_n) &= s_1f(\mathbf{u}_1)+\cdots+s_nf(\mathbf{u}_n) \\ &= s_1\mathbf{0}+\cdots+s_k\mathbf{0}+s_{k+1}f(\mathbf{u}_{k+1})+\cdots+s_nf(\mathbf{u}_n) \\ &= s_{k+1}f(\mathbf{u}_{k+1})+\cdots+s_nf(\mathbf{u}_n). \end{aligned}$$

This shows that $\operatorname{Im} f$ is spanned by the vectors $f(\mathbf{u}_{k+1}),\ldots,f(\mathbf{u}_n)$. In fact, this last expression is only zero when $s_1\mathbf{u}_1+\cdots+s_n\mathbf{u}_n$ lies in $\operatorname{Ker} f$ which can only happen if $s_{k+1}=\cdots=s_n=0$. So the vectors $f(\mathbf{u}_{k+1}),\ldots,f(\mathbf{u}_n)$ are linearly independent and therefore form a basis for $\operatorname{Im} f$. Thus $\operatorname{rank} f=n-k$ and the result follows. □

We can also reformulate the results of Theorem 3.15.

Theorem 3.20. *Let* $f\colon V \longrightarrow W$ *be a linear transformation.*
(a) f *is injective if and only if* $\operatorname{null} f = 0$.
(b) f *is surjective if and only if* $\operatorname{rank} f = \dim W$.
(c) f *is bijective if and only if* $f = 0$ *and* $\operatorname{rank} f = \dim W$.

Corollary 3.21. *Let* $f\colon V \longrightarrow W$ *be a linear transformation with* V *and* W *finite dimensional vector spaces. The following hold:*
(a) *if* f *is an injection, then* $\dim_F V \leqslant \dim_F W$*;*
(b) *if* f *is a surjection, then* $\dim_F V \geqslant \dim_F W$*;*
(c) *if* f *is a bijection, then* $\dim_F V = \dim_F W$.

Proof. This makes use of the rank-nullity Theorem 3.19 and Theorem 3.20.
(a) Since $\operatorname{Im} f$ is a subspace of W, we have $\dim \operatorname{Im} f \leqslant \dim W$, hence if f is injective,

$$\dim V = \operatorname{rank} f \leqslant \dim W.$$

(b) This time, if f is surjective we have

$$\dim V \geqslant \operatorname{rank} f = \dim W.$$

Combining (a) and (b) gives (c). □

Example 3.22. Let $\mathcal{D}(\mathbb{R})$ be the vector space of differentiable real functions considered in Example 1.11.
(a) For $n \geqslant 1$, define the function

$$D^n\colon \mathcal{D}(\mathbb{R}) \longrightarrow \mathcal{D}(\mathbb{R}); \quad D^n(f) = \frac{d^n f}{dx}.$$

Show that this is a linear transformation and that

$$D^n = D \circ D^{n-1} = D^{n-1} \circ D$$

and so D^n is the n-fold composition $D \circ \cdots \circ D$. Identify the kernel and image of D^n.
(b) More generally, for real numbers $a_0, a_1, \ldots, a_n$, define the function

$$a_0 + a_1 D + \cdots + a_n D^n\colon \mathcal{D}(\mathbb{R}) \longrightarrow \mathcal{D}(\mathbb{R});$$
$$(a_0 + a_1 D + \cdots + a_n D^n)(f) = a_0 f + a_1 D(f) + \cdots + a_n D^n(f),$$

and show that this is a linear transformation and identify its kernel and image.

Solution.
(a) When $n = 1$, D has the effect of differentiation on functions. For scalars s, t and functions f, g in $\mathcal{D}(\mathbb{R})$,

$$D(sf + tg) = \frac{d(sf + tg)}{dx} = s\frac{df}{dx} + t\frac{dg}{dx} = sD(f) + tD(g).$$

Hence D is a linear transformation. The formulae for D^n are clear and compositions of linear transformations are also linear transformations by Theorem 3.11.

To understand the kernel of D^n, notice that $f \in \operatorname{Ker} D^n$ if and only if $D^n(f) = 0$, so we are looking for the general solution of the differential equation

$$\frac{d^n f}{dx^n} = 0$$

which has general solution

$$f(x) = c_0 + c_1 x + \cdots c_{n-1} x^{n-1} \quad (c_0, \ldots, c_{n-1} \in \mathbb{R}).$$

To find the image of D^n, first notice that for any real function g we have

$$D\left(\int_0^x g(t)\,dt\right) = \frac{d}{dx}\int_0^x g(t)\,dt = g(x).$$

This shows that $\operatorname{Im} D = \mathcal{D}(\mathbb{R})$, *i.e.*, D is surjective. More generally, we find that $\operatorname{Im} D^n = \mathcal{D}(\mathbb{R})$.

(b) This uses Theorem 3.12. The point is that

$$a_0 + a_1 D + \cdots + a_n D^n = a_0 \operatorname{Id}_{\mathcal{D}(\mathbb{R})} + a_1 D + \cdots a_n D^n.$$

We have $f \in \operatorname{Ker}(a_0 + a_1 D + \cdots + a_n D^n)$ if and only if

$$a_0 f + a_1 \frac{df}{dx} + \cdots + a_n \frac{d^n f}{dx^n} = 0,$$

so the kernel is the same as the set of solutions of the latter differential equation. Again we find that this linear transformation is surjective and we leave this as an exercise. $\square$

EXAMPLE 3.23. Let F be a field of scalars. For $n \geqslant 1$, consider the vector space $\mathrm{M}_n(F)$. Define the *trace function* $\operatorname{tr}\colon \mathrm{M}_n(F) \longrightarrow F$ by

$$\operatorname{tr}([a_{ij}]) = \sum_{r=1}^{n} a_{rr} = a_{11} + \cdots + a_{nn}.$$

Show that tr is linear transformation and find bases for its kernel and image.

SOLUTION. For $[a_{ij}], [b_{ij}] \in \mathrm{M}_n(F)$ and $s, t \in F$ we have

$$\begin{aligned}
\operatorname{tr}(s[a_{ij}] + t[b_{ij}]) = \operatorname{tr}([s a_{ij} + t b_{ij}]) &= \sum_{r=1}^{n} (s a_{rr} + t b_{rr}) \\
&= \sum_{r=1}^{n} s a_{rr} + \sum_{r=1}^{n} t b_{rr} \\
&= s \sum_{r=1}^{n} a_{rr} + t \sum_{r=1}^{n} b_{rr} \\
&= s \operatorname{tr}([a_{ij}]) + t \operatorname{tr}([b_{ij}]),
\end{aligned}$$

hence tr is a linear transformation.

Notice that $\dim_F \mathrm{M}_n(F) = n^2$ and $\dim_F F = 1$. Also, for $t \in F$,

$$\operatorname{tr}\begin{bmatrix} t & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} = t.$$

This shows that $\operatorname{Im} \operatorname{tr} = F$, *i.e.*, tr is surjective and $\operatorname{Im} \operatorname{tr} = F$ has basis 1.

By the rank-nullity Theorem 3.19,

$$\operatorname{null} \operatorname{tr} = \dim_F \operatorname{Ker} \operatorname{tr} = n^2 - 1.$$

Now the matrices E^{rs} $(r, s = 1, \ldots, n-1,\ r \neq s)$ and $E^{rr} - E^{(r+1)(r+1)}$ $(r = 1, \ldots, n-1)$ lie in $\operatorname{Ker} \operatorname{tr}$ and are linearly independent. There are

$$n(n-1) + (n-1) = (n+1)(n-1) = n^2 - 1$$

of these, so they form a basis of $\operatorname{Ker} \operatorname{tr}$. $\square$

EXAMPLE 3.24. Let F be a field and let $m, n \geqslant 1$. Consider the *transpose function*

$$(-)^T \colon \mathrm{M}_{m\times n}(F) \longrightarrow \mathrm{M}_{n\times m}(F)$$

defined by

$$[a_{ij}]^T = [a_{ji}],$$

i.e., the (i, j)-th entry $[a_{ij}]^T$ is the (j, i)-th entry of $[a_{ij}]$.

Show that $(-)^T$ is a linear transformation which is an isomorphism. What is its inverse?

SOLUTION. For $[a_{ij}], [b_{ij}] \in \mathrm{M}_{m\times n}(F)$,

$$([a_{ij}] + [b_{ij}])^T = [a_{ij} + b_{ij}]^T = [a_{ji} + b_{ji}] = [a_{ji}] + [b_{ji}] = [a_{ij}]^T + [b_{ij}]^T,$$

and if $s \in F$,

$$(s[a_{ij}])^T = [sa_{ij}]^T = [sa_{ji}] = s[a_{ji}] = s[a_{ji}]^T.$$

Thus $(-)^T$ is a linear transformation.

If $[a_{ij}] \in \mathrm{Ker}(-)^T$ then $[a_{ji}] = O_{n\times m}$ (the $n\times m$ zero matrix), hence $a_{ij} = 0$ for $i = 1, \dots, m$ and $j = 1, \dots, n$ and so $[a_{ji}] = O_{m\times n}$. Thus $\mathrm{null}(-)^T = 0$ and by the rank-nullity theorem,

$$\mathrm{rank}(-)^T = \dim \mathrm{M}_{m\times n}(F) = mn = \dim \mathrm{M}_{n\times m}(F).$$

This shows that $\mathrm{Im}(-)^T = \mathrm{M}_{n\times m}(F)$. Hence $(-)^T$ is an isomorphism. Its inverse is

$$(-)^T \colon \mathrm{M}_{n\times m}(F) \longrightarrow \mathrm{M}_{m\times n}(F). \qquad \square$$

Similar arguments apply to the next example.

EXAMPLE 3.25. Let $m, n \geqslant 1$ and consider the *Hermitian conjugate function*

$$(-)^* \colon \mathrm{M}_{m\times n}(\mathbb{C}) \longrightarrow \mathrm{M}_{n\times m}(\mathbb{C})$$

defined by

$$[a_{ij}]^* = [\overline{a_{ji}}],$$

i.e., the (i, j)-th entry $[a_{ij}]^*$ is the complex conjugate of the (j, i)-th entry of $[a_{ij}]$.

If $\mathrm{M}_{m\times n}(\mathbb{C})$ and $\mathrm{M}_{n\times m}(\mathbb{C})$ are viewed as real vector spaces, then $(-)^*$ is a linear transformation which is an isomorphism.

There are some useful rules for dealing with transpose and Hermitian conjugates of products of matrices.

PROPOSITION 3.26 (Reversal rules).
(a) *If F is any field, the transpose operation satisfies*

$$(AB)^T = B^T A^T$$

for all matrices $A \in \mathrm{M}_{\ell\times m}(F)$ and $B \in \mathrm{M}_{m\times n}(F)$.
(b) *For all complex matrices $P \in \mathrm{M}_{\ell\times m}(\mathbb{C})$ and $\mathrm{M}_{m\times n}(\mathbb{C})$,*

$$(PQ)^* = Q^* P^*.$$

PROOF. (a) Writing $A = [a_{ij}]$ and $B = [b_{ij}]$, we obtain

$$\begin{aligned}(AB)^T = ([a_{ij}][b_{ij}])^T = \left[\sum_{r=1}^{m} a_{ir}b_{rj}\right]^T &= \left[\sum_{r=1}^{m} a_{jr}b_{ri}\right]\\ &= \left[\sum_{r=1}^{m} b_{ri}a_{jr}\right]\\ &= [b_{ji}][a_{ji}] = [b_{ij}]^T[a_{ij}]^T = B^T A^T.\end{aligned}$$

(b) is proved in a similar way but with the appearance of complex conjugates in the formulae. □

DEFINITION 3.27. For an arbitrary field F, an matrix $A \in \mathrm{M}_n(F)$ is called

- *symmetric* if $A^T = A$,
- *skew-symmetric* if $A^T = -A$.

A complex matrix $B \in \mathrm{M}_n(\mathbb{C})$ is called

- *Hermitian* if $B^* = A$,
- *skew-Hermitian* if $B^* = -A$.

3.3. Working with bases and coordinates

When performing calculations with a linear transformation it is often useful to use coordinates with respect to bases of the domain and codomain. In fact, this reduces every linear transformation to one associated to a matrix as described in Example 3.10.

LEMMA 3.28. *Let $f, g\colon V \longrightarrow W$ be linear transformations and let $S\colon \mathbf{v}_1, \dots, \mathbf{v}_n$ be a basis for V. Then $f = g$ if and only if the values of f and g agree on the elements of S. Hence a linear transformation is completely determined by its values on the elements of a basis for its domain.*

PROOF. Every element $\mathbf{v} \in V$ can be uniquely expressed as a linear combination

$$\mathbf{v} = s_1\mathbf{v}_1 + \cdots + s_n\mathbf{v}_n$$

Now

$$\begin{aligned}f(\mathbf{v}) &= f(s_1\mathbf{v}_1 + \cdots + s_n\mathbf{v}_n)\\ &= s_1 f(\mathbf{v}_1) + \cdots + s_n f(\mathbf{v}_n)\end{aligned}$$

and similarly

$$\begin{aligned}g(\mathbf{v}) &= g(s_1\mathbf{v}_1 + \cdots + s_n\mathbf{v}_n)\\ &= s_1 g(\mathbf{v}_1) + \cdots + s_n g(\mathbf{v}_n).\end{aligned}$$

So the functions f and g agree on $\mathbf{v}$ if and only if for each i, $f(\mathbf{v}_i) = g(\mathbf{v}_i)$. We conclude that f and g agree on V if and only if they agree on S. □

Here is another useful result involving bases.

LEMMA 3.29. *Let V and W be finite dimensional vector spaces over F, and let $S\colon \mathbf{v}_1, \dots, \mathbf{v}_n$ be a basis for V. Suppose that $f\colon V \longrightarrow W$ is a linear transformation. Then*
(a) *if $f(\mathbf{v}_1), \dots, f(\mathbf{v}_n)$ is linearly independent in W, then f is an injection;*
(b) *if $f(\mathbf{v}_1), \dots, f(\mathbf{v}_n)$ spans W, then f is surjective;*
(b) *if $f(\mathbf{v}_1), \dots, f(\mathbf{v}_n)$ forms a basis for W, then f is bijective.*

PROOF. (a) The equation

$$f(s_1\mathbf{v}_1) + \cdots + s_n\mathbf{v}_n) = \mathbf{0}$$

is equivalent to

$$s_1 f(\mathbf{v}_1) + \cdots + s_n f(\mathbf{v}_n) = \mathbf{0}$$

and so if the vectors $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)$ are linearly independent, then this equation has unique solution $s_1 = \cdots = s_n = 0$. This shows that $\operatorname{Ker} f = \{\mathbf{0}\}$ and hence by Theorem 3.15 we see that f is injective.
(b) $\operatorname{Im} f \subseteq W$ contains every vector of form

$$f(t_1\mathbf{v}_1 + \cdots + t_n\mathbf{v}_n) = t_1 f(\mathbf{v}_1) + \cdots + t_n f(\mathbf{v}_n) \quad (t_1, \ldots, t_n \in F).$$

If the vectors $f(\mathbf{v}_1), \ldots, f(\mathbf{v}_n)$ span W, this shows that $W \subseteq \operatorname{Im} f$ and so f is surjective.
(c) This follows by combining (a) and (b). □

THEOREM 3.30. *Let $S\colon \mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis for the vector space V. Let W be a second vector space and let $T\colon \mathbf{w}_1, \ldots, \mathbf{w}_n$ be any sequence of vectors in W. Then there is a unique linear transformation $f\colon V \longrightarrow W$ satisfying*

$$f(\mathbf{v}_i) = \mathbf{w}_k \quad (i = 1, \ldots, n).$$

PROOF. For every vector $\mathbf{v} \in V$, there is a unique expression

$$\mathbf{v} = s_1\mathbf{v}_1 + \cdots + s_k\mathbf{v}_k,$$

where $s_i \in F$. Then we can set

$$f(\mathbf{v}) = s_1\mathbf{w}_1 + \cdots + s_k\mathbf{w}_k$$

since the right hand side is clearly well defined in terms of $\mathbf{v}$. In particular, if $\mathbf{v} = \mathbf{v}_i$, then $f(\mathbf{v}_i) = \mathbf{w}_i$ and that f is a linear transformation $f\colon V \longrightarrow W$. Uniqueness follows from Lemma 3.28. □

A useful consequence of this result is the following.

THEOREM 3.31. *Suppose that the vector space V is a direct sum $V = V_1 \oplus V_2$ of two subspaces and let $f_1\colon V_1 \longrightarrow W$ and $f_2\colon V_2 \longrightarrow W$ be two linear transformations. Then there is a unique linear transformation $f\colon V \longrightarrow W$ extending f_1, f_2,* i.e., *satisfying*

$$f(\mathbf{u}) = \begin{cases} f_1(\mathbf{u}) & \text{if } \mathbf{u} \in V_1, \\ f_2(\mathbf{u}) & \text{if } \mathbf{u} \in V_2. \end{cases}$$

PROOF. Use bases of V_1 and V_2 to form a basis of V, then use Theorem 3.30. □

Here is another useful result.

THEOREM 3.32. *Let V and W be finite dimensional vector spaces over F with $n = \dim_F V$ and $m = \dim_F W$. Suppose that $S\colon \mathbf{v}_1, \ldots, \mathbf{v}_n$ and $T\colon \mathbf{w}_1, \ldots, \mathbf{w}_m$ are bases for V and W respectively. Then there is a unique transformation $f\colon V \longrightarrow W$ with the following properties:*
(a) *if $n \leqslant m$ then*

$$f(\mathbf{v}_j) = \mathbf{w}_j \quad \text{for } j = 1, \ldots, n;$$

(b) *if $n > m$ then*

$$f(\mathbf{v}_j) = \begin{cases} \mathbf{w}_j & \text{when } j = 1, \ldots, m, \\ \mathbf{0} & \text{when } j > m. \end{cases}$$

Furthermore, f also satisfies:
(c) *if $n \leqslant m$ then f is injective;*
(d) *if $n > m$ then f is surjective;*
(e) *if $m = n$ then f is an isomorphism.*

PROOF. (a) Starting with the sequence $\mathbf{w}_1, \dots, \mathbf{w}_n$, we can apply Theorem 3.30 there is a unique linear transformation $f\colon V \longrightarrow W$ with the stated properties.
(b) Starting with the sequence $\mathbf{w}_1, \dots, \mathbf{w}_m, \mathbf{0}, \dots, \mathbf{0}$ of length n, we can apply Theorem 3.30 there is a unique linear transformation $f\colon V \longrightarrow W$ with the stated properties.
(c), (d) and (e) now follow from Lemma 3.29. □

The next result provides a convenient way to decide if two vector spaces are isomorphic: simply show that they have the same dimension.

THEOREM 3.33. *Let V and W be finite dimensional vector spaces over the field F. Then there is an isomorphism $V \longrightarrow W$ if and only if $\dim_F V = \dim_F W$.*

PROOF. If there is an isomorphism $f\colon V \longrightarrow W$, then $\dim_F V = \dim_F W$ by Corollary 3.21(c).
For the converse, if $\dim_F V = \dim_F W$ then by Theorem 3.32(e) there is an isomorphism $V \longrightarrow W$. □

Now we will see how to make use of bases for the domain and codomain to work with a linear transformation by using matrices.

Let V and W be two finite dimensional vector spaces. Let $S\colon \mathbf{v}_1, \dots, \mathbf{v}_n$ and $T\colon \mathbf{w}_1, \dots, \mathbf{w}_m$ be bases for V and W respectively. Suppose that $f\colon V \longrightarrow W$ is a linear transformation. Then there are unique scalars $t_{ij} \in F$ $(i = 1, \dots, m,\ j = 1, \dots, n)$ for which

$$f(\mathbf{v}_j) = t_{1j}\mathbf{w}_1 + \cdots + t_{mj}\mathbf{w}_m. \tag{3.1}$$

We can form the $n \times m$ matrix

$$_T[f]_S = [t_{ij}] = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{m1} & \cdots & t_{mn} \end{bmatrix}. \tag{3.2}$$

REMARK 3.34.
(a) Notice that this matrix depends on the bases S and T as well as f.
(b) The coefficients are easily found: the j-th column is made up from the coefficients that occur when expressing $f(\mathbf{v}_j)$ in terms of the $\mathbf{w}_i$.

In the next result we will also assume that there are bases $S'\colon \mathbf{v}'_1, \dots, \mathbf{v}'_n$ and $T'\colon \mathbf{w}'_1, \dots, \mathbf{w}'_m$ for V and W respectively, for which

$$\mathbf{v}'_j = a_{1j}\mathbf{v}_1 + \cdots + a_{nj}\mathbf{v}_n = \sum_{i=1}^{n} a_{ij}\mathbf{v}_i \qquad (j = 1, \dots, n),$$

$$\mathbf{w}'_k = b_{1k}\mathbf{v}_1 + \cdots + b_{mk}\mathbf{w}_m = \sum_{i=1}^{n} b_{ik}\mathbf{w}_i \qquad (k = 1, \dots, m).$$

Working with these bases we obtain the matrix

$$_{T'}[f]_{S'} = [t'_{ij}] = \begin{bmatrix} t'_{11} & \cdots & t'_{1n} \\ \vdots & \ddots & \vdots \\ t'_{m1} & \cdots & t'_{mn} \end{bmatrix}.$$

We will set

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, \quad B = [b_{ij}] = \begin{bmatrix} b_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix}.$$

THEOREM 3.35.
(a) *Let $\mathbf{x} \in V$ have coordinates $x_1, \ldots, x_n$ with respect to the basis $S\colon \mathbf{v}_1, \ldots, \mathbf{v}_n$ of V and let $f(\mathbf{x}) \in W$ have coordinates $y_1, \ldots, y_m$ with respect to the basis $T\colon \mathbf{w}_1, \ldots, \mathbf{w}_m$ of W. Then*

(3.3a)
$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = {}_T[f]_S \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{m1} & \cdots & t_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

(b) *The matrices $_T[f]_S$ and $_{T'}[f]_{S'}$ are related by the formula*

(3.3b)
$$_{T'}[f]_{S'} = B^{-1}{}_T[f]_S A.$$

PROOF. These follow from calculations with the above formulae. □

The formulae of (3.3), allow us to reduce calculations to ones with coordinates and matrices.

EXAMPLE 3.36. Consider the linear transformation

$$f\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2; \quad f(x, y, z) = (2x - 3y + 7z, x - z)$$

between the real vector spaces $\mathbb{R}^3$ and $\mathbb{R}^2$.
(a) Write down the matrix of f with respect to the standard bases of $\mathbb{R}^3$ and $\mathbb{R}^2$.
(b) Determine the matrix $_T[f]_S$, where S and T are the bases

$$S\colon (1, 0, 1), (0, 1, 1), (0, 0, 1), \quad T\colon (3, 1), (2, 1).$$

SOLUTION.
(a) In terms of the standard basis vectors $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ and $(1, 0), (0, 1)$ we have

$$\begin{aligned} f(1, 0, 0) &= (2, 1) = 2(1, 0) + 1(0, 1), \\ f(0, 1, 0) &= (-3, 0) = -3(1, 0) + 0(0, 1), \\ f(0, 0, 1) &= (7, -1) = 7(1, 0) + (-1)(0, 1), \end{aligned}$$

hence the matrix is $\begin{bmatrix} 2 & -3 & 7 \\ 1 & 0 & -1 \end{bmatrix}$.
(b) We have

$$\begin{aligned} (1, 0, 1) &= 1(1, 0, 0) + 0(0, 1, 0) + 1(0, 0, 1), \\ (0, 1, 1) &= 0(1, 0, 0) + 1(0, 1, 0) + 1(0, 0, 1), \\ (0, 0, 1) &= 0(1, 0, 0) + 0(0, 1, 0) + 1(0, 0, 1), \end{aligned}$$

so we take

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

We also have

$$(3,1) = 3(1,0) + 1(0,1), \quad (2,1) = 2(1,0) + 1(0,1),$$

and so we take

$$B = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}.$$

The inverse of B is

$$B^{-1} = \frac{1}{3-2} \begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix},$$

so we have

$$\begin{aligned} {}_T[f]_S &= B^{-1} \begin{bmatrix} 2 & -3 & 7 \\ 1 & 0 & -1 \end{bmatrix} A \\ &= \begin{bmatrix} 1 & -2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 2 & -3 & 7 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 9 & 6 & 9 \\ -9 & -7 & -10 \end{bmatrix}. \end{aligned}$$

□

THEOREM 3.37. *Suppose that $f\colon V \longrightarrow W$ and $g\colon U \longrightarrow V$ are linear transformations and that R, S, T are bases for U, V, W respectively. Then the matrices ${}_T[f]_S$, ${}_S[g]_R$, ${}_T[f \circ g]_R$ are related by the matrix equation*

$${}_T[f \circ g]_R = {}_T[f]_S {}_S[g]_R.$$

PROOF. This follows by a calculation using the formula (3.1). Notice that the right hand matrix product does make sense because ${}_T[f]_S$ is $\dim W \times \dim V$, ${}_S[g]_R$ and is $\dim V \times \dim U$, while ${}_T[f \circ g]_R$ is $\dim W \times \dim U$. □

3.4. Application to matrices and systems of linear equations

Let F be a field and consider an $m \times n$ matrix $A \in \mathrm{M}_{m\times n}(F)$. Recall that by performing a suitable sequence of *elementary row operations* we may obtain a reduced echelon matrix A' which is row equivalent to A, *i.e.*, $A' \sim A$. This can be used to solve a linear system of the form $A\mathbf{x} = \mathbf{b}$.

It is also possible to perform *elementary column operations* on A, or equivalently, to form A^T then perform elementary row operations until we get a reduced echelon matrix A'', then transpose back to get another *column reduced echelon matrix* $(A'')^T$.

DEFINITION 3.38. The number of non-zero rows in A' is called the *row rank of* A. The number of non-zero columns in $(A'')^T$ is called the *column rank of* A.

EXAMPLE 3.39. Taking $F = \mathbb{R}$, find the row and column ranks of the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \\ 5 & 1 \end{bmatrix}.$$

SOLUTION. For the row rank we have

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \\ 5 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 \\ 0 & 3 \\ 0 & -9 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 0 & -9 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = A',$$

so the row rank is 2.

For the column rank,

$$A^T = \begin{bmatrix} 1 & 0 & 5 \\ 2 & 3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 5 \\ 0 & 3 & -9 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & -3 \end{bmatrix} = A'',$$

which gives

$$(A'')^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 5 & -3 \end{bmatrix}$$

and the column rank is 2. □

THEOREM 3.40. *Let* $A \in \mathrm{M}_{m\times n}(F)$. *Then*

$$\text{row rank of } A = \text{column rank of } A.$$

PROOF OF THEOREM 3.40. We will make use of ideas from Example 3.10. Recall the linear transformation $f_A \colon F^n \longrightarrow F^m$ determined by

$$f_A(\mathbf{x}) = A\mathbf{x} \quad (\mathbf{x} \in F^n).$$

First we identify the value of the row rank of A in other terms. Let r be the row rank of A. Then

$$\operatorname{Ker} f_A = \{\mathbf{x} \in F^n : A\mathbf{x} = \mathbf{0}\}$$

is the set of solution of the associated system of homogeneous linear equations. Recall that once the reduced echelon matrix A' has been found, the general solution can be expressed in terms of $n - r$ basic solutions each obtained by setting one of the parameters equal to 1 and the rest equal to 0. These basic solutions form a spanning sequence for $\operatorname{Ker} f_A$ which linearly independent, hence is a basis for $\operatorname{Ker} f_A$. So we have

$$\text{row rank of } A = r = n - \dim \operatorname{Ker} f_A. \tag{3.4}$$

On the other hand, if c is the column rank of A, we find that the columns of $(A'')^T$ are a spanning sequence for the subspace of F^m spanned by the columns of A, and in fact, they are linearly independent.

As the columns of A span $\operatorname{Im} f_A$,

$$\text{column rank of } A = c = \dim \operatorname{Im} f_A. \tag{3.5}$$

But now we can use the rank-nullity Theorem 3.19 to deduce that

$$n = \dim \operatorname{Ker} f_A + \dim \operatorname{Im} f_A = (n - r) + c = n + (c - r),$$

giving $c = r$ as claimed. □

REMARK 3.41. In this proof we saw that the columns of A span $\operatorname{Im} f_A$ and also how to find a basis for $\operatorname{Ker} f_A$ by the method of solution of linear systems using elementary row operations.

DEFINITION 3.42. The *rank* of A is the number

$$\operatorname{rank} A = \operatorname{rank} f_A = \dim \operatorname{Im} f_A = \text{column rank of } A = \text{row rank of } A,$$

where $\operatorname{rank} f_A$ was defined in Definition 3.18. It agrees with the number of non-zero rows in the reduced echelon matrix similar to A.

The *kernel of* A is $\operatorname{Ker} A = \operatorname{Ker} f_A$ and the *image of* A is $\operatorname{Im} A = \operatorname{Im} f_A$.

3.5. Geometric linear transformations

When considering a linear transformation $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ on the real vector space $\mathbb{R}^n$ it is often important or useful to understand its geometric content.

EXAMPLE 3.43. For $\theta \in \mathbb{R}$, consider the linear transformation $\rho_\theta\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ given by

$$\rho_\theta(x, y) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta).$$

Investigate the geometric effect of ρ_θ on the plane.

SOLUTION. The effect on the standard basis vectors is

$$\rho_\theta(1, 0) = (\cos\theta, \sin\theta), \quad \rho_\theta(0, 1) = (-\sin\theta, \cos\theta).$$

Using the dot product of vectors it is possible to check that the angle between $\rho_\theta(x, y)$ and (x, y) is always θ if $(x, y) \neq (0, 0)$, and that the lengths of $\rho_\theta(x, y)$ and (x, y) are equal. So ρ_θ has the effect of rotating the plane about the origin through the angle θ measured in the anti-clockwise direction.

Notice that the matrix of ρ_θ with respect to the standard basis of $\mathbb{R}^2$ is the *rotation matrix*

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

EXAMPLE 3.44. For $\theta \in \mathbb{R}$, consider the linear transformation $\sigma_\theta\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ given by

$$\sigma_\theta(x, y) = (x\cos\theta + y\sin\theta, x\sin\theta - y\cos\theta).$$

Investigate the geometric effect of σ_θ on the plane.

SOLUTION. The effect on the standard basis vectors is

$$\sigma_\theta(1, 0) = (\cos\theta, \sin\theta), \quad \sigma_\theta(0, 1) = (\sin\theta, -\cos\theta).$$

The angle between every pair of vectors is preserved by σ_θ, as is the length of a vector. We also find that

$$\begin{aligned} \sigma_\theta(\cos(\theta/2), \sin(\theta/2)) &= (\cos(\theta/2), \sin(\theta/2)), \\ \sigma_\theta(-\sin(\theta/2), \cos(\theta/2)) &= (\sin(\theta/2), -\cos(\theta/2)) = -(-\sin(\theta/2), \cos(\theta/2)), \end{aligned}$$

from which it follows that σ_θ has the effect of reflecting the plane in the line through the origin and parallel to $(\cos(\theta/2), \sin(\theta/2))$.

Notice that the matrix of σ_θ with respect to the standard basis of $\mathbb{R}^2$ is the *reflection matrix*

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

EXAMPLE 3.45. Let $S = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$ and let $f_S \colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be the real linear transformation given by

$$f_S(\mathbf{x}) = S\mathbf{x}.$$

Then f_S has the geometric effect of a *shearing* parallel to the x-axis.

SOLUTION. This can be seen by plotting the effect of f_S on points. For example, points on the x-axis are fixed by f_S, while points on the y-axis get moved parallel to the x-axis into the line of slope t. □

EXAMPLE 3.46. Let $D = \begin{bmatrix} d & 0 \\ 0 & 1 \end{bmatrix}$ where $d > 0$ and let $f_D \colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be the real linear transformation defined by

$$f_D(\mathbf{x}) = D\mathbf{x}.$$

Then f_D has the geometric effect of a *dilation* parallel to the x-axis.

SOLUTION. When applying f_D, vectors along the x-axis get dilated by a factor d, while those on the y-axis are unchanged. □

EXAMPLE 3.47. Let V a finite dimensional vector space over a field F. Suppose that $V = V' \oplus V''$ is the direct sum of two subspaces V', V'' (see Definition 2.38). Then there is a linear transformation $p_{V',V''} \colon V \longrightarrow V$ given by

$$p_{V',V''}(\mathbf{v}' + \mathbf{v}'') = \mathbf{v}' \quad (\mathbf{v}' \in V',\ \mathbf{v}'' \in V'').$$

Then $\operatorname{Ker} p_{V',V''} = V''$ and $\operatorname{Im} p_{V',V''} = V'$.

$p_{V',V''}$ is sometimes called the *projection* of V onto V' with kernel V''. The case where V'' is a line (*i.e.*, $\dim V'' = 1$) is particularly important.

EXAMPLE 3.48. Find a formula for the projection of $\mathbb{R}^3$ onto the plane P through the origin and perpendicular to $(1, 0, 1)$ with kernel equal to the line L spanned by $(1, 0, 1)$.

SOLUTION. Let this projection be $p \colon \mathbb{R}^3 \longrightarrow \mathbb{R}^3$. Then every vector $\mathbf{x} \in \mathbb{R}^3$ can be expressed as

$$\mathbf{x} = \Big(\mathbf{x} - [(1/\sqrt{2}, 0, 1/\sqrt{2}) \cdot \mathbf{x}](1/\sqrt{2}, 0, 1/\sqrt{2})\Big) + [(1/\sqrt{2}, 0, 1/\sqrt{2}) \cdot \mathbf{x}](1/\sqrt{2}, 0, 1/\sqrt{2}),$$

where

$$\mathbf{x} - [(1/\sqrt{2}, 0, 1/\sqrt{2}) \cdot \mathbf{x}](1/\sqrt{2}, 0, 1/\sqrt{2}) \in P$$

and

$$[(1/\sqrt{2}, 0, 1/\sqrt{2}) \cdot \mathbf{x}](1/\sqrt{2}, 0, 1/\sqrt{2}) \in L.$$

This corresponds to the direct sum decomposition $\mathbb{R}^3 = P \oplus L$.

The projection p is given by

$$p(\mathbf{x}) = \mathbf{x} - [(1/\sqrt{2}, 0, 1/\sqrt{2}) \cdot \mathbf{x}](1/\sqrt{2}, 0, 1/\sqrt{2}).$$ □

CHAPTER 4

Determinants

For 2×2 and 3×3 matrices determinants are defined by the formulae

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \tag{4.1}$$

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} - a_{12} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + a_{13} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}. \tag{4.2}$$

We will introduce determinants for all $n \times n$ matrices over any field of scalars and these will have properties analogous to that of (4.2) which allows the calculation of 3×3 determinants from 2×2 ones.

Definition 4.1. For the $n \times n$ matrix $A = [a_{ij}]$ with entries in F, for each pair $r, s = 1, \ldots, n$, the *cofactor matrix* A^{rs} is the $(n-1) \times (n-1)$ matrix obtained by deleting the r-th row and s-th column of A.

Warning. When working with the determinant of a matrix $[a_{ij}]$ we will often write $|A|$ for $\det(A)$ and display it as an array with vertical lines rather than square brackets as in (4.1) and (4.2). But beware, the two expressions

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

and

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

are very different things: the first is a *matrix*, whereas the second is a *scalar*. ***They should not be confused!***

4.1. Definition and properties of determinants

Let F be a field of scalars.

Theorem 4.2. *For each $n \geqslant 1$, and every $n \times n$ matrix $A = [a_{ij}]$ with entries in F, there is a uniquely determined scalar*

$$\det A = \det(A) = |A| \in F$$

which satisfies the following properties.

(A) *For the $n \times n$ identity matrix I_n,*

$$\det(I_n) = 1.$$

(B) *If A, B are two $n \times n$ matrices then*

$$\det(AB) = \det(A)\det(B).$$

(C) *Let A be an $n \times n$ matrix. Then we have*
Expansion along the r-th row:

$$\det(A) = \sum_{j=1}^{n} (-1)^{r+j} a_{rj} \det(A^{rj})$$

for any $r = 1, \ldots, n$.
Expansion along the s-th column:

$$\det(A) = \sum_{i=1}^{n} (-1)^{s+i} a_{is} \det(A^{is})$$

for any $s = 1, \ldots, n$. The signs in these expansions can be obtained from the pattern

$$\begin{vmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ \vdots & & \ddots & & \vdots \end{vmatrix}$$

(D) *For an $n \times n$ matrix A,*

$$\det(A^T) = \det(A).$$

Notice that the formula

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

from (C) is a direct generalisation of (4.2).

Here are some important results on determinants of invertible matrices that follow from the above properties.

PROPOSITION 4.3. *Let A be an invertible $n \times n$ matrix.*
(i) $\det(A) \neq 0$ *and*

$$\det(A^{-1}) = \det(A)^{-1} = \frac{1}{\det(A)}.$$

(ii) *If B is any $n \times n$ matrix then*

$$\det(ABA^{-1}) = \det(B).$$

PROOF.
(i) By (B),

$$\det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1}),$$

hence using (A) we find that

$$\det(A)\det(A^{-1}) = 1,$$

so $\det(A^{-1}) \neq 0$ and

$$\det(A^{-1}) = \det(A)^{-1} = \frac{1}{\det(A)}.$$

(ii) By (B) and (i),

$$\begin{aligned}\det(ABA^{-1}) &= \det(AB)\det(A^{-1})\\ &= \det(A)\det(B)\det(A^{-1})\\ &= \det(B)\det(A)\det(A^{-1}) = \det(B).\end{aligned}$$ □

Let us relate this to elementary row operations and elementary matrices, whose basic properties we will assume. Suppose that R is an elementary row operation and $E(R)$ is the corresponding elementary matrix. Then the matrix A' obtained from A by applying R satisfies

$$A' = E(R)A,$$

hence

$$\det(A') = \det(E(R)A) = \det(E(R))\det(A). \tag{4.3}$$

Proposition 4.4. *If $A' = E(R)A$ is the result of applying the elementary row operation R to the $n \times n$ matrix A, then*

$$\det(A') = \begin{cases} -\det(A) & \text{if } R = (R_r \leftrightarrow R_s),\\ \lambda\det(A) & \text{if } R = (R_r \rightarrow \lambda R_r),\\ \det(A) & \text{if } R = (R_r \rightarrow R_r + \lambda R_s).\end{cases}$$

In particular,

$$\det(E(R)) = \begin{cases} -1 & \text{if } R = (R_r \leftrightarrow R_s),\\ \lambda & \text{if } R = (R_r \rightarrow \lambda R_r),\\ 1 & \text{if } R = (R_r \rightarrow R_r + \lambda R_s).\end{cases}$$

Proof. If $R = R_r \leftrightarrow R_s$, then it is possible to prove by induction on the size n that

$$\det(E(R_r \leftrightarrow R_s)) = -1.$$

The initial case is $n = 2$ where

$$\det(E(R_1 \leftrightarrow R_2)) = \begin{vmatrix} 0 & 1\\ 1 & 0\end{vmatrix} = -1.$$

If $R = R_r \rightarrow \lambda R_r$ for $\lambda \neq 0$, then expanding along the r-th row gives

$$\det(E(R_r \rightarrow \lambda R_r)) = \lambda\det(I_n) = \lambda.$$

Finally, if $R = R_r \rightarrow R_r + \lambda R_s$ with $r \neq s$, then expanding along the r-th row gives

$$\det(E(R_r \rightarrow R_r + \lambda R_s)) = \det(I_{n-1}) + \lambda\det(I_n^{rs}) = 1 + \lambda \times 0 = 1.$$ □

Thus to calculate a determinant, first find any sequence $R_1, \ldots R_k$ of elementary row operations so that the combined effect of applying these successively to A is an upper triangular matrix L. Then $L = E(R_k)\cdots E(R_1)A$ and so

$$\det(L) = \det(E(R_k))\cdots\det(E(R_1))\det(A),$$

giving

$$\det(A) = \det(E(R_k))^{-1}\cdots\det(E(R_1))^{-1}\det(L),$$

To calculate $\det(L)$ is easy, since

$$\det(L) = \begin{vmatrix} \ell_1 & * & * & * & * & * & * \\ 0 & \ell_2 & * & * & * & * & * \\ 0 & 0 & \ell_3 & * & * & * & * \\ \vdots & \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \ell_n \end{vmatrix} = \ell_1 \begin{vmatrix} \ell_2 & * & * & * & * & * \\ 0 & \ell_3 & * & * & * & * \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \ell_n \end{vmatrix}$$

by definition and repeating this gives $\det(L) = \ell_1\ell_2 \cdots \ell_n$.

Using this we can obtain a useful criterion for invertibility of a square matrix.

PROPOSITION 4.5. *Let A be an $n \times n$ matrix with entries in a field of scalars F. Then* $\det(A) = 0$ *if and only if A is singular. Equivalently,* $\det(A) \neq 0$ *if and only if A is invertible.*

PROOF. By Proposition 4.3 we know that if A is invertible, then $\det(A) \neq 0$.

On the other hand, if $\det(A) \neq 0$ then suppose that the reduced echelon form of A is A'. If there is a zero row in A' then $\det(A') = 0$. But if $A' = E(R_k) \cdots E(R_1)A$, then

$$0 = \det(A') = \det(E(R_k)) \cdots \det(E(R_1)) \det(A),$$

where the scalars $\det(E(R_k)), \ldots, \det(E(R_1)), \det(A)$ are all non-zero. Hence this cannot happen and so $A' = I_n$, showing that A is invertible. □

There is an explicit formula, *Cramer's Rule*, for finding the inverse of a matrix using determinants. However, it is not much use for computations with large matrices and methods based on row and column operations are usually much more efficient. In practise, evaluation of determinants is often simplified by using elementary row operations to create zero's or otherwise reduced calculations.

EXAMPLE 4.6. Evaluate the following determinants:

$$\begin{vmatrix} 4 & 5 & 1 & 2 \\ 0 & 3 & 2 & 1 \\ 3 & 0 & 1 & 3 \\ 2 & 7 & 0 & 3 \end{vmatrix}, \quad \begin{vmatrix} x & y^2 & 1 \\ y & 2 & x \\ x+y & 0 & x \end{vmatrix}.$$

SOLUTION.

$$
\begin{aligned}
\begin{vmatrix} 4 & 5 & 1 & 2 \\ 0 & 3 & 2 & 1 \\ 3 & 0 & 1 & 3 \\ 2 & 7 & 0 & 3 \end{vmatrix}
&= \begin{vmatrix} 4 & 0 & 3 & 2 \\ 5 & 3 & 0 & 7 \\ 1 & 2 & 1 & 0 \\ 2 & 1 & 3 & 3 \end{vmatrix} \quad \text{[transposing]} \\
&= \begin{vmatrix} 4 & 0 & 3 & 2 \\ -1 & 0 & -9 & -2 \\ -3 & 0 & -5 & -6 \\ 2 & 1 & 3 & 3 \end{vmatrix} \quad [R_2 \to R_2 - 3R_4,\ R_3 \to R_3 - 2R_4] \\
&= \begin{vmatrix} 4 & 3 & 2 \\ -1 & -9 & -2 \\ -3 & -5 & -6 \end{vmatrix} \quad \text{[expanding along column 2]} \\
&= \begin{vmatrix} 4 & 3 & 2 \\ 1 & 9 & 2 \\ 3 & 5 & 6 \end{vmatrix} \quad \text{[multiplying rows 2 and 3 by } (-1)] \\
&= \begin{vmatrix} 4 & 3 & 2 \\ -3 & 6 & 0 \\ 0 & -22 & 0 \end{vmatrix} \quad [R_2 \to R_2 - R_1,\ R_3 \to R_3 - 3R_1] \\
&= 22 \begin{vmatrix} 4 & 2 \\ -3 & 0 \end{vmatrix} = 22 \times 6 = 132, \quad \text{[expanding along row 2]}
\end{aligned}
$$

$$
\begin{aligned}
\begin{vmatrix} x & y^2 & 1 \\ y & 2 & x \\ x+y & 0 & x \end{vmatrix}
&= (x+y) \begin{vmatrix} y^2 & 1 \\ 2 & x \end{vmatrix} + x \begin{vmatrix} x & y^2 \\ y & 2 \end{vmatrix} \quad \text{[expanding along row 3]} \\
&= (x+y)(xy^2 - 2) + x(2x - y^3) \\
&= x^2y^2 - 2x + xy^3 - 2y + 2x^2 - xy^3 = x^2y^2 - 2x - 2y + 2x^2.
\end{aligned}
$$

□

EXAMPLE 4.7. Solve the equation

$$
\begin{vmatrix} 1 & 2 & -1 \\ x & x+1 & x^2 \\ 1 & 1 & 5 \end{vmatrix} = 0.
$$

SOLUTION. Expanding the determinant gives

$$
\begin{aligned}
\begin{vmatrix} 1 & 2 & -1 \\ x & x+1 & x^2 \\ 1 & 1 & 5 \end{vmatrix}
&= \begin{vmatrix} 1 & 2 & -1 \\ 0 & 1-x & x^2+x \\ 0 & -1 & 6 \end{vmatrix} \quad [R_2 \to R_2 - xR_1,\ R_3 \to R_3 - R_1] \\
&= \begin{vmatrix} 1-x & x^2+x \\ -1 & 6 \end{vmatrix} \quad \text{[expanding along column 1]} \\
&= x^2 + x + 6 - 6x = x^2 - 5x + 6 = (x-2)(x-3).
\end{aligned}
$$

So the solutions are $x = 2, 3$. □

EXAMPLE 4.8. Evaluate the *Vandermonde determinant* $\begin{vmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{vmatrix}$. Determine when the matrix $\begin{bmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{bmatrix}$ is singular.

SOLUTION.

$$\begin{aligned}
\begin{vmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{vmatrix} = \begin{vmatrix} 1 & x & x^2 \\ 0 & y-x & y^2-x^2 \\ 0 & z-x & z^2-x^2 \end{vmatrix} &= \begin{vmatrix} 1 & x & x^2 \\ 0 & y-x & (y-x)(y+x) \\ 0 & z-x & (z-x)(z+x) \end{vmatrix} \\
&= (y-x)(z-x) \begin{vmatrix} 1 & x & x^2 \\ 0 & 1 & y+x \\ 0 & 1 & z+x \end{vmatrix} \\
&= (y-x)(z-x) \begin{vmatrix} 1 & y+x \\ 0 & z-y \end{vmatrix} \\
&= (y-x)(z-x)(z-y).
\end{aligned}$$

The matrix is singular precisely when its determinant vanishes, and this happens when at least one of $(y-x), (z-x), (z-y)$ is zero, *i.e.*, when two of the scalars x, y, z are equal. $\square$

REMARK 4.9. For $n \geqslant 1$, there is a general formula for an $n \times n$ *Vandermonde determinant*,

$$\begin{vmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & & \ddots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{vmatrix} = \prod_{1 \leqslant r < s \leqslant n} (x_s - x_r).$$

The proof is similar to that for the case $n = 3$. Again we can see when the matrix

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & & \ddots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}$$

is singular, namely when $x_r = x_s$ for some pair of distinct indices r, s.

4.2. Determinants of linear transformations

Let V be a finite dimensional vector space over the field of scalars F and suppose that $f\colon V \longrightarrow V$ is a linear transformation. Write $n = \dim V$.

If we choose a basis $S\colon \mathbf{v}_1, \ldots \mathbf{v}_n$ for V, then there is a matrix ${}_S[f]_S$ and we can find its determinant $\det({}_S[f]_S)$. A second basis $T\colon \mathbf{w}_1, \ldots \mathbf{w}_n$ leads to another determinant $\det({}_T[f]_T)$. Appealing to Theorem 3.35(b) we find that for a suitable invertible matrix P,

$${}_T[f]_T = P^{-1}{}_S[f]_S P,$$

hence by Proposition 4.3,

$$\det({}_T[f]_T) = \det(P^{-1}{}_S[f]_S P) = \det({}_S[f]_S).$$

This shows that $\det({}_S[f]_S)$ does not depend on the choice of basis, only on f, hence it is an *invariant* of f.

DEFINITION 4.10. Let $f\colon V \longrightarrow V$ is a linear transformation on a finite dimensional vector space. Then the *determinant* of f is $\det f = \det(f) = \det({}_S[f]_S)$, where S is any basis of V. This value only depends on f and not on the basis S.

REMARK 4.11. Suppose that $V = \mathbb{R}^n$. Then there is a geometric interpretation of $\det f$ (at least for $n = 2, 3$, but it makes sense in the general case if we suitably interpret volume in $\mathbb{R}^n$). The standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ form the edges of a unit cube based at the origin, with volume 1. The vectors $f(\mathbf{e}_1), \ldots, f(\mathbf{e}_n)$ form edges of a parallellipiped based at the origin. The volume of this parallellipiped is $|\det f|$ (note that volumes are non-negative!). This is used in change of variable formulae for multivariable integrals. The value of $\det f$ (or more accurately of $|\det f|$) indicates how much volumes are changed by applying f. The sign of $\det f$ indicates whether or not 'orientations' change and this is also encountered in multivariable integration formulae.

4.3. Characteristic polynomials and the Cayley-Hamilton theorem

For a field F, let A be an $n \times n$ matrix with entries in F.

DEFINITION 4.12. The *characteristic polynomial* of A is the polynomial $\chi_A(t) \in F[t]$ (often denoted $\operatorname{char}_A(t)$) defined by

$$\chi_A(t) = \det(tI_n - A).$$

PROPOSITION 4.13. *The polynomial $\chi_A(t)$ has degree n and is monic,* i.e.,

$$\chi_A(t) = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0,$$

where $c_0, c_1, \ldots, c_{n-1} \in F$. Furthermore,

$$c_{n-1} = -\operatorname{tr} A, \quad c_0 = (-1)^n \det(A).$$

PROOF. To see the first part, write

$$\det(tI_n - A) = \begin{vmatrix} t - a_{11} & \cdots & -a_{1n} \\ \vdots & \ddots & \vdots \\ -a_{n1} & \cdots & t - a_{nn} \end{vmatrix}$$

and then apply property (C) repeatedly.

The formula for c_{n-1} follows from the fact that the terms contributing to the t^{n-1} term are obtained by multiplying $n-1$ of the diagonal terms and then multiplying by the constant term in the remaining one. The result is then

$$c_{n-1} = -(a_{11} + \cdots + a_{nn}) = -\operatorname{tr}(A).$$

For the constant term c_0, taking $t = 0$ and using Proposition 4.4 we obtain

$$c_0 = \det(-A) = (-1)^n \det(A). \qquad \square$$

PROPOSITION 4.14. *Let B be an invertible $n \times n$ matrix, then*

$$\chi_{BAB^{-1}}(t) = \chi_A(t).$$

Proof. This follows from the fact that

$$\chi_{BAB^{-1}}(t) = \det(tI_n - BAB^{-1}) = \det\left(B(tI_n - A)B^{-1}\right) = \det(tI_n - A)$$

by Proposition 4.3(ii). □

This allows us to make the following definition. See Definition 4.10 for a similar idea. Let V be a finite dimensional vector space over a field F and let $f\colon V \longrightarrow V$ be a linear transformation.

Definition 4.15. The *characteristic polynomial* of f is

$$\chi_f(t) = \chi_{{}_S[f]_S}(t) = \det(tI_n - {}_S[f]_S),$$

where S is any basis of V.

We now come to an important result which is useful in connection with the notion of eigenvalues to be considered in Chapter 5.

Let $p(t) = a_k t^k + a_{k-1}t^{k-1} + \cdots + a_1 t + a_0 \in F[t]$ be a polynomial (so $a_0, a_1, \ldots, a_k \in F$). For an $n \times n$ matrix A we write

$$p(A) = a_k A^k + a_{k-1}A^{k-1} + \cdots + a_1 A + a_0 I_n,$$

which is also an $n \times n$ matrix. If $p(A) = O_n$ then we say that A satisfies this *polynomial identity.* Similarly, if $f\colon V \longrightarrow V$ is a linear transformation,

$$p(f) = a_k f^k + a_{k-1}f^{k-1} + \cdots + a_1 f + a_0 \operatorname{Id}_V .$$

Theorem 4.16 (Cayley-Hamilton theorem).

(i) *Let A be an $n \times n$ matrix. Then*

$$\chi_A(A) = O_n.$$

(ii) *Let V be a finite dimensional vector space of dimension n and let $f\colon V \longrightarrow V$ be a linear transformation. Then*

$$\chi_f(f) = 0,$$

where the right hand side is the constant function taking value $\mathbf{0}$.

Informally, these results are sometimes summarised by saying that a square matrix A or a linear transformation $f\colon V \longrightarrow V$ *satisfies its own characteristic equation.*

Example 4.17. Show that the matrix $A = \begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 3 \end{bmatrix}$ satisfies a polynomial identity of degree 3. Hence show that A is invertible and express A^{-1} as a polynomial in A.

Solution. The characteristic polynomial of A is

$$\chi_A(t) = \begin{vmatrix} t & 0 & -2 \\ -1 & t & 0 \\ 0 & -1 & t-3 \end{vmatrix} = t^3 - 3t^2 - 2.$$

So $\det(A) = 2 \neq 0$ and A is invertible. Also,

$$A^3 - 3A^2 - 2I_3 = O_3$$

so

$$A(A^2 - 3A) = (A^2 - 3A)A = A^3 - 3A^2 = 2I_3,$$

giving

$$A^{-1} = \frac{1}{2}(A^2 - 3A). \qquad \square$$

CHAPTER 5

Eigenvalues and eigenvectors

5.1. Eigenvalues and eigenvectors for matrices

Let F be a field and $A \in \mathrm{M}_n(F) = \mathrm{M}_{n\times n}(F)$ be a square matrix. We know that there is an associated linear transformation $f_A \colon F^n \longrightarrow F^n$ given by $f_A(\mathbf{x}) = A\mathbf{x}$. We can ask whether there are any vectors which are fixed by f_A, *i.e.*, which satisfy the equation

$$f_A(\mathbf{x}) = \mathbf{x}.$$

Of course, $\mathbf{0}$ is always fixed so we should ask this question with the extra requirement that $\mathbf{x}$ be non-zero. Usually there will be no such vectors, but a slightly more general situation that might occur is that f_A will *dilate* some non-zero vectors, *i.e.*, $f_A(\mathbf{x}) = \lambda\mathbf{x}$ for some $\lambda \in F$ and non-zero vector $\mathbf{x}$. This amounts to saying that f_A sends all vectors in some line through the origin into that line. Of course, such a line is a 1-dimensional subspace of F^n.

EXAMPLE 5.1. Let $F = \mathbb{R}$ and consider the 2×2 matrix $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$.

(i) Show that the linear transformation $f_A \colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ does not fix any non-zero vectors.
(ii) Show thatf_A dilates all vectors on the x-axis by a factor of 2.
(iii) Find all the non-zero vectors which f_A dilates by a factor of 3.

SOLUTION.

(i) Suppose that a vector $\mathbf{x} \in \mathbb{R}^2$ satisfies $A\mathbf{x} = \mathbf{x}$, then it also satisfies the homogenous equation

$$(I_2 - A)\mathbf{x} = \mathbf{0}, \qquad i.e., \quad \begin{bmatrix} -1 & -1 \\ 0 & -2 \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

But the only solution of this is $\mathbf{x} = \mathbf{0}$.

(ii) We have

$$\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix} = \begin{bmatrix} 2x \\ 0 \end{bmatrix} = 2 \begin{bmatrix} x \\ 0 \end{bmatrix},$$

so vectors on the x-axis get dilated under f_A by a factor of 2.

(iii) Suppose that $A\mathbf{x} = 3\mathbf{x}$, then

$$(3I_2 - A)\mathbf{x} = \mathbf{0}, \qquad i.e., \quad \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

The set of all solution vectors of this is the line $L = \{(t,t) : t \in \mathbb{R}\}$. $\square$

DEFINITION 5.2. Let F be a field and $A \in \mathrm{M}_n(F)$. Then $\lambda \in F$ is an *eigenvalue* for A if there is a non-zero vector $\mathbf{v} \in F^n$ for which $Av = \lambda v$; such a vector v is called an *eigenvalue* associated with λ.

Definition 5.2 crucially depends on the field F as the following example shows.

EXAMPLE 5.3. Let $A = \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$.

(i) Show that A has no eigenvalues in $\mathbb{Q}$.
(ii) Show that A has two eigenvalues in $\mathbb{R}$.

SOLUTION. Notice that

$$A^2 - 2I_2 = O_2,$$

so if λ is a real eigenvalue with associated eigenvector $\mathbf{v}$, then

$$(A^2 - 2I_2)\mathbf{v} = \mathbf{0},$$

which gives

$$(\lambda^2 - 2)\mathbf{v} = \mathbf{0}.$$

Thus we must have $\lambda^2 = 2$, hence $\lambda = \pm\sqrt{2}$. However, neither of $\pm\sqrt{2}$ is a rational number although both are real.

Working in $\mathbb{R}^2$ we have

$$A \begin{bmatrix} \sqrt{2} \\ 1 \end{bmatrix} = \sqrt{2} \begin{bmatrix} \sqrt{2} \\ 1 \end{bmatrix}, \quad A \begin{bmatrix} -\sqrt{2} \\ 1 \end{bmatrix} = -\sqrt{2} \begin{bmatrix} -\sqrt{2} \\ 1 \end{bmatrix}. \qquad \square$$

A similar problem occurs with the real matrix $\begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}$ which has the two imaginary eigenvalues $\pm i\sqrt{2}$.

Because of this it is preferable to work in an *algebraically closed field* such as the complex numbers. So from now on we will work with complex matrices and consider eigenvalues in $\mathbb{C}$ and eigenvectors in $\mathbb{C}^n$. When discussing real matrices we need to take care to check for real eigenvalues. This is covered by the next result which can be proved easily from the definition.

LEMMA 5.4. *Let A be an $n \times n$ real matrix which we can also view as a complex matrix. Suppose that $\lambda \in \mathbb{R}$ is an eigenvalue for which there is an associated eigenvector in $\mathbb{C}^n$. Then there is an associated eigenvector in $\mathbb{R}^n$.*

THEOREM 5.5. *Let A be an $n \times n$ complex matrix and let $\lambda \in \mathbb{C}$. Then λ is an eigenvalue of A if and only if λ is a root of the characteristic polynomial $\chi_A(t)$,* i.e., *$\chi_A(\lambda) = 0$.*

PROOF. From Definition 5.2, λ is an eigenvalue for A if and only if $\lambda I_n - A$ is singular, *i.e.*, is not invertible. By Proposition 4.5, this is equivalent to requiring that

$$\chi_A(\lambda) = \det(\lambda I_n - A) = 0. \qquad \square$$

5.2. Some useful facts about roots of polynomials

Recall that every complex (which includes real) polynomial $p(t) \in \mathbb{C}[t]$ of positive degree n admits a factorisation into linear polynomials

$$p(t) = c(t - \alpha_1) \cdots (t - \alpha_n),$$

where $c \neq 0$ and the roots α_k are unique apart from the order in which they occur. It is often useful to write

$$c(t - \beta_1)^{r_1} \cdots (t - \beta_m)^{r_m},$$

where $\beta_1, \ldots, \beta_m$ are the distinct complex roots of $p(t)$ and $r_k \geqslant 1$. Then r_k is the (*algebraic*) *multiplicity* of the root β_k. Again, this factorisation is unique apart form the order in which the distinct roots are listed.

When $p(t)$ is a real polynomial, *i.e.*, $p(t) \in \mathbb{R}[t]$, the complex roots fall into two types: *real roots* and *non-real roots* which come in complex conjugate pairs. Thus for such a real polynomial we have

$$p(t) = c(t - \alpha_1) \cdots (t - \alpha_r)(t - \gamma_1)(t - \overline{\gamma}_1) \cdots (t - \gamma_s)(t - \overline{\gamma}_s),$$

where $\alpha_1, \ldots, \alpha_r$ are the real roots (possibly with repetitions) and $\gamma_1, \overline{\gamma}_1 \ldots, \gamma_s, \overline{\gamma}_s$ are the non-real roots which occur in complex conjugate pairs $\gamma_k, \overline{\gamma}_j$. For a non-real complex number γ,

$$(t - \gamma)(t - \overline{\gamma}) = t^2 - (\gamma + \overline{\gamma})t + \gamma\overline{\gamma} = t^2 - (2 \operatorname{Re} \gamma)t + |\gamma|^2,$$

where $\operatorname{Re} \gamma$ is the real part of γ and $|\gamma|$ is the modulus of γ.

Proposition 5.6. *Let $p(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 \in \mathbb{C}[t]$ be a monic complex polynomial which factors completely as*

$$p(t) = (t - \alpha_1)(t - \alpha_2) \cdots (t - \alpha_n),$$

where $\alpha_1, \ldots, \alpha_n$ are the complex roots, occurring with multiplicities. The following formulae apply for the sum and product of these roots:

$$\sum_{i=1}^{n} \alpha_i = \alpha_1 + \cdots + \alpha_n = -a_{n-1}, \qquad \prod_{i=1}^{n} \alpha_i = \alpha_1 \cdots \alpha_n = (-1)^n a_0.$$

Example 5.7. Let $p(t)$ be a complex polynomial of degree 3. Suppose that

$$p(t) = t^3 - 6t^2 + at - 6,$$

where $a \in \mathbb{C}$, and also that 1 is a root of $p(t)$. Find the other roots of $p(t)$.

Solution. Suppose that the other roots are α and β. Then by Proposition 5.6,

$$-(\alpha + \beta + 1) = -6, \qquad \textit{i.e.}, \quad \beta = 5 - \alpha;$$

also,

$$(-1) \times 1 \times \alpha\beta = -6, \qquad \textit{i.e.}, \quad \alpha\beta = 6.$$

Hence we find that

$$(5 - \alpha)\alpha = 6, \qquad \textit{i.e.}, \quad \alpha^2 - 5\alpha + 6 = 0.$$

The roots of this are $2, 3$, therefore we obtain these as the remaining roots of $p(t)$. □

Of course we can use this when working with the characteristic polynomial of an $n \times n$ matrix A when the roots are the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A occurring with suitable multiplicities. If

$$\chi_A(t) = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0,$$

then by Proposition 4.13,

$$\lambda_1 + \cdots + \lambda_n = \operatorname{tr} A = -c_{n-1}, \tag{5.1}$$

$$\lambda_1 \cdots \lambda_n = \det A = (-1)^n c_0. \tag{5.2}$$

5.3. Eigenspaces and multiplicity of eigenvalues

In this section we will work over the field of complex numbers $\mathbb{C}$.

DEFINITION 5.8. Let A be an $n \times n$ complex matrix and let $\lambda \in \mathbb{C}$ be an eigenvalue of A. Then the *λ-eigenspace of A* is

$$\operatorname{Eig}_A(\lambda) = \{\mathbf{v} \in \mathbb{C}^n : A\mathbf{v} = \lambda\mathbf{v}\} = \{\mathbf{v} \in \mathbb{C}^n : (\lambda I_n - A)\mathbf{v} = \mathbf{0}\},$$

the set of all eigenvectors associated with λ together with the zero vector $\mathbf{0}$.

If A is a real matrix and if $\lambda \in \mathbb{R}$ is a real eigenvalue, then the *real λ-eigenspace of A* is

$$\operatorname{Eig}^{\mathbb{R}}_A(\lambda) = \{\mathbf{v} \in \mathbb{R}^n : A\mathbf{v} = \lambda\mathbf{v}\} = \operatorname{Eig}_A(\lambda) \cap \mathbb{R}^n.$$

PROPOSITION 5.9. *Let A be an $n \times n$ complex matrix and let $\lambda \in \mathbb{C}$ be an eigenvalue of A. Then $\operatorname{Eig}_A(\lambda)$ is a subspace of $\mathbb{C}^n$.*

If A is a real matrix and λ is a real eigenvalue, then $\operatorname{Eig}^{\mathbb{R}}_A(\lambda)$ is a subspace of the real vector space $\mathbb{R}^n$.

PROOF. By definition, $\mathbf{0} \in \operatorname{Eig}_A(\lambda)$. Now notice that if $\mathbf{u}, \mathbf{v} \in \operatorname{Eig}_A(\lambda)$ and $z, w \in \mathbb{C}$, then

$$\begin{aligned} A(z\mathbf{u} + w\mathbf{v}) &= A(z\mathbf{u}) + A(w\mathbf{v}) \\ &= zA\mathbf{u} + wA\mathbf{v} \\ &= z\lambda\mathbf{u} + w\lambda\mathbf{v} \\ &= \lambda(z\mathbf{u} + w\mathbf{v}). \end{aligned}$$

Hence $\operatorname{Eig}_A(\lambda)$ is a subspace of $\mathbb{C}^n$.

The proof in the real case easily follows from the fact that $\operatorname{Eig}^{\mathbb{R}}_A(\lambda) = \operatorname{Eig}_A(\lambda) \cap \mathbb{R}^n$. □

DEFINITION 5.10. Let A be an $n \times n$ complex matrix and let $\lambda \in \mathbb{C}$ be an eigenvalue of A. The *geometric multiplicity* of λ is $\dim_{\mathbb{C}} \operatorname{Eig}_A(\lambda)$.

Note that $\dim_{\mathbb{C}} \operatorname{Eig}_A(\lambda) \geqslant 1$. In fact there are some other constraints on geometric multiplicity which we will not prove.

THEOREM 5.11. *Let A be an $n \times n$ complex matrix and let $\lambda \in \mathbb{C}$ be an eigenvalue of A.*
(i) *Let the algebraic multiplicity of λ in the characteristic polynomial $\chi_A(t)$ be r_λ. Then*

$$\text{geometric multiplicity of } \lambda = \dim_{\mathbb{C}} \operatorname{Eig}_A(\lambda) \leqslant r_\lambda.$$

(ii) *If A is a real matrix, then*

$$\dim_{\mathbb{C}} \operatorname{Eig}_A(\lambda) = \dim_{\mathbb{R}} \operatorname{Eig}^{\mathbb{R}}_A(\lambda).$$

EXAMPLE 5.12. For the real matrix $A = \begin{bmatrix} 2 & 1 & 0 \\ -1 & 0 & 1 \\ 2 & 0 & 0 \end{bmatrix}$, determine its complex eigenvalues. For each eigenvalue λ, find the corresponding eigenspace $\operatorname{Eig}_A(\lambda)$. For each real eigenvalue λ, find $\operatorname{Eig}^{\mathbb{R}}_A(\lambda)$.

SOLUTION. We have

$$\begin{aligned} \chi_A(t) = \begin{vmatrix} t-2 & -1 & 0 \\ 1 & t & -1 \\ -2 & 0 & t \end{vmatrix} &= (t-2) \begin{vmatrix} t & -1 \\ 0 & t \end{vmatrix} + \begin{vmatrix} 1 & -1 \\ -2 & t \end{vmatrix} \\ &= (t-2)t^2 + (t-2) = (t-2)(t^2+1) = (t-2)(t-i)(t+i). \end{aligned}$$

So the complex roots of $\chi_A(t)$ are $2, i, -i$ and there is one real root 2. Now we can find the three eigenspaces.

$\mathrm{Eig}_A(2)$: Working over $\mathbb{C}$, we need to solve

$$(2I_3 - A)\mathbf{z} = \begin{bmatrix} 2-2 & -1 & 0 \\ 1 & 2 & -1 \\ -2 & 0 & 2 \end{bmatrix} \mathbf{z} = \mathbf{0},$$

i.e.,

$$\begin{bmatrix} 0 & -1 & 0 \\ 1 & 2 & -1 \\ -2 & 0 & 2 \end{bmatrix} \mathbf{z} = \mathbf{0}.$$

This is equivalent to the equation

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

in which the matrix is reduced echelon, and the general solution is $\mathbf{z} = (z, 0, z)$ $(z \in \mathbb{C})$. Hence,

$$\mathrm{Eig}_A(2) = \{(z, 0, z) \in \mathbb{C}^3 : z \in \mathbb{C}\}$$

and so $\dim_{\mathbb{C}} \mathrm{Eig}_A(2) = 1$. Since 2 is real, we can form

$$\mathrm{Eig}_A^{\mathbb{R}}(2) = \{(t, 0, t) \in \mathbb{R}^3 : t \in \mathbb{R}\}$$

which also has $\dim_{\mathbb{R}} \mathrm{Eig}_A^{\mathbb{R}}(2) = 1$.

$\mathrm{Eig}_A(i)$: Working over $\mathbb{C}$, we need to solve

$$(iI_3 - A)\mathbf{z} = \begin{bmatrix} i-2 & -1 & 0 \\ 1 & i & -1 \\ -2 & 0 & i \end{bmatrix} \mathbf{z} = \mathbf{0},$$

which is equivalent to the reduced echelon equation

$$\begin{bmatrix} 1 & 0 & -i/2 \\ 0 & 1 & (1+2i)/2 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{z} = \mathbf{0}.$$

This has general solution

$$(iz, -(1+2i)z, 2z) \quad (z \in \mathbb{C}),$$

so

$$\mathrm{Eig}_A(i) = \{(iz, -(1+2i)z, 2z) \in \mathbb{C}^3 : z \in \mathbb{C}\}$$

and $\dim_{\mathbb{C}} \mathrm{Eig}_A(i) = 1$.

$\mathrm{Eig}_A(i)$: Similarly, we have

$$\mathrm{Eig}_A(-i) = \{(-iz, -(1-2i)z, 2z) \in \mathbb{C}^3 : z \in \mathbb{C}\}$$

and $\dim_{\mathbb{C}} \mathrm{Eig}_A(-i) = 1$. □

Notice that in this example, each eigenvalue has geometric multiplicity equal to 1 which is also its algebraic multiplicity. Before giving a general result on this, here is an important observation.

PROPOSITION 5.13. *Let A be an $n \times n$ complex matrix.*
(i) *Suppose that $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues for A with associated eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is linearly independent.*
(ii) *The sum of the eigenspaces $\mathrm{Eig}_A(\lambda_1) + \cdots + \mathrm{Eig}_A(\lambda_k)$ is a direct sum,* i.e.,

$$\mathrm{Eig}_A(\lambda_1) + \cdots + \mathrm{Eig}_A(\lambda_k) = \mathrm{Eig}_A(\lambda_1) \oplus \cdots \oplus \mathrm{Eig}_A(\lambda_k).$$

PROOF.
(i) Choose the largest $r \leqslant k$ for which $\mathbf{v}_1, \ldots, \mathbf{v}_r$ is linearly independent. Then if $r < k$, $\mathbf{v}_1, \ldots, \mathbf{v}_{r+1}$ is linearly dependent. Suppose that for some $z_1, \ldots, z_{r+1} \in \mathbb{C}$, not all zero,

$$z_1\mathbf{v}_1 + \cdots + z_{r+1}\mathbf{v}_{r+1} = \mathbf{0}. \tag{5.3}$$

Multiplying by A we obtain

$$z_1 A\mathbf{v}_1 + \cdots + z_{r+1} A\mathbf{v}_{r+1} = \mathbf{0}$$

and hence

$$z_1\lambda_1\mathbf{v}_1 + \cdots + z_{r+1}\lambda_{r+1}\mathbf{v}_{r+1} = \mathbf{0}. \tag{5.4}$$

Now subtracting $\lambda_{r+1}\times$ (5.3) from (5.4) we obtain

$$z_1(\lambda_1 - \lambda_{r+1})\mathbf{v}_1 + \cdots + z_r(\lambda_r - \lambda_{r+1})\mathbf{v}_r + 0\mathbf{v}_{r+1} = \mathbf{0}.$$

But since $\lambda_1, \ldots, \lambda_r$ is linearly independent, this means that

$$z_1(\lambda_1 - \lambda_{r+1}) = \cdots = z_r(\lambda_r - \lambda_{r+1}) = 0,$$

and so since the λ_j are distinct,

$$z_1 = \cdots = z_r = 0.$$

hence we must have $z_{r+1}\mathbf{v}_{r+1} = \mathbf{0}$ which implies that $z_{r+1} = 0$ since $\mathbf{v}_{r+1} \neq \mathbf{0}$. So we must have $r = k$ and $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is linearly independent.
(ii) This follows by a similar argument to (i). $\square$

THEOREM 5.14. *Let A be an $n \times n$ complex matrix. Suppose that the distinct complex eigenvalues of A are $\lambda_1, \ldots, \lambda_\ell$ and that for each $j = 1, \ldots, \ell$,*

$$\text{geometric multiplicity of } \lambda_j = \text{algebraic multiplicity of } \lambda_j.$$

Then $\mathbb{C}^n$ is the direct sum of the eigenspaces $\mathrm{Eig}_A(\lambda_j)$, i.e.,

$$\mathbb{C}^n = \mathrm{Eig}_A(\lambda_1) \oplus \cdots \oplus \mathrm{Eig}_A(\lambda_\ell).$$

In particular, $\mathbb{C}^n$ has a basis consisting of eigenvectors of A.

EXAMPLE 5.15. Let $A = \begin{bmatrix} -1 & 2 & 2 \\ 2 & 2 & 2 \\ -3 & -6 & -6 \end{bmatrix}$. Find the eigenvalues of A and show that $\mathbb{C}^3$ has a basis consisting of eigenvectors of A. Show that the real vector space $\mathbb{R}^3$ also has a basis consisting of eigenvectors of A.

SOLUTION. The characteristic polynomial of A is

$$\begin{aligned}
\chi_A(t) &= \begin{vmatrix} t+1 & -2 & -2 \\ -2 & t-2 & -2 \\ 3 & 6 & t+6 \end{vmatrix} \\
&= \begin{vmatrix} t+1 & -2 & 0 \\ -2 & t-2 & -t \\ 3 & 6 & t \end{vmatrix} \\
&= (t+1) \begin{vmatrix} t-2 & -t \\ 6 & t \end{vmatrix} + 2 \begin{vmatrix} -2 & -t \\ 3 & t \end{vmatrix} \\
&= t \left((t+1) \begin{vmatrix} t-2 & -1 \\ 6 & 1 \end{vmatrix} + 2 \begin{vmatrix} -2 & -1 \\ 3 & 1 \end{vmatrix} \right) \\
&= t\big((t+1)(t+4)+2\big) = t(t^2+5t+6) = t(t+2)(t+3).
\end{aligned}$$

Thus the roots of $\chi_A(t)$ are $0, -2, -3$ which are all real numbers. Notice that these each have algebraic multiplicity 1.

$\mathrm{Eig}_A(0)$: We need to find the solutions of

$$(0I_3 - A) \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 1 & -2 & -2 \\ -2 & -2 & -2 \\ 3 & 6 & 6 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which is equivalent to

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and the general solution is $(u, v, w) = (0, z, -z) \in \mathbb{C}^3$. Thus a basis for $\mathrm{Eig}_A(0)$ is $(0, 1, -1)$ which is also a basis for the real vector space $\mathrm{Eig}_A^{\mathbb{R}}(0)$. So

$$\dim_{\mathbb{C}} \mathrm{Eig}_A(0) = \dim_{\mathbb{R}} \mathrm{Eig}_A^{\mathbb{R}}(0) = 1.$$

$\mathrm{Eig}_A(-2)$: We need to find the solutions of

$$((-2)I_3 - A) \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} -1 & -2 & -2 \\ -2 & -4 & -2 \\ 3 & 6 & 4 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which is equivalent to

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and the general solution is $(u, v, w) = (2z, -z, 0) \in \mathbb{C}^3$. Thus a basis for $\mathrm{Eig}_A(-2)$ is $(2, -1, 0)$ which is also a basis for the real vector space $\mathrm{Eig}_A^{\mathbb{R}}(-2)$. So

$$\dim_{\mathbb{C}} \mathrm{Eig}_A(-2) = \dim_{\mathbb{R}} \mathrm{Eig}_A^{\mathbb{R}}(-2) = 1.$$

$\mathrm{Eig}_A(-3)$: We need to find the solutions of

$$((-3)I_3 - A) \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} -2 & -2 & -2 \\ -2 & -5 & -2 \\ 3 & 6 & 3 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which is equivalent to

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

and the general solution is $(u, v, w) = (z, 0, -z) \in \mathbb{C}^3$. Thus a basis for $\operatorname{Eig}_A(-2)$ is $(1, 0, -1)$ which is also a basis for the real vector space $\operatorname{Eig}_A^{\mathbb{R}}(-3)$. So

$$\dim_{\mathbb{C}} \operatorname{Eig}_A(-3) = \dim_{\mathbb{R}} \operatorname{Eig}_A^{\mathbb{R}}(-3) = 1.$$

Then $(0, 1, -1), (2, -1, 0), (1, 0, -1)$ is a basis for the complex vector space $\mathbb{C}^3$ and for the real vector space $\mathbb{R}^3$. $\square$

EXAMPLE 5.16. Let $B = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 2 \\ -2 & -2 & 0 \end{bmatrix}$. Find the eigenvalues of B and show that $\mathbb{C}^3$ has a basis consisting of eigenvectors of B. For each real eigenvalue λ determine $\operatorname{Eig}_B^{\mathbb{R}}(\lambda)$.

SOLUTION. The characteristic polynomial of B is

$$\chi_B(t) = \begin{vmatrix} t & -1 & -2 \\ 1 & t & -2 \\ 2 & 2 & t \end{vmatrix} = t^3 + 9t = t(t^2 + 9) = t(t - 3i)(t + 3i),$$

so the eigenvalues are 0, $3i$, $-3i$. We find that

$$\begin{aligned} \operatorname{Eig}_B(0) &= \{(2z, -2z, z) : z \in \mathbb{C}\}, \\ \operatorname{Eig}_B(3i) &= \{((1 + 3i)z, (-1 + 3i)z, -4z) : z \in \mathbb{C}\}, \\ \operatorname{Eig}_B(-3i) &= \{((1 - 3i)z, (-1 - 3i)z, -4z)) : z \in \mathbb{C}\}. \end{aligned}$$

For the real eigenvalue 0,

$$\operatorname{Eig}_B^{\mathbb{R}}(0) = \{(2t, -2t, t) : t \in \mathbb{R}\}.$$

Note that vectors $(2t, -2t, t)$ with $t \in \mathbb{R}$ and $t \neq 0$ are the only real eigenvectors of B.

Then $(2, -2, 1), (1 + 3i, -1 + 3i, -4), (1 - 3i, -1 - 3i, -4)$ is a basis for the complex vector space $\mathbb{C}^3$. $\square$

EXAMPLE 5.17. Let $A = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}$. Find the eigenvalues of A and show that $\mathbb{C}^3$ has a basis consisting of eigenvectors of A. For each real eigenvalue λ determine $\operatorname{Eig}_A^{\mathbb{R}}(\lambda)$. Show that $\mathbb{C}^3$ has no basis consisting of eigenvalues for A.

SOLUTION. The characteristic polynomial of A is

$$\chi_A(t) = \begin{vmatrix} t & 0 & 1 \\ 0 & t - 2 & 0 \\ -1 & 0 & t - 2 \end{vmatrix} = (t - 1)^2(t - 2),$$

hence the eigenvalues are 1 with algebraic multiplicity 2, and 2 with algebraic multiplicity 1.

$\operatorname{Eig}_A(1)$: We have to solve the equation

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

which is equivalent to

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

This gives

$$\mathrm{Eig}_A(1) = \{(z, 0, -z) : z \in \mathbb{C}\}.$$

Notice that $\dim \mathrm{Eig}_A(1) = 1$ which is less than the algebraic multiplicity of this eigenvalue.
$\mathrm{Eig}_A(2)$: We have to solve the equation

$$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

which is equivalent to

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

This gives

$$\mathrm{Eig}_A(2) = \{(0, z, 0) : z \in \mathbb{C}\}.$$

Thus we obtain

$$\begin{aligned} \dim(\mathrm{Eig}_A(1) + \mathrm{Eig}_A(2)) &= \dim \mathrm{Eig}_A(1) + \dim \mathrm{Eig}_A(2) - \dim \mathrm{Eig}_A(1) \cap \mathrm{Eig}_A(2) \\ &= 1 + 1 - 0 = 2. \end{aligned}$$

So the eigenvalues do not span $\mathbb{C}^3$.

Similarly,

$$\mathrm{Eig}_A^{\mathbb{R}}(1) = \{(s, 0, -s) : s \in \mathbb{R}\}, \quad \mathrm{Eig}_A^{\mathbb{R}}(2) = \{(0, t, 0) : t \in \mathbb{R}\},$$

and

$$\begin{aligned} \dim_{\mathbb{R}}(\mathrm{Eig}_A^{\mathbb{R}}(1) + \mathrm{Eig}_A^{\mathbb{R}}(2)) &= \dim_{\mathbb{R}} \mathrm{Eig}_A^{\mathbb{R}}(1) + \dim_{\mathbb{R}} \mathrm{Eig}_A^{\mathbb{R}}(2) - \dim_{\mathbb{R}} \mathrm{Eig}_A^{\mathbb{R}}(1) \cap \mathrm{Eig}_A^{\mathbb{R}}(2) \\ &= 1 + 1 - 0 = 2. \end{aligned}$$ □

This shows that eigenvalues cannot span $\mathbb{C}^3$.

These examples show that when the algebraic multiplicity of an eigenvalue is greater than 1, its geometric multiplicity need not be the same as its algebraic multiplicity.

5.4. Diagonalisability of square matrices

Now we will consider the implications of the last section for *diagonalisability* of square matrices.

DEFINITION 5.18. Let $A, B \in \mathrm{M}_n(F)$ be $n \times n$ matrices with entries in a field F. Then A *is similar to* B if there is an invertible $n \times n$ matrix $P \in \mathrm{M}_n(F)$ for which

$$B = P^{-1}AP.$$

REMARK 5.19. Notice that if $B = P^{-1}AP$ then

$$A = PBP^{-1} = (P^{-1})^{-1}BP^{-1},$$

so B is also similar to A. If we take $P = I_n = I_n^{-1}$, then

$$A = I_n^{-1}AI_n,$$

so A is similar to itself. Finally, if A is similar to B (say $B = P^{-1}AP$) and B is similar to C (say $C = Q^{-1}BQ$), then

$$C = Q^{-1}(P^{-1}AP)Q = (Q^{-1}P^{-1})A(PQ) = (PQ)^{-1}A(PQ),$$

hence A is similar to C. These three observations show that the notion of similarity defines an *equivalence relation* on $\mathrm{M}_n(F)$.

DEFINITION 5.20. Let $\lambda_1, \ldots, \lambda_n \in F$ be scalars. The *diagonal matrix* $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is the $n \times n$ matrix

$$\operatorname{diag}(\lambda_1, \ldots, \lambda_n) = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

with the scalars $\lambda_1, \ldots, \lambda_n$ down the main diagonal and 0 everywhere else. An $n \times n$ matrix $A \in \mathrm{M}_n(F)$ is *diagonalisable* (over F) if it is similar to a diagonal matrix, *i.e.*, if there is an invertible matrix P for which $P^{-1}AP$ is a diagonal matrix.

Now we can ask the question: When is a square matrix similar to a diagonal matrix? This is answered by our next result.

THEOREM 5.21. *A matrix $A \in \mathrm{M}_n(F)$ is diagonalisable over F if and only F^n has a basis consisting of eigenvectors of A.*

PROOF. If F^n has a basis of eigenvectors of A, say $\mathbf{v}_1, \ldots, \mathbf{v}_n$ with associated eigenvalues $\lambda_1, \ldots, \lambda_n$, let P be the invertible matrix with these vectors as its columns. We have

$$\begin{aligned} AP &= \begin{bmatrix} A\mathbf{v}_1 & \cdots & A\mathbf{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\mathbf{v}_1 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix} \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \\ &= P \operatorname{diag}(\lambda_1, \ldots, \lambda_n), \end{aligned}$$

from which we obtain

$$P^{-1}AP = \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

hence A is similar to the diagonal matrix $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. On the other hand, if A is similar to a diagonal matrix, say

$$P^{-1}AP = \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

then

$$AP = P \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

and the columns of P are eigenvectors associated with the eigenvalues $\lambda_1, \ldots, \lambda_n$ and they also form a basis of F^n. $\square$

The process of finding a diagonal matrix similar to a given matrix is called *diagonalisation*. It usually works best over $\mathbb{C}$, but sometimes a real matrix can be diagonalised over $\mathbb{R}$ if all its eigenvalues are real.

EXAMPLE 5.22. Diagonalise the real matrix

$$A = \begin{bmatrix} 2 & -2 & 3 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{bmatrix}.$$

SOLUTION. First we find the complex eigenvalues of A. The characteristic polynomial is

$$\begin{aligned}
\chi_A(t) &= \begin{vmatrix} t-2 & 2 & -3 \\ -1 & t-1 & -1 \\ -1 & -3 & t+1 \end{vmatrix} \\
&= \begin{vmatrix} t-2 & 2 & -3 \\ 0 & t+2 & -t-2 \\ -1 & -3 & t+1 \end{vmatrix} \\
&= (t+2) \begin{vmatrix} t-2 & 2 & -3 \\ 0 & 1 & -1 \\ -1 & -3 & t+1 \end{vmatrix} \\
&= (t+2) \left((t-2) \begin{vmatrix} 1 & -1 \\ -3 & t+1 \end{vmatrix} - \begin{vmatrix} 2 & -3 \\ 1 & -1 \end{vmatrix} \right) \\
&= (t+2) \left((t-2)(t+1-3) - (-2+3) \right) \\
&= (t+2) \left((t-2)^2 - 1 \right) = (t+2)(t^2-4t+4-1) = (t+2)(t^2-4t+3) \\
&= (t+2)(t-3)(t-1).
\end{aligned}$$

Hence the roots of $\chi_A(t)$ are the real numbers $-2, 1, 3$ and these are the eigenvalues of A.

Now solving the equation $((-2)I_3 - A)\mathbf{x} = \mathbf{0}$ we find that the general real solution is $\mathbf{x} = (11t, t, -14t)$ with $t \in \mathbb{R}$. So an eigenvector for the eigenvalue -2 is $(11, 1, -14)$.

Solving $(I_3 - A)\mathbf{x} = \mathbf{0}$ we find that the general real solution is $\mathbf{x} = (t, -t, -t)$ with $t \in \mathbb{R}$. So an eigenvector for the eigenvalue 1 is $(1, -1, -1)$.

Solving $(3I_3 - A)\mathbf{x} = \mathbf{0}$ we find that the general real solution is $\mathbf{x} = (t, t, t)$ with $t \in \mathbb{R}$. So an eigenvector for the eigenvalue 3 is $(1, 1, 1)$.

The vectors $(11, 1, -14), (1, -1, -1), (1, 1, 1)$ form a basis for the real vector space $\mathbb{R}^3$ and the matrix $P = \begin{bmatrix} 11 & 1 & 1 \\ 1 & -1 & 1 \\ -14 & -1 & 1 \end{bmatrix}$ satisfies

$$A \begin{bmatrix} 11 & 1 & 1 \\ 1 & -1 & 1 \\ -14 & -1 & 1 \end{bmatrix} = \begin{bmatrix} -22 & 1 & 3 \\ -2 & -1 & 3 \\ 28 & -1 & 3 \end{bmatrix} = \begin{bmatrix} 11 & 1 & 1 \\ 1 & -1 & 1 \\ -14 & -1 & 1 \end{bmatrix} \operatorname{diag}(-2, 1, 3),$$

hence

$$AP = P \operatorname{diag}(-2, 1, 3).$$

Therefore $\operatorname{diag}(-2, 1, 3) = P^{-1}AP$, showing that A is diagonalisable. $\square$

APPENDIX A

Complex solutions of linear ordinary differential equations

It often makes sense to look for complex valued solutions of differential equations of the form

$$\frac{d^n f}{dx^n} + a_{n-1}\frac{d^{n-1} f}{dx^{n-1}} + \cdots + a_1\frac{df}{dx} + a_0 f = 0, \tag{A.1}$$

where the a_r are real or complex numbers which do not depend on x. The solutions are functions defined on $\mathbb{R}$ and taking values in $\mathbb{C}$. The number n is called the *degree* of the differential equation.

When $n = 1$, for any complex number a, the differential equation

$$\frac{df}{dx} + af = 0$$

has as its solutions all the functions of the form te^{-ax}, for $t \in \mathbb{C}$. Here we recall that for a complex number $z = x + yi$ with $x, y \in \mathbb{R}$,

$$e^z = e^{x+yi} = e^x e^{yi} = e^x(\cos y + i \sin y) = e^x \cos y + ie^x \sin y.$$

When $n = 2$, for any complex numbers a, b, we consider the differential equation

$$\frac{d^2 f}{dx^2} + a\frac{df}{dx} + bf = 0, \tag{A.2}$$

where we assume that the polynomial $z^2 + az + b$ has the complex roots α, β, which might be equal.

- If $\alpha \neq \beta$, then (A.2) has as its solutions all the functions of the form $se^{\alpha x} + tse^{\beta x}$, for $s, t \in \mathbb{C}$.
- If $\alpha = \beta$, then (A.2) has as its solutions all the functions of the form $se^{\alpha x} + txe^{\alpha x}$, for $s, t \in \mathbb{C}$.

When a, b are real, then either both roots are real or there are two complex conjugate roots $\alpha, \overline{\alpha}$; suppose that $\alpha = u + vi$ where $u, v \in \mathbb{R}$, hence the complex conjugate of α is $\overline{\alpha} = u - vi$. In this case we can write the solutions in the form

- $pe^{ux} \cos vx + qe^{ux} \sin vx$ for $p, q \in \mathbb{C}$ if $v \neq 0$,
- $pe^{ux} + qxe^{ux}$ for $p, q \in \mathbb{C}$ if $v = 0$.

In either case, to get real solutions we have to take $p, q \in \mathbb{R}$.

Bibliography

[BLA] T. S. Blyth & E. F. Robertson, Basic Linear Algebra, Springer-Verlag (1998), ISBN 3540761225.

[FLA] T. S. Blyth & E. F. Robertson, Further Linear Algebra, Springer-Verlag (2002), ISBN 1852334258.

[LADR] S. Axler, Linear Algebra Done Right, 2nd edition, corrected 7th printing, Springer-Verlag (2004), ISBN 0387982582.

[Schaum] S. Lipschutz & M. Lipson, Schaum's Outline of Linear algebra, McGraw-Hill (2000), ISBN 0071362002.

[Strang] G. Strang, Linear Algebra and Its Applications, 4th edition, BrooksCole, ISBN 0534422004.