

University of Melbourne  
Department of Mathematics and Statistics

Notes for 620-222:  
Linear and Abstract Algebra  
Semester 2, 2002

## **Abstract**

These lecture notes were compiled in the Department of Mathematics and Statistics in the University of Melbourne for the use of students in the subject 620-222. Copyright J. R. J. Groves, C. D. Hodgson, 2002.

# Contents

<b>1</b>	<b>Linear Algebra</b>	<b>5</b>
1.1	Fields . . . . .	5
1.1.1	Definition and examples of fields . . . . .	5
1.1.2	The integers modulo a prime . . . . .	7
1.1.3	Algebraically closed fields . . . . .	9
1.1.4	Exercises . . . . .	9
1.2	Revision . . . . .	10
1.2.1	Vector spaces and subspaces . . . . .	10
1.2.2	Spanning, linear dependence, bases . . . . .	13
1.2.3	Linear transformations . . . . .	16
1.2.4	Matrix representations . . . . .	18
1.2.5	Change of basis . . . . .	19
1.2.6	Exercises . . . . .	21
1.3	Normal forms . . . . .	23
1.3.1	Eigenvalues and eigenspaces, invariant subspaces . . . . .	23
1.3.2	Minimal polynomials. . . . .	27
1.3.3	Triangular form and the Cayley-Hamilton Theorem . . . . .	31
1.3.4	Jordan normal form . . . . .	34
1.3.5	Exercises . . . . .	38
1.4	Inner product spaces . . . . .	40
1.4.1	Complex inner products . . . . .	40
1.4.2	Orthogonal complements . . . . .	46
1.4.3	Adjoints, self-adjoint, Hermitian, normal . . . . .	46
1.4.4	Exercises . . . . .	51
1.5	The Spectral Theorem and applications . . . . .	53
1.5.1	The Theorem . . . . .	53
1.5.2	Polar form . . . . .	55
1.5.3	Commuting normal matrices . . . . .	57
1.5.4	Exercises . . . . .	58

<b>2</b>	<b>Groups</b>	<b>61</b>
2.1	Symmetries . . . . .	61
2.1.1	Exercises . . . . .	63
2.2	What groups are, and some examples . . . . .	64
2.2.1	Definition of a group . . . . .	64
2.2.2	Examples of groups . . . . .	65
2.2.3	Matrix groups . . . . .	66
2.2.4	Groups with at most 4 elements . . . . .	67
2.2.5	Permutations . . . . .	68
2.2.6	Exercises . . . . .	71
2.3	Group discussion; some terminology for groups . . . . .	72
2.3.1	Subgroups . . . . .	72
2.3.2	Cyclic subgroups . . . . .	74
2.3.3	Isomorphism . . . . .	76
2.3.4	Products of groups . . . . .	78
2.3.5	Exercises . . . . .	80
2.4	Lagrange's Theorem . . . . .	81
2.4.1	Cosets . . . . .	81
2.4.2	The Theorem . . . . .	83
2.4.3	Some applications . . . . .	84
2.4.4	Exercises . . . . .	84
2.5	Quotient groups . . . . .	86
2.5.1	Normal subgroups . . . . .	86
2.5.2	Trivialising subgroups . . . . .	88
2.5.3	Homomorphisms . . . . .	89
2.5.4	Exercises . . . . .	91
2.6	Groups that can act . . . . .	93
2.6.1	Definition and examples . . . . .	93
2.6.2	Orbits and stabilizers . . . . .	94
2.6.3	Exercises . . . . .	96
2.7	Groups acting on themselves . . . . .	97
2.7.1	Left multiplication . . . . .	97
2.7.2	Conjugation . . . . .	99
2.7.3	Some consequences for group theory . . . . .	100
2.7.4	Exercises . . . . .	103
2.8	Keeping our distance; isometries . . . . .	104
2.8.1	Definition and basic properties . . . . .	104
2.8.2	What isometries are there? . . . . .	106
2.8.3	Classification of finite symmetry groups . . . . .	109
2.8.4	Isometries of Euclidean space $\mathbb{E}^n$ . . . . .	110
2.8.5	Exercises . . . . .	112

2.9	Wallpaper groups . . . . .	112
2.9.1	Lattices and point groups . . . . .	113
2.9.2	Simple properties of wallpaper groups . . . . .	115
2.9.3	Exercises . . . . .	116
2.10	Classifying wallpaper groups . . . . .	118
2.10.1	The point groups and the lattices . . . . .	118
2.10.2	Understanding the rotations in $W$ . . . . .	119
2.10.3	Classifying the cases when the point group contains a reflection . . . . .	120
2.10.4	The case when $\pi(W)$ is $D_4$ and $W$ contains a corresponding reflection . . . . .	120
2.10.5	The case when $\pi(W)$ is $D_4$ and $W$ does not contain a corresponding reflection . . . . .	121
2.10.6	Summary of the classification . . . . .	122
2.10.7	Exercises . . . . .	123
<b>3</b>	<b>Hints and answers to Exercises</b>	<b>125</b>
3.1	Linear Algebra . . . . .	125
3.1.1	Exercises in Section 1.1.4 on page 9 . . . . .	125
3.1.2	Exercises in Section 1.2.6 on page 21 . . . . .	126
3.1.3	Exercises in Section 1.3.5 on page 38 . . . . .	127
3.1.4	Exercises in Section 1.4.4 on page 51 . . . . .	129
3.1.5	Exercises in Section 1.5.4 on page 58 . . . . .	132
3.2	Groups . . . . .	133
3.2.1	Exercises in Section 2.1.1 on page 63 . . . . .	133
3.2.2	Exercises in Section 2.2.6 on page 71 . . . . .	134
3.2.3	Exercises in Section 2.3.5 on page 80 . . . . .	135
3.2.4	Exercises in Section 2.4.4 on page 84 . . . . .	137
3.2.5	Exercises in Section 2.5.4 on page 91 . . . . .	138
3.2.6	Exercises in Section 2.6.3 on page 96 . . . . .	140
3.2.7	Exercises in Section 2.7.4 on page 103 . . . . .	140
3.2.8	Exercises in Section 2.8.5 on page 112 . . . . .	143
3.2.9	Exercises in Section 2.9.3 on page 116 . . . . .	144
3.2.10	Exercises in Section 2.10.7 on page 123 . . . . .	146
<b>4</b>	<b>Old Examinations etc.</b>	<b>147</b>
4.1	The 1998 Mid-Semester test . . . . .	147
4.2	The 1999 Mid-Semester test . . . . .	149
4.3	The 2000 Mid-Semester test . . . . .	151
4.4	Solutions to the 1998 Mid-Semester test . . . . .	153
4.5	Solutions to the 1999 Mid-Semester test . . . . .	155

4.6	Solutions to the 2000 Mid-Semester test . . . . .	157
4.7	The 1997 examination . . . . .	159
4.8	The 1998 examination . . . . .	163
4.9	The 1999 examination . . . . .	168
4.10	The 2000 examination . . . . .	172
4.11	Solutions to the 1997 Examination . . . . .	176
4.12	Solutions to the 1998 Examination . . . . .	181
4.13	Solutions to the 1999 Examination . . . . .	187
4.14	Solutions to the 2000 Examination . . . . .	191

# Chapter 1

## Linear Algebra

### 1.1 Fields

It is likely that, so far in your study of linear algebra, you have seen very few examples of vector spaces where the scalars are not the real numbers. But vector spaces, and matrices, have very wide application in mathematics and the physical sciences and the ideas are frequently needed in a context where the scalars are other than the real numbers, or even the complex numbers.

#### 1.1.1 Definition and examples of fields

Roughly speaking, a field is a mathematical system where the notions of addition, multiplication, subtraction and division, make sense in the same way that they do for real numbers. But that is clearly not a precise enough definition for mathematical purposes.

**Definition 1.1.1.** *A field  $F$  consists of a set with two operations, called addition ‘+’ and multiplication ‘ $\times$ ’ (thus if  $a, b \in F$  then  $a + b$  and  $a \times b$  are well defined elements of  $F$ ) satisfying the following:*

#### Properties of addition

- (1)  $a + (b + c) = (a + b) + c$  for all  $a, b, c \in F$ ;
- (2) there is an element  $0 \in F$  satisfying  $0 + a = a + 0 = a$  for all  $a \in F$ ;
- (3) for all  $a \in F$ , there is an element  $-a \in F$  satisfying  
 $a + (-a) = (-a) + a = 0$ ;
- (4)  $a + b = b + a$  for all  $a, b \in F$ ;

**Properties of multiplication**

- (5)  $a \times (b \times c) = (a \times b) \times c$  for all  $a, b, c \in F$ ;
- (6) there is an element  $1 \in F$  with  $1 \neq 0$  satisfying  $1 \times a = a \times 1 = a$  for all  $a \in F$ ;
- (7) for all  $a \in F$  with  $a \neq 0$ , there is an element  $a^{-1} \in F$  satisfying  $a \times (a^{-1}) = (a^{-1}) \times a = 1$ ;
- (8)  $a \times b = b \times a$  for all  $a, b \in F$ ;

**Connecting addition and multiplication**

- (9)  $a \times (b + c) = (a \times b) + (a \times c)$  for all  $a, b, c \in F$ .

In practice, we often write  $ab$  or  $a.b$  rather than  $a \times b$ .

This is a lot of rules (axioms). If in doubt, remember the examples and the non-examples, to give you an idea of what a field is.

**Examples:**

- (1) The real numbers  $\mathbb{R}$ ;
- (2) The complex numbers  $\mathbb{C}$ ;
- (3) The rational numbers  $\mathbb{Q}$ ;
- (4) The collection of all expressions  $p(x)/q(x)$  where  $p(x)$  and  $q(x)$  are polynomials in  $x$  with real coefficients and  $q(x)$  is not the zero polynomial. This is called the *field of rational functions* (with coefficients from  $\mathbb{R}$ ).
- (5) The set of integers with the usual addition and multiplication does **not** give us a field.
- (6) Let  $F$  have two elements  $\{0, 1\}$  and the following addition and multiplication tables.

+	0	1	×	0	1
0	0	1	0	0	0
1	1	0	1	0	1

This is the simplest case of a very important example of fields which we deal with in the next subsection.

### 1.1.2 The integers modulo a prime

Many of you will have seen ‘modulo arithmetic’. We want to put this idea on a slightly firmer foundation so that we can use ‘numbers modulo a prime’ as scalars for a vector space. We will begin by considering arithmetic modulo any natural number.

The fundamental idea of ‘arithmetic modulo  $n$ ’ is that we regard two numbers that differ by a multiple of  $n$  as being essentially the same. If  $a - b$  is a multiple of  $n$  we often write  $a \equiv b \pmod{n}$  and say that “ $a$  is congruent to  $b$  modulo  $n$ ”.

The classic example is that two periods of time, when displayed on a (12 hour) clock, are indistinguishable if they differ by a multiple of 12. The technical way to achieve this identification of all numbers which differ by a multiple of  $n$  is to make a set which contains *all* numbers which differ from a fixed number by a multiple of  $n$ .

**Definition 1.1.2.** *Let  $n$  be a natural number. If  $a$  is an integer, then the collection of all integers  $b$  such that  $b - a$  is an integral multiple of  $n$  is denoted  $[a]_n$ . The set of all possible sets  $[a]_n$  (for fixed  $n$  but variable  $a$ ) is denoted  $\mathbb{Z}_n$  and is called the integers modulo  $n$ .*

**Example:** The integers modulo 3,  $\mathbb{Z}_3$ , has 3 elements:

$$\begin{aligned} [0]_3 &= \{ \dots, -9, -6, -3, 0, 3, 6, 9, 12, \dots \} \\ [1]_3 &= \{ \dots, -8, -5, -2, 1, 4, 7, 10, 13, \dots \} \\ [2]_3 &= \{ \dots, -7, -4, -1, 2, 5, 8, 11, 14, \dots \}. \end{aligned}$$

Note that each of the three sets has infinitely many descriptions in the form  $[l]_3$ . For example,

$$[0]_3 = [-9]_3 = [12]_3 = [-3174]_3 \text{ and } [2]_3 = [-7]_3 = [8]_3 = [477287]_3.$$

We can use the equality sign here because the *sets* represented by the square brackets are equal. In general,  $[a]_n = [b]_n$  if and only if  $a \equiv b \pmod{n}$ . Thus  $\mathbb{Z}_n$  has  $n$  elements  $[0]_n, [1]_n, \dots, [n-1]_n$ .

We want to be able to carry the normal arithmetic operations over to  $\mathbb{Z}_n$ . There is really only one sensible way to do it.

**Definition 1.1.3.** *For  $[a]_n, [b]_n \in \mathbb{Z}_n$ , define addition and multiplication by*

$$[a]_n + [b]_n = [a + b]_n \quad \text{and} \quad [a]_n \times [b]_n = [ab]_n.$$

There is a problem, however. We know that  $[0]_3 = [-3174]_3$  and  $[2]_3 = [477287]_3$ . Our definitions insist that  $[0]_3 + [2]_3 = [2]_3$  and that  $[-3174]_3 + [477287]_3 = [474113]_3$ . Is it true that  $[2]_3 = [474113]_3$ ? The answer is ‘yes’.

**Lemma 1.1.1.** *If  $a, b, c, d \in \mathbb{Z}$  and  $[a]_n = [b]_n$  and  $[c]_n = [d]_n$  then*

$$[a + c]_n = [b + d]_n \quad \text{and} \quad [ac]_n = [bd]_n.$$

So we have a ‘well-defined’ addition and multiplication on  $\mathbb{Z}_n$ . We also have subtraction via  $[a]_n - [b]_n = [a - b]_n$ . We should not really expect division since we cannot always divide one integer by another to obtain a third integer.

**Example:** In  $\mathbb{Z}_4$ , the element  $[2]_4$  has no multiplicative inverse  $[a]_4$  such that  $[2]_4 \times [a]_4 = [1]_4$ . So  $\mathbb{Z}_4$  is *not* a field.

But in some important special cases, we do get division and so a field.

**Lemma 1.1.2.** *If  $p$  is a prime and  $[a]_p \neq [0]_p$  then we can find an integer  $c$  such that  $[a]_p \times [c]_p = [1]_p$ .*

*Proof.* Consider all possible products of  $[a]_p$  with the elements of  $\mathbb{Z}_p$ :

$$\{[a \cdot 0]_p, [a \cdot 1]_p, \dots, [a \cdot (p - 1)]_p\}. \quad (*)$$

We claim that these are all **different**: If  $[ai]_p = [aj]_p$  then  $[ai - aj]_p = [0]_p$  and so  $p$  divides  $ai - aj = a(i - j)$ .

**Fact:** If  $p$  is prime and  $p$  divides  $cd$  then  $p$  divides  $c$  or  $p$  divides  $d$ .

As  $p$  does not divide  $a$ , it follows that  $p$  divides  $(i - j)$ . Hence  $[i]_p = [j]_p$  and  $i = j$ .

Thus the list (\*) consists of  $p$  different elements of  $\mathbb{Z}_p$  (in some order), so contains *all* the elements of  $\mathbb{Z}_p$ . In particular  $[1]_p$  is included in this list. So  $[1]_p = [ai]_p = [a]_p \times [i]_p$  for some  $i$ . This proves the lemma.  $\square$

**Notation:** We usually write  $a|b$  if ‘ $a$  divides  $b$ ’, i.e.  $b$  is an exact multiple of  $a$ .

**Theorem 1.1.3.** *Let  $p$  be a prime. With the given operations of addition and multiplication given above,  $\mathbb{Z}_p$  is a field.*

*Proof.* This is mostly routine checking using the properties of addition and multiplication for the integers. The two hardest parts are checking that the operations are well-defined and that division is possible. The former is Lemma 1.1.1 and the latter follows quickly from Lemma 1.1.2.  $\square$

**Exercise:** Show that if  $n$  is not a prime, then  $\mathbb{Z}_n$  is not a field.

### 1.1.3 Algebraically closed fields

Often we'll be interested in solving polynomial equations in a field. There is a special property of the complex numbers which we will sometimes use and which is worth isolating.

**Definition 1.1.4.** *A field  $F$  is said to be algebraically closed if every non-constant polynomial with coefficients in  $F$  has a root in  $F$ . In other words, if  $p(x) = a_0 + a_1x + \dots + a_nx^n$  with  $a_0, \dots, a_n \in F$ ,  $n \geq 1$  and  $a_n \neq 0$ , then there exists  $x_0 \in F$  such that  $p(x_0) = 0$ .*

We can give little in the way of examples. It is true, but not easy to prove, that the complex numbers are algebraically closed. The real numbers are not algebraically closed because, for example,  $x^2 + 1$  is a polynomial with coefficients in  $\mathbb{R}$  which has no root in  $\mathbb{R}$ . Apart from the definition above, we can summarise what we need in the following theorem.

**Theorem 1.1.4.** (1)  $\mathbb{C}$  is algebraically closed.

(2) Every field lies inside an algebraically closed field.

### 1.1.4 Exercises

- (1) Show that the set of all real numbers of the form  $a + b\sqrt{2}$  with  $a, b \in \mathbb{Q}$  forms a field with the usual operations of addition and multiplication of the real numbers. (This is a *subfield* of  $\mathbb{R}$ .)
- (2) Show that the set of all real numbers of the form  $a + b\sqrt[3]{2}$  with  $a, b \in \mathbb{Q}$  does not form a field with the usual operations of addition and multiplication of the real numbers. Is there a way to make a field, similar to the previous example, but which contains  $\sqrt[3]{2}$  as well as the rational numbers?
- (3) Write down the multiplication table for  $\mathbb{Z}_7$ . Find an element  $a$  of  $\mathbb{Z}_7$  so that every non-zero element of  $\mathbb{Z}_7$  is a power of  $a$ .
- (4) Show that  $\mathbb{Z}_9$ , with the usual operations of addition and multiplication modulo 9, does not form a field.
- (5) Show that the set of all polynomials, with coefficients from the real numbers, does not form a field.
- (6) (Harder) Let  $\mathbb{C}((t))$  denote the set of all power series of the form

$$c_{-k}t^{-k} + c_{-k+1}t^{-k+1} + \dots + c_0 + c_1t + \dots + c_s t^s + \dots$$

with the usual operations of addition and multiplication of power series. Show that  $\mathbb{C}((t))$  forms a field. You should ignore the question of whether the power series are convergent.

- (7) Show that the field in Exercise 1 is not algebraically closed.
- (8) (Harder) Show that, for every prime  $p$ ,  $\mathbb{Z}_p$  is not algebraically closed.

## 1.2 Revision

### 1.2.1 Vector spaces and subspaces

We begin with the formal definition of vector spaces.

**Definition 1.2.1.** *Let  $F$  be a field. A vector space over  $F$  consists of a set  $V$  together with two operations, addition of the form  $V \times V \rightarrow V$  and scalar multiplication of the form  $F \times V \rightarrow V$ . These satisfy the following axioms:*

**Properties of addition:**

- (1)  $u + (v + w) = (u + v) + w$  for all  $u, v, w \in V$ ;
- (2) there is an element  $0 \in V$  satisfying  $0 + v = v + 0 = v$  for all  $v \in V$ ;
- (3) for each  $v \in V$ , there is an element  $-v \in V$  such that  $v + (-v) = (-v) + v = 0$ ;
- (4)  $u + v = v + u$  for all  $u, v \in V$ ;

**Properties of scalar multiplication:**

- (5)  $a(u + v) = au + av$  for all  $a \in F, u, v \in V$ ;
- (6)  $(a + b)v = av + bv$  for all  $a, b \in F, v \in V$ ;
- (7)  $(ab)v = a(bv)$  for all  $a, b \in F, v \in V$ ;
- (8)  $1v = v$  for all  $v \in V$ .

**Examples:**

- (1) Set  $F = \mathbb{R}$  and  $V = \{(x, y, z) : x, y, z \in \mathbb{R}\}$  with addition and scalar multiplication defined by:

$$(x, y, z) + (x', y', z') = (x+x', y+y', z+z') \quad \text{and} \quad c \cdot (x, y, z) = (cx, cy, cz).$$

This is the standard vector space  $\mathbb{R}^3$ .

- (2) Let  $F$  be an arbitrary field and  $V = \{(a_1, a_2, \dots, a_n) : a_1, \dots, a_n \in F\}$  with addition and scalar multiplication defined by:

$$\begin{aligned} (a_1, \dots, a_n) + (b_1, \dots, b_n) &= (a_1 + b_1, \dots, a_n + b_n) \quad \text{and} \\ c \cdot (a_1, \dots, a_n) &= (ca_1, \dots, ca_n) \quad \text{for all } c \in F. \end{aligned}$$

Denote this vector space by  $F^n$ . The first example is a special case of this.

- (3) Let  $F = \mathbb{R}$  and let  $M_{m \times n}(\mathbb{R})$  denote the set of  $m \times n$  matrices with entries from  $\mathbb{R}$ . Then  $M_{m \times n}(\mathbb{R})$ , furnished with the usual addition and scalar multiplication of matrices, is a vector space. This example also works when we replace  $\mathbb{R}$  by a general field.
- (4) Let  $F$  be a field. Then the set of polynomials with coefficients in  $F$ , together with the usual addition and scalar multiplication of polynomials, form a vector space  $\mathcal{P}(F)$ .
- (5) As the previous example, but consider only polynomials of degree at most  $n$ , for some fixed natural number  $n$ . Call the resulting space  $\mathcal{P}_n(F)$ .
- (6) The set  $\mathbb{R}^{\mathbb{R}} = \mathcal{F}(\mathbb{R}, \mathbb{R})$  of all functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  forms a vector space over the field of real numbers. Addition of two such functions  $f$  and  $g$  is given by:

$$f + g \text{ is the function defined by } (f + g) : x \mapsto f(x) + g(x)$$

and scalar multiplication, for  $a \in \mathbb{R}$  is given by:

$$af \text{ is the function defined by } (af) : x \mapsto af(x).$$

- (7) As the previous example, but allow the set  $F^S = \mathcal{F}(S, F)$  of functions  $f : S \rightarrow F$  where  $S$  is an arbitrary set and  $F$  is a field. This is then a vector space over  $F$ .

- (8) The set of solutions  $y$  of the differential equation

$$\frac{d^2y}{dx^2} + 7\frac{dy}{dx} + 23y = 0$$

forms a vector space if we use the addition and scalar multiplication of functions defined above.

- (9) Let  $F = \mathbb{R}$  and let  $V = \mathbb{R}^\infty$  be the set of all sequences  $\{a_n\}$ ,  $a_n \in \mathbb{R}$ . Define addition and scalar multiplication by:

$$\{a_n\} + \{b_n\} = \{a_n + b_n\} \quad \text{and} \quad x\{a_n\} = \{xa_n\}.$$

Note that this is really a special case of Example 7 since we can regard such a sequence as a function  $\mathbb{N} \rightarrow \mathbb{R}$ .

- (10) As above but restrict to sequences which satisfy  $\lim_{n \rightarrow \infty} a_n = 0$ .
- (11) If we repeat as in Example 9 but restrict to sequences which satisfy  $\lim_{n \rightarrow \infty} a_n = 1$  then we do **not** obtain a vector space.

**Definition 1.2.2.** *Let  $V$  be a vector space over the field  $F$ . A subspace of  $V$  is a subset  $W$  of  $V$  which is itself a vector space using the operations of addition and scalar multiplication from  $V$ .*

If we take a subset of  $\mathbb{R}^3$ , say  $\{(a, b, c) : a, b, c \in \mathbb{R}, a + b + c = 0\}$  and start checking whether it is a subspace, we find that many of the checks are essentially trivial. Briefly, we know that the operations behave well because the overlying space, in this case  $\mathbb{R}^3$ , is a vector space. When we eliminate all of the things we don't need to check for this reason, we are left with the following.

**Lemma 1.2.1.** *Let  $V$  be a vector space over  $F$ . A subset  $W$  of  $V$  is a subspace if and only if the following three conditions are satisfied:*

- (1)  $W$  is non-empty;
- (2) if  $u, w \in W$  then  $u + w \in W$ ;
- (3) if  $a \in F$  and  $w \in W$  then  $aw \in W$ .

**Examples:**

- (1) Set  $W = \{(a, b, c) : a, b, c \in \mathbb{R}, a + b + c = 0\}$ . Then  $W$  is a subspace of  $\mathbb{R}^3$ .
- (2) The set of matrices of trace zero is a subspace of the vector space  $M_{n \times n}(\mathbb{R})$ .
- (3) The set of polynomials with zero constant term is a subspace of the space  $\mathcal{P}(\mathbb{R})$ .
- (4) The set of differentiable functions is a subspace of the space  $\mathbb{R}^{\mathbb{R}} = \mathcal{F}(\mathbb{R}, \mathbb{R})$ .
- (5) The set of sequences with  $\lim_{n \rightarrow \infty} a_n = 0$  is a subspace of the space of all sequences.

**1.2.2 Spanning, linear dependence, bases**

**Definition 1.2.3.** *If  $S$  is a subset of a vector space  $V$  then a linear combination of  $S$  is an finite sum of the form*

$$\sum_{i=1}^n a_i s_i \quad \text{where } a_i \in F, s_i \in S.$$

*The set of all linear combinations of elements of  $S$  is called the span of  $S$  and is denoted by  $\langle S \rangle$ . We also say that  $S$  is a spanning set for  $\langle S \rangle$ .*

**Lemma 1.2.2.** *If  $S$  is a non-empty subset of  $V$ , then  $\langle S \rangle$  is a subspace of  $V$ .*

**Examples:**

- (1) The set of all linear combinations of the vectors  $(1, -2, 3)$  and  $(0, 2, 1)$  in  $\mathbb{R}^3$  is the set  $\{(a, -2a + 2b, 3a + b) : a, b \in \mathbb{R}\}$ .
- (2) The set of all linear combinations of the matrices

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

in  $M_{3 \times 3}(\mathbb{R})$  is the set of all matrices of the form

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

where  $a, b, c \in \mathbb{R}$ .

**Definition 1.2.4.** We say that a subset  $S$  of a vector space  $V$  is linearly dependent if some non-zero linear combination gives the zero vector:

$$\sum_{i=1}^n a_i s_i = 0 \quad \text{where } a_i \in F, s_i \in S \text{ and not all } a_i \text{ are zero.}$$

Otherwise,  $S$  is linearly independent.

**Examples:**

- (1) The set  $\{(1, 2, 3), (2, -1, 0), (-1, 8, 9)\}$  is linearly dependent in  $\mathbb{R}^3$ .
- (2) The set  $\{1, x, x^2, 1 + x^3\}$  is linearly independent in  $\mathcal{P}(\mathbb{R})$ .
- (3) The set  $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 2 & -29 \\ 0 & 0 \end{bmatrix} \right\}$  is linearly dependent in  $M_{2 \times 2}(\mathbb{R})$ .

**Lemma 1.2.3.** A subset  $S$  of a vector space  $V$  is linearly dependent if and only if, some element  $s$  of  $S$  is a linear combination of the others.

In this case removing  $s$  from  $S$  gives a *smaller* spanning set for the subspace  $\langle S \rangle$ . Making the spanning set as small as possible leads to the idea of basis.

**Definition 1.2.5.** A basis for a vector space  $V$  is a linearly independent spanning set.

**Examples:**

- (1) The standard basis for  $F^n$  is the set  $\{e_1 = (1, 0, 0, \dots, 0), e_2 = (0, 1, 0, \dots, 0), \dots, e_n = (0, 0, \dots, 1)\}$ .
- (2) The set  $\{(2, 1, 3), (1, 2, 3), (1, 0, 0)\}$  is a basis of  $\mathbb{R}^3$ .
- (3) The set  $\{1, x, x^2, 1 + x^3\}$  is a basis of  $\mathcal{P}_3(\mathbb{R})$ .
- (4) The set  $\{1, x, x^2, x^3, x^4, \dots, x^n, \dots\}$  is a basis of  $\mathcal{P}(\mathbb{R})$ .

**Theorem 1.2.4.** Every vector space has a basis. In fact, every spanning set contains a basis and every linearly independent set can be extended to a basis.

**Theorem 1.2.5.** If  $\mathcal{B}_1$  and  $\mathcal{B}_2$  are two bases of a vector space then they have the same number of elements. (In general, this means that there exists a function  $f : \mathcal{B}_1 \rightarrow \mathcal{B}_2$  which is 1-1 and onto, i.e. a bijection.)

**Definition 1.2.6.** The dimension of a vector space  $V$  is the number of elements in a basis. We usually write this as  $\dim V$ .

By Theorem 1.2.5, we know that this number will not depend on the particular choice of basis.

**Examples:** For the examples after Definition 1.2.1:

- (1)  $\mathbb{R}^3$  has dimension 3.
- (2)  $F^n$  has dimension  $n$ .
- (3)  $M_{m \times n}(\mathbb{R})$  has dimension  $mn$ .
- (4)  $\mathcal{P}_n(F)$  has dimension  $n + 1$ .
- (5) Example 8 has dimension 2 (although this needs a bit of work).
- (6) All of the other examples have infinite dimension.

**Combining subspaces:** Let  $U$  and  $W$  be subspaces of a vector space  $V$ . Then the *intersection*  $U \cap W = \{v \in V : v \in U \text{ and } v \in W\}$  and the *sum*  $U + W = \{u + w : u \in U, w \in W\}$  are both *subspaces* of  $V$ . (See exercises 1, 2.) In fact  $U + W$  is the smallest subspace containing both  $U$  and  $W$ .

**Lemma 1.2.6.** *Let  $U$  and  $W$  be subspaces of a vector space  $V$  and assume that  $U + W$  is finite dimensional. Then*

$$\dim(U + W) + \dim(U \cap W) = \dim U + \dim W.$$

*Proof.* Let  $\{v_1, \dots, v_l\}$  be a basis of  $U \cap W$ . Then  $\{v_1, \dots, v_l\}$  is a linearly independent set in  $U$  and so can be extended to a basis  $\{v_1, \dots, v_l, u_1, \dots, u_m\}$  of  $U$ . Similarly  $\{v_1, \dots, v_l\}$  can be extended to a basis  $\{v_1, \dots, v_l, w_1, \dots, w_n\}$  of  $W$ . We claim that  $\{v_1, \dots, v_l, u_1, \dots, u_m, w_1, \dots, w_n\}$  is a basis of  $U + W$ .

Since every element of  $U$  is a linear combination of  $\{v_1, \dots, v_l, u_1, \dots, u_m\}$  and every element of  $W$  is a linear combination of

$$\{v_1, \dots, v_l, w_1, \dots, w_n\},$$

it is clear that the sum of an element of  $U$  and an element of  $W$  is a linear combination of

$$\begin{aligned} \{v_1, \dots, v_l, u_1, \dots, u_m\} \cup \{v_1, \dots, v_l, w_1, \dots, w_n\} \\ = \{v_1, \dots, v_l, u_1, \dots, u_m, w_1, \dots, w_n\}. \end{aligned} \quad (1.1)$$

So  $\{v_1, \dots, v_l, u_1, \dots, u_m, w_1, \dots, w_n\}$  spans  $U + W$ .

Suppose that we have

$$\sum_i a_i v_i + \sum_j b_j u_j + \sum_k c_k w_k = 0 \text{ with } a_i, b_j, c_k \in F.$$

Then  $\sum_k c_k w_k$  is a linear combination of elements of  $U$  and so lies in  $U \cap W$ . Thus  $\sum_k c_k w_k$  can be written as a linear combination of the basis  $\{v_1, \dots, v_l\}$  of  $U \cap W$ . Thus we have

$$\sum_k c_k w_k = \sum_i d_i v_i \text{ for some } d_i \in F.$$

But

$$\{v_1, \dots, v_l, w_1, \dots, w_n\}$$

is a basis of  $W$  and so linearly independent. Thus each  $c_k$  and each  $d_i$  is zero. Now we have  $\sum_i a_i v_i + \sum_j b_j u_j = 0$ . But  $\{v_1, \dots, v_l, u_1, \dots, u_m\}$  is a basis of  $U$  and so linearly independent. Thus each  $a_i$  and  $b_j$  is zero. Hence

$$\{v_1, \dots, v_l, u_1, \dots, u_m, w_1, \dots, w_n\}$$

is linearly independent and so is a basis for  $U + W$ .

We now have  $\dim(U \cap W) = l$ ,  $\dim U = l + m$ ,  $\dim W = l + n$  and  $\dim(U + W) = l + m + n$ . The result follows immediately.  $\square$

### 1.2.3 Linear transformations

Informally, a linear transformation is a function between vector spaces over the same field which preserves the operations of addition and scalar multiplication.

**Definition 1.2.7.** *Let  $V$  and  $W$  be vector spaces over the same field  $F$ . A function  $f : V \rightarrow W$  is a linear transformation if*

- (1)  $f(u + v) = f(u) + f(v)$  for all  $u, v \in V$ ;
- (2)  $f(av) = af(v)$  for all  $a \in F, v \in V$ .

Note: Taking  $a = 0$  in (2) shows that  $f(\mathbf{0}) = \mathbf{0}$ .

**Examples:**

- (1) Rotation about the origin through a fixed angle  $\theta$  is a linear transformation on  $\mathbb{R}^2$ .
- (2) Rotation about any line through the origin and through a fixed angle  $\theta$  is a linear transformation on  $\mathbb{R}^3$ .
- (3) Differentiation is a linear transformation on  $\mathcal{P}(\mathbb{R})$ .

- (4) Let  $\mathcal{C}$  denote the subspace of  $\mathbb{R}^{\mathbb{R}}$  consisting of continuous functions. Let the function  $I : \mathcal{C} \rightarrow \mathcal{C}$  be given by defining  $I(f)$  to be the function whose value at  $t$  is

$$I(f)[t] = \int_0^t f(x) dx.$$

Then  $I$  is a linear transformation.

- (5) The functions  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  where  $f(x) = x^2$  and  $g(x) = x + 2$  are *not* linear transformations.

**Definition 1.2.8.** Let  $f : V \rightarrow W$  be a linear transformation. The nullspace (or kernel) of  $f$  is  $\{v \in V : f(v) = 0\}$ . The range (or image) of  $f$  is  $\{f(v) : v \in V\}$ .

**Examples:**

- (1) Rotation in  $\mathbb{R}^2$  has nullspace  $0$  and range the whole of  $\mathbb{R}^2$ .  
 (2) Differentiation on  $\mathcal{P}(\mathbb{R})$  has nullspace  $\langle 1 \rangle$  and range  $\mathcal{P}(\mathbb{R})$ .

The following should not be too surprising, or too hard to prove.

**Lemma 1.2.7.** Let  $f : V \rightarrow W$  be a linear transformation. The nullspace of  $f$  is a subspace of  $V$  and the range of  $f$  is a subspace of  $W$ .

**Definition 1.2.9.** Let  $f : V \rightarrow W$  be a linear transformation. The dimension of the nullspace of  $f$  is called the nullity of  $f$  and the dimension of the range of  $f$  is called the rank of  $f$ .

**Lemma 1.2.8.** Let  $f : V \rightarrow W$  be a linear transformation and assume that  $V$  is finite dimensional. The nullity of  $f$  plus the rank of  $f$  is equal to the dimension of  $V$ .

*Sketch of proof.* Denote the nullspace of  $f$  by  $N$ . Since it is a subspace of  $V$  it will have a basis  $\mathcal{B} = \{v_1, \dots, v_m\}$ . So  $m$  is the nullity of  $f$ . Since  $\mathcal{B}$  is a basis of  $N$ , it is linearly independent in  $N$ . Since  $N$  is a subspace of  $V$ ,  $\mathcal{B}$  is also linearly independent in  $V$ . So we can extend  $\mathcal{B}$  to a basis of  $\{v_1, \dots, v_m, v_{m+1}, \dots, v_n\}$  of  $V$ . So the dimension of  $V$  is  $n$ .

We claim that  $\{f(v_{m+1}), \dots, f(v_n)\}$  is a basis of the range of  $V$ . We must show that  $\{f(v_{m+1}), \dots, f(v_n)\}$  is linearly independent and that every element of the range of  $V$  can be expressed as a linear combination of  $\{f(v_{m+1}), \dots, f(v_n)\}$ . We leave the details as exercise 16.

We will have shown that  $f$  has nullity  $m$  and rank  $n - m$  where  $n$  is the dimension of  $V$ . The theorem now follows.  $\square$



**Interpretation of the matrix representation:**

Given a basis  $\mathcal{B}_V = \{v_1, v_2, \dots, v_m\}$  for a vector space  $V$ , each vector  $v \in V$  can be written *uniquely* as a linear combination

$$v = a_1 v_1 + \dots + a_m v_m, \quad \alpha_i \in F.$$

This allows us to introduce *coordinates* on  $V$ : the column vector

$$[v]_{\mathcal{B}_V} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} \in F^m$$

is called the *coordinate vector* of  $v$  with respect to the basis  $\mathcal{B}_V$ .

Then the effect a linear transformation  $f : V \rightarrow W$  on coordinate vectors is just multiplication by the matrix  $A$  representing  $f$ :

$$[f(v)]_{\mathcal{B}_W} = A[v]_{\mathcal{B}_V}.$$

In summary, we have

$$\begin{array}{ccc} v \in V & \xrightarrow{\text{apply } f} & f(v) \in W \\ \text{take coords} \downarrow & & \downarrow \text{take coords} \\ [v]_{\mathcal{B}_V} \in F^m & \xrightarrow{\text{mult by } A} & [f(v)]_{\mathcal{B}_W} \in F^n. \end{array}$$

**1.2.5 Change of basis**

Any linear transformation will have different matrices for different bases of the underlying vector spaces. It is very useful to be able to choose a basis so that the matrix is as simple as possible. To do this, we need to be able to see the effect on the matrix of changing the basis.

Let  $V, W$  be finite dimensional vector spaces over a field  $F$  and let  $f : V \rightarrow W$  be a linear transformation. Let  $\mathcal{B}_V = \{v_1, v_2, \dots, v_m\}$  be a basis for  $V$  and  $\mathcal{B}_W = \{w_1, w_2, \dots, w_n\}$  be a basis for  $W$ . Suppose that  $\mathcal{B}'_V = \{v'_1, v'_2, \dots, v'_m\}$  is a new basis for  $V$  and  $\mathcal{B}'_W = \{w'_1, w'_2, \dots, w'_n\}$  is a new basis for  $W$ . Then we can convert  $\mathcal{B}_V$ -coordinates to  $\mathcal{B}'_V$ -coordinates using the matrix  $P$  with  $i$ th column  $[v_i]_{\mathcal{B}'_V}$ . Similarly we can convert  $\mathcal{B}_W$ -coordinates to  $\mathcal{B}'_W$ -coordinates using the matrix  $Q$  with  $i$ th column  $[w_i]_{\mathcal{B}'_W}$ .

Explicitly,  $P = (p_{ij})$  and  $Q = (q_{ij})$  where

$$v_i = \sum_{j=1}^m p_{ji} v'_j \quad \text{and} \quad w_i = \sum_{j=1}^n q_{ji} w'_j.$$

**Theorem 1.2.9.** *The matrices  $P$  and  $Q$  are invertible and the matrix of  $f$  with respect to the bases  $\mathcal{B}'_V$  and  $\mathcal{B}'_W$  is*

$$QAP^{-1},$$

where  $A$  is the matrix of  $f$  with respect to the bases  $\mathcal{B}_V$  and  $\mathcal{B}_W$ .

Thus we have the following diagram:

$$\begin{array}{ccc} [v]_{\mathcal{B}_V} & \xrightarrow{A} & [f(v)]_{\mathcal{B}_W} \\ P \downarrow & & \downarrow Q \\ [v]_{\mathcal{B}'_V} & \xrightarrow{QAP^{-1}} & [f(v)]_{\mathcal{B}'_W}. \end{array}$$

In the most important case where  $V = W$  and  $\mathcal{B}_V = \mathcal{B}_W$ , we also have  $P = Q$  and so, if  $A$  is the matrix of  $f$  with respect to the old basis then  $PAP^{-1}$  is the matrix of  $f$  with respect to the new basis.

**Example:** Let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be the linear transformation defined by  $f(x, y) = (3x - y, -x + 3y)$ . Using the standard basis  $\mathcal{B} = \{(1, 0), (0, 1)\}$  we find the matrix of  $f$  is

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

Now let's calculate the matrix with respect to the basis  $\mathcal{B}' = \{(1, 1), (-1, 1)\}$ . We have

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

and

$$f(-1, 1) = (-4, 4) = 0(1, 1) + 4(-1, 1).$$

Thus the matrix for  $f$  with respect to basis  $\mathcal{B}'$  is the diagonal matrix

$$A' = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}.$$

This makes it easy to understand the effect of the transformation  $f$ : it just stretches by a factor 2 in the  $(1, 1)$  direction and by a factor 4 in the  $(-1, 1)$  direction.

Alternatively we can use the change of basis formula in the previous theorem. The transition matrix from  $\mathcal{B}'$  to the standard basis  $\mathcal{B}$  is  $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$  so the transition matrix from  $\mathcal{B}$  to  $\mathcal{B}'$  is the *inverse* of this:

$$P = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Then

$$A' = PAP^{-1} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix},$$

as before.

**Definition 1.2.11.** Two  $n \times n$  matrices  $A$  and  $B$  are said to be similar if  $B = PAP^{-1}$  for some invertible matrix  $P$ .

Thus similar matrices represent the same linear transformation with respect to different bases.

### 1.2.6 Exercises

- (1) If  $U$  and  $W$  are subspaces of a vector space  $V$ , show that  $U + W = \{u + w : u \in U, w \in W\}$  is also a subspace.
- (2) Show that, if  $U_1$  and  $U_2$  are subspaces of a vector space  $V$  then  $U_1 \cap U_2$  is also a subspace.
- (3) If  $U_1$  and  $U_2$  are subspaces of a vector space  $V$  and  $U_1 \cup U_2 = V$ , show that either  $U_1 = V$  or  $U_2 = V$ .
- (4) Decide whether the following sets of vectors are (i) linearly dependent and (ii) a basis, in  $\mathbb{Z}_7^4$ .
  - (a)  $\{([1]_7, [3]_7, [0]_7, [2]_7), ([2]_7, [1]_7, [3]_7, [0]_7)\}$ ;
  - (b)  $\{([1]_7, [2]_7, [3]_7, [1]_7), ([4]_7, [6]_7, [2]_7, [0]_7), ([0]_7, [1]_7, [5]_7, [1]_7)\}$ ;
  - (c)  $\{([1]_7, [2]_7, [3]_7, [1]_7), ([4]_7, [6]_7, [2]_7, [0]_7), ([0]_7, [1]_7, [5]_7, [2]_7), ([0]_7, [1]_7, [0]_7, [0]_7), ([0]_7, [1]_7, [0]_7, [1]_7)\}$ .
- (5) Decide whether the following sets of matrices are linearly independent in the space  $M_{2 \times 2}(\mathbb{R})$ :
  - (a)  $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$ ;
  - (b)  $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ ;
  - (c)  $\left\{ \begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 4 & -6 \\ 3 & 8 \end{bmatrix} \right\}$ .
- (6) Show that any subset of a linearly independent set is also linearly independent.

(7) Let  $F$  be a field and let  $E_{ij} \in M_{m \times n}(F)$  be the matrix with 1 in the  $i, j$  position and 0 elsewhere. Show that  $\{E_{i,j} : 1 \leq i \leq m, 1 \leq j \leq n\}$  is a basis of  $M_{m \times n}(F)$ .

(8) Show that the space  $\mathcal{P}(F)$  does not have finite dimension.

(9) What is the dimension of the space  $M_{3 \times 3}(\mathbb{Z}_5)$ ?

(10) Let  $B$  be the matrix  $\begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}$ . Show that the function  $g : M_{2 \times 2}(\mathbb{R}) \rightarrow M_{2 \times 2}(\mathbb{R})$  given by  $A \mapsto AB$  for  $A \in M_{2 \times 2}(\mathbb{R})$  is a linear transformation.

(11) Find the matrix of the linear transformation of Question 10 with respect to the basis

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

found in Question 7.

(12) Find the matrix, with respect to the standard basis of  $\mathbb{R}^2$ , of the reflection in the  $x$ -axis. Now let  $\mathcal{B}$  be the basis  $\{(a, b), (c, d)\}$ ,  $ad - bc \neq 0$  of  $\mathbb{R}^2$ . Write down a change of basis matrix for the change from the standard basis to  $\mathcal{B}$  and so calculate the matrix of the reflection with respect to this new basis.

(13) Calculate the nullity and rank of the linear transformation  $f$  on  $\mathbb{R}^3$  given by (here  $e_1, e_2, e_3$  is the standard basis)

$$f(e_1) = e_1 - e_2; f(e_2) = e_2 - e_3; f(e_3) = e_1 - e_3.$$

(14) Calculate the nullity and rank of the linear transformation  $f$  on  $\mathbb{Z}_7^3$  given by

$$\begin{aligned} f([1]_7, [0]_7, [0]_7) &= ([1]_7, [2]_7, [3]_7); \\ f([0]_7, [1]_7, [0]_7) &= ([3]_7, [4]_7, [5]_7); \\ f([0]_7, [0]_7, [1]_7) &= ([5]_7, [1]_7, [4]_7). \end{aligned}$$

(15) Let  $f : V \rightarrow V$  be a linear transformation on a finite dimensional vector space  $V$ . Show that the nullity of  $f$  is zero if and only if  $f$  is surjective.

(16) Complete the proof of Lemma 1.2.8

## 1.3 Normal forms

In this section, we shall consider a linear transformation  $f : V \rightarrow V$  on a vector space  $V$  and study the problem of finding a basis  $\mathcal{B}$  of  $V$  so that the matrix of  $f$  with respect to  $\mathcal{B}$  is as simple as possible. Often we can choose a *diagonal* matrix representing  $f$ , but in general the best we can find is the *Jordan normal form* (see section 1.3.4).

### 1.3.1 Eigenvalues and eigenspaces, invariant subspaces

**Definition 1.3.1.** *Suppose that  $f(v) = av$  for some non-zero  $v \in V$  and some  $a \in F$ . Then  $a$  is called an eigenvalue of  $f$  and  $v$  is said to be an eigenvector corresponding to  $a$ . The set of all solutions to the equation  $f(v) = av$  which correspond to a fixed eigenvalue  $a$  is a subspace of  $V$ , called the eigenspace corresponding to  $a$ .*

Similarly we define eigenvalues, eigenvectors and eigenspaces for any square matrix  $A$ .

**Examples:**

- (1) A rotation fixing the origin on  $\mathbb{R}^3$  has an eigenvalue of 1 and a corresponding eigenspace of dimension 1, the axis of the rotation.
- (2) A reflection fixing the origin on  $\mathbb{R}^2$  has two eigenvalues, 1 and -1. The eigenspace corresponding to 1 is the line of reflection. The eigenspace corresponding to -1 is the perpendicular to the line of reflection.
- (3) A reflection fixing the origin on  $\mathbb{R}^3$  has three eigenvalues, 1 (twice) and -1. The eigenspace corresponding to 1 is the plane of reflection. The eigenspace corresponding to -1 is the line perpendicular to the plane of reflection.
- (4) The eigenvalues of the (linear transformation corresponding to the) matrix

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

satisfy  $\det(A - \lambda I) = 0$ . So the eigenvalues are 2, -1, 0. The corresponding eigenspaces are generated by the eigenvectors

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix}, \begin{bmatrix} -7 \\ 8 \\ 2 \end{bmatrix}$$

respectively. Each eigenspace has dimension 1.

- (5) Let  $\mathcal{D}$  denote the subspace of functions in  $\mathbb{R}^{\mathbb{R}}$  which are differentiable infinitely often. Then differentiation gives a linear transformation  $\delta : \mathcal{D} \rightarrow \mathcal{D}$  and

$$\delta(e^{ax}) = \frac{d}{dx}e^{ax} = ae^{ax}.$$

So *every* real number  $a$  is an eigenvalue of  $\delta$  with corresponding eigenvector  $e^{ax}$ .

Observe that, if  $a$  is an eigenvalue and  $V_a$  is the corresponding eigenspace, then  $v \in V_a$  implies  $f(v) \in V_a$  (because  $f(v) = av$ ). Subspaces with this property are of special interest.

**Definition 1.3.2.** Let  $f$  be a linear transformation on a vector space  $V$  and let  $W$  be a subspace of  $V$ . We say that  $W$  is  $f$ -invariant if  $f(w) \in W$  for every  $w \in W$ .

If  $W$  is  $f$ -invariant then  $f$  defines a linear transformation  $f_W : W \rightarrow W$  called the *restriction* of  $f$  to  $W$  (forget all elements of  $V$  which are not in  $W$ ). Invariant subspaces are useful when we are trying to pick a ‘good’ basis with which to represent a linear transformation.

**Lemma 1.3.1.** Let  $f : V \rightarrow V$  be a linear transformation and let  $W$  be an  $f$ -invariant subspace, with  $\dim V = n$  and  $\dim W = m$ . Let  $\mathcal{B}_1 = \{w_1, \dots, w_m\}$  be a basis for  $W$ , and extend it to a basis  $\mathcal{B} = \{w_1, \dots, w_m, w_{m+1}, \dots, w_n\}$  for  $V$ . Then the matrix of  $f$  with respect to  $\mathcal{B}$  is of the “block” form

$$\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

where  $A, B, D$  are matrices and  $A$  is the  $m \times m$  matrix of  $f_W$ .

*Proof.* The first  $m$  columns represent the images of the elements of  $\mathcal{B}_1$ . These images lie in  $W$ , as  $W$  is  $f$ -invariant, and so can be expressed in terms of the elements of  $\mathcal{B}_1$  only. It follows that the first  $m$  columns have non-zero entries only in the first  $m$  rows.  $\square$

**Examples:**

- (1) Let  $f$  be a rotation on  $\mathbb{R}^3$ . Then the plane perpendicular to the axis of rotation is an invariant subspace of  $f$ . If we use two (orthonormal) vectors from this plane together with a unit vector along the axis of rotation, the matrix for the rotation becomes

$$\begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

- (2) Suppose that a linear transformation on  $\mathbb{R}^3$  has matrix

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

with respect to the basis  $\{e_1, e_2, e_3\}$ . Then the subspace  $W = \langle e_1, e_2 \rangle$  is  $f$ -invariant and the matrix of  $f_W$  with respect to the basis  $\{e_1, e_2\}$  of  $W$  is

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

In order to make more progress with this, we need the idea of a complement to a subspace.

**Definition 1.3.3.** *Let  $V$  be a vector space and let  $W$  be a subspace of  $V$ . Then a subspace  $U$  of  $V$  is a complement to  $W$  if  $U \cap W = \{0\}$  and  $U + W = V$ . We then write  $U \oplus W = V$ , and say that  $V$  is a direct sum of  $U$  and  $W$ .*

**Examples:**

- (1) In  $\mathbb{R}^3$ , a complement to a plane through the origin is any line through the origin which does not lie in the plane.
- (2) In  $\mathbb{R}^4$ , the subspaces  $\langle (1, 0, 0, 0), (0, 1, 0, 0) \rangle$  and  $\langle (0, 0, 1, 0), (0, 0, 0, 1) \rangle$  are complementary.
- (3) In  $\mathcal{P}(\mathbb{R})$ , the subspaces  $\langle 2, 1 + x, 1 + x + x^3 \rangle$  and  $\langle x^2 + 3x^4, x^4, x^5, x^6, \dots, x^n, \dots \rangle$  are complementary.

**Lemma 1.3.2.** *Let  $V$  be a finite dimensional vector space and let  $U, W$  be subspaces of  $V$ . The following are equivalent.*

- (1)  $U$  is a complement of  $W$ ;
- (2) There is a basis  $\mathcal{B}$  of  $V$  of the form  $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$  where  $\mathcal{B}_1$  is a basis of  $U$ ,  $\mathcal{B}_2$  is a basis of  $W$  and  $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$ .
- (3)  $U \cap W = \{0\}$  and  $\dim U = \dim V - \dim W$ ;
- (4)  $V = U + W$  and  $\dim U = \dim V - \dim W$ .

*Proof.* We shall make frequent use of Lemma 1.2.6.

(1) implies (2). Let  $\mathcal{B}_1$  be a basis of  $U$  and  $\mathcal{B}_2$  be a basis of  $W$ . Then it is easy to check that  $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$  and  $\mathcal{B}_1 \cup \mathcal{B}_2$  spans  $U + W = V$ . But  $\dim V = \dim(U + W) = \dim U + \dim W$  and so  $\mathcal{B}_1 \cup \mathcal{B}_2$  is a spanning set of  $V$  which has the same number of elements as a basis. It must therefore be a basis.

(2) implies (3). From (2) we can quickly deduce that  $\dim V = \dim U + \dim W$  and that  $V = U + W$ . Thus  $\dim(U \cap W) = \dim U + \dim W - \dim(U + W) = 0$ ; hence  $U \cap W = \{0\}$ .

(3) implies (4). Assuming (3) we have

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W) = \dim U + \dim W = \dim V.$$

Thus  $U + W = V$ .

(4) implies (1). Assuming (4) we have that  $V = U + W$ . Also,  $\dim(U \cap W) = \dim U + \dim W - \dim V = 0$  and so  $U \cap W = \{0\}$ .  $\square$

We can now obtain an even better version of Lemma 1.3.1.

**Lemma 1.3.3.** *Let  $V$  be a vector space and let  $f$  be a linear transformation on  $V$ . Let  $U, W$  be complementary subspaces of  $V$  (i.e.  $U \oplus W = V$ ). Suppose that both  $U$  and  $W$  are  $f$ -invariant. Choose an ordered basis  $\mathcal{B}$  of  $V$  of the form  $\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2)$  where  $\mathcal{B}_1$  is a basis of  $U$  and  $\mathcal{B}_2$  is a basis of  $W$ . Then the matrix of  $f$  with respect to  $\mathcal{B}$  is of the “block diagonal” form:*

$$\begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}$$

where  $A$  is the matrix of  $f_U$  and  $D$  is the matrix of  $f_W$ .

*Proof.* See Exercise 7.  $\square$

Thus we can now see a way to simplify the matrix for a linear transformation  $f$ . We must find complementary  $f$ -invariant subspaces. The next section will give us a way to do that.

### 1.3.2 Minimal polynomials.

Let  $V$  be a vector space of finite dimension  $n$  over a field  $F$ . Let  $f : V \rightarrow V$  be a linear transformation, and let  $A$  be a matrix representing  $f$  with respect to some basis. Given any polynomial

$$p(X) = a_0 + a_1X + \dots + a_kX^k, a_i \in F$$

we can apply it to the matrix  $A$ :

$$p(A) = a_0I + a_1A + a_2A^2 + \dots + a_kA^k,$$

where  $I$  is the  $n \times n$  identity matrix. Similarly we define

$$p(f) = a_01_V + a_1f + a_2f^2 + \dots + a_kf^k,$$

where  $1_V : V \rightarrow V$  is the identity transformation and  $f^k = f \circ \dots \circ f : V \rightarrow V$  is  $f$  composed with itself  $k$  times.

Now  $A$  lies in the vector space  $M_{n \times n}(F)$  of all  $n \times n$  matrices over  $F$ . The powers of  $A$ , which represent the powers of  $f$ , also lie in  $M_{n \times n}(F)$ . So  $\{I_n, A, A^2, \dots, A^k, \dots\}$  is an apparently infinite subset of the finite dimensional vector space  $M_{n \times n}(F)$ . Thus these powers must be linearly dependent and so there is some expression  $\sum_i a_i A^i = 0$  with  $a_i \in F$ , not all zero. There is therefore a similar expression  $\sum_i a_i f^i = 0$ . If we let  $q(X)$  be the polynomial  $\sum_i a_i X^i = 0$  we can write this as  $q(f) = 0$ . Observe that we can always divide through  $q(X)$  by the coefficient of the term of highest degree to ensure that, in the result, this coefficient is 1. We call such a polynomial *monic*. That is, a *monic polynomial* is one in which the term of highest degree has coefficient 1.

**Definition 1.3.4.** *Let  $f$  be a linear transformation on a vector space  $V$  of finite dimension. The minimal polynomial of  $f$  is the monic polynomial  $m(X)$  of lowest degree such that  $m(f) = 0$ .*

We define the minimal polynomial of a matrix  $A$  in a similar way.

#### Examples:

- (1) Let  $f$  denote a reflection, in a line through the origin, in  $\mathbb{R}^2$ . Then the minimal polynomial of  $f$  is  $X^2 - 1$ . It is clear that  $f^2$  is the identity on  $\mathbb{R}^2$ . Also, if the minimal polynomial were to have smaller degree, then it would be of degree 1 and  $f$  would be a scalar multiple of the identity, which is false.

(2) Let  $f$  be a linear transformation with matrix

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then the minimal polynomial of  $f$  is  $(X - 2)(X - 3)X$ .

(3) Let  $f$  be a linear transformation with matrix

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

Then the minimal polynomial of  $f$  is  $(X - 2)(X - 1)$ .

**Lemma 1.3.4.** *If  $m(X)$  is the minimal polynomial of  $f$  and if  $q(X)$  is a polynomial with coefficients in  $F$  such that  $q(f) = 0$  then  $m(x)$  divides  $q(X)$ .*

*Proof.* Use ‘division with remainder’ of polynomials to write

$$q(X) = s(X)m(X) + r(X)$$

where  $r(X)$  is a polynomial which is either 0 or is of degree less than the degree of  $m(X)$ . But then  $q(f) = s(f)m(f) + r(f)$ ; that is,  $0 = s(f)0 + r(f)$ . Thus  $r(f) = 0$ . But this will contradict the definition of  $m(X)$  unless  $r(X) = 0$ . So  $r(X) = 0$  and so  $q(X) = s(X)m(X)$ ; that is,  $m(X)$  divides  $q(X)$ .  $\square$

This helps a little in finding the minimal polynomial. If we can find some polynomial  $q(X)$  with  $q(f) = 0$  then we can look amongst its factors for the minimal polynomial. In fact we shall see later that the characteristic polynomial is a candidate for  $q(X)$ . The next lemma tells us that only factors big enough to have all of the eigenvalues as roots need be checked.

**Lemma 1.3.5.** *The roots of the minimal polynomial of  $f$  are precisely the eigenvalues of  $f$ .*

*Proof.* Let  $m(X)$  be the minimal polynomial of  $f$ .

Let  $a$  be an eigenvalue of  $f$  with corresponding eigenvector  $v \neq 0$ . Then  $f(v) = av$ . It is easily checked that  $f^k(v) = a^k v$  and so  $m(f)(v) = m(a)v$ . But  $m(f) = 0$  and so  $m(a)v = 0$ . Since  $v \neq 0$  we must have  $m(a) = 0$ ; that is,  $a$  is a root of  $m(X)$ .

Suppose, conversely, that  $a$  is a root of  $m(X)$ . Write  $m(X) = (X - a)^k p(X)$  where  $X - a$  does not divide  $p(X)$ . Since  $m(X)$  is the *minimal*

polynomial,  $p(f)$  is not the zero linear transformation on  $V$ . So there exists  $u \in V$  such that  $v = p(f)(u) \neq 0$ . Observe that

$$0 = m(f)(u) = (f - a)^k (p(f)(u)) = (f - a)^k (v).$$

Choose  $l$  least so that  $(f - a)^l (v) = 0$  and set  $w = (f - a)^{l-1} (v)$ . then  $w \neq 0$  and  $(f - a)(w) = 0$ . Hence  $f(w) = aw$ ; that is,  $a$  is an eigenvalue of  $f$ .  $\square$

Although the lemma tells us that the minimal polynomial has the eigenvalues as roots, it does not tell us about the multiplicity of these roots. The roots of the characteristic polynomial are also the eigenvalues, with possibly different multiplicity. We will come to the exact relationship between the two polynomials later.

The following result is an important way of constructing more invariant subspaces.

**Lemma 1.3.6.** *Let  $f : V \rightarrow V$  be a linear transformation on a vector space  $V$  over  $F$ . Let  $p(X)$  be any polynomial with coefficients in  $F$ . Then the null-space of  $p(f)$  is an  $f$ -invariant subspace of  $V$ .*

*Proof.* See Exercise 8.  $\square$

The following lemma, although rather technical, is the key step in decomposing a linear transformation into simple pieces.

**Lemma 1.3.7.** *Suppose that the minimal polynomial  $m(X)$  of  $f$  can be factored as a product  $m(X) = p(X)q(X)$  where  $p(X)$  and  $q(X)$  are polynomials, with coefficients from  $F$ , which have no common factor (except constants). Then  $V$  is a direct sum of  $f$ -invariant subspaces*

$$V = W_p \oplus W_q,$$

where  $W_p$  and  $W_q$  are the nullspaces of  $p(f)$  and  $q(f)$  respectively. Further, the restrictions  $f_{W_p}$  and  $f_{W_q}$  have minimal polynomials  $p(x)$  and  $q(x)$  respectively.

*Proof.* We will use the following **fact** that may not be familiar to everybody: If  $p(X)$  and  $q(X)$  are two polynomials which have no common factor (except constants), then there are polynomials  $k(X)$  and  $l(X)$  such that

$$k(X)p(X) + l(X)q(X) = 1.$$

(There is a similar statement if  $p$  and  $q$  are integers with no common factor. Either statement can be proved using the ‘Euclidean algorithm’ for finding the greatest common factor — based on repeated use of division with remainder. See 321 or 351 for the details.)

Lemma 1.3.6 tells us that  $W_p$  and  $W_q$  are  $f$ -invariant. We must show that  $V = W_p \oplus W_q$ .

(1) To show  $W_p \cap W_q = \{0\}$ : Suppose that  $v \in W_p \cap W_q$ . Because  $v \in W_p$ , we have that  $p(f)(v) = 0$ ; because  $v \in W_q$ , we have that  $q(f)(v) = 0$ . But, we also have

$$\text{id}_V = k(f)p(f) + l(f)q(f).$$

Thus

$$\begin{aligned} v &= (k(f)p(f) + l(f)q(f))(v) \\ &= k(f)(p(f)(v)) + l(f)(q(f)(v)) \\ &= k(f)(0) + l(f)(0) = 0 + 0 = 0 \end{aligned}$$

and so  $W_p \cap W_q = \{0\}$ .

(2) To show  $W_p + W_q = V$ : Suppose now that  $v \in V$ . Then

$$\text{id}_V = p(f)k(f) + q(f)l(f)$$

so

$$v = p(f)(k(f)(v)) + q(f)(l(f)(v)) = p(f)(u) + q(f)(w)$$

where we have put  $u = k(f)(v)$  and  $w = l(f)(v)$ . But  $q(f)(p(f)(u)) = m(f)(u) = 0$  and so  $p(f)(u) \in W_q$ . Similarly,  $q(f)(w) \in W_p$ . Thus

$$v = p(f)(u) + q(f)(w) \in W_q + W_p$$

and so  $V = W_q + W_p$ . Thus  $W_q \oplus W_p = V$ .

(3) Restricting  $f$  to  $W_p$  we see that  $p(f_{W_p}) = p(f)_{W_p} = 0$ . We must show that no polynomial  $p_1(X)$  of degree smaller than the degree of  $p(X)$  satisfies  $p_1(f_{W_p}) = 0$ . As in the last paragraph we have that, for every  $v \in V$ ,  $p(f)(q(f)(v)) = m(f)(v) = 0$  and so  $q(f)(v) \in W_p$ . If  $p_1(f_{W_p}) = 0$  for some polynomial  $p_1(X)$  of degree smaller than that of  $p(X)$ , then we would have that  $p_1(f)(q(f)(v)) = 0$  for all  $v \in V$ . That is, setting  $m_1(X) = p_1(X)q(X)$ ,  $m_1(f) = 0$ . But since  $p_1(X)$  has degree smaller than that of  $p(X)$ ,  $m_1(X)$  would have degree smaller than that of  $m(X)$  which is not possible. So  $p(X)$  is the minimal polynomial of  $f_{W_p}$  and, similarly,  $q(X)$  is the minimal polynomial of  $f_{W_q}$ .  $\square$

Combining the previous lemmas now shows us how to choose a basis so that the matrix has ‘block diagonal’ form.

**Theorem 1.3.8.** *Let  $f$  be a linear transformation on a finite dimensional vector space. Suppose that the minimal polynomial of  $f$  takes the form  $m(X) = q_1(X) \dots q_k(X)$  where  $q_i(X)$  has no common factor with  $q_j(X)$  if*

$i \neq j$ . Let  $W_i$  be the nullspace of  $q_i(f)$ . Suppose that  $\mathcal{B}_i$  is an ordered basis for  $W_i$ . Then

$$\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_k)$$

is an ordered basis for  $V$  and the matrix of  $f$  with respect to this basis is of the form

$$\begin{bmatrix} A_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A_k \end{bmatrix}$$

where  $A_i$  is the matrix of  $f_{W_i}$  with respect to  $\mathcal{B}_i$ .

*Proof.* We combine the result of Lemma 1.3.7 with the result of Lemma 1.3.3, using induction on  $k$ .  $\square$

The final step in our argument is to work out the simplest possibilities for the diagonal blocks  $A_i$  that are guaranteed by this last theorem.

### 1.3.3 Triangular form and the Cayley-Hamilton Theorem

We have seen in Theorem 1.3.8 that any linear transformation on a finite dimensional vector space can be represented by a matrix in ‘block diagonal’ form. We are therefore left with the problem of understanding these diagonal blocks or, equivalently, linear transformations which have a minimal polynomial which is a power of an *irreducible* polynomial (i.e. one with no lower degree factors).

We shall now **assume**, for this subsection and the next, that **the field  $F$  of scalars is algebraically closed**; for example,  $F$  could be the field  $\mathbb{C}$  of complex numbers.

Then an irreducible polynomial is of the form  $X - a$  for some  $a \in F$  and we want to consider minimal polynomials of the type  $(X - a)^m$ .

**Definition 1.3.5.** A matrix  $A = (a_{ij})$  is (upper) triangular if  $a_{ij} = 0$  for all  $i > j$ .

**Lemma 1.3.9.** Suppose that  $f$  has a minimal polynomial of the form  $(X - a)^m$ . Then there is a basis of  $V$  with respect to which the matrix of  $f$  is triangular. Further,  $m \leq \dim V$ .

*Proof.* We define a set of subspaces of  $V$  as follows.

$$V_i = \{v \in V : (f - a)^i(v) = 0\}, i = 1, 2, \dots, m.$$

Note that

$$\{0\} \subseteq V_1 \subseteq V_2 \subseteq \cdots \subseteq V_m = V.$$

Note also that if  $v \in V_i$  then  $f(v) \in V_i$  and  $(f - a)(v) \in V_{i-1}$ .

Choose a basis of  $V$  by starting with a basis  $\mathcal{B}_1$  of  $V_1$ . Extend this to a basis  $\mathcal{B}_2$  of  $V_2$ , then to a basis  $\mathcal{B}_3$  of  $V_3$  and so on until we have a basis  $\mathcal{B}_m$  of  $V_m = V$ .

Now we look at the matrix for  $f$  using the basis  $\mathcal{B}_m$ . If  $v \in \mathcal{B}_i$  but  $v \notin \mathcal{B}_{i-1}$ , then

$$f(v) = av + \sum_j^n a_j v_j \quad \text{for some } a_j \in F \quad \text{and } v_j \in \mathcal{B}_{i-1}$$

Thus the matrix  $A$  of  $f$  with respect to this basis is upper triangular with  $a$ 's on the main diagonal. Note that  $A - aI_n$  is upper triangular with zeroes on the diagonal. We leave as Exercise 5 the fact that  $(A - aI_n)^n = 0$ . Then,  $(f - a)^n$  is zero and so  $(X - a)^m$  divides  $(X - a)^n$  by Lemma 1.3.4. That is,  $m \leq n$ .  $\square$

**Theorem 1.3.10 (Triangular Form Theorem).** *Let  $V$  be a finite dimensional vector space over an algebraically closed field  $F$  and let  $f$  be a linear transformation on  $V$ . Then there is a basis of  $V$  so that the matrix of  $f$  with respect to this basis is triangular.*

*Proof.* This just combines Theorem 1.3.8 with Lemma 1.3.9.  $\square$

**Definition 1.3.6.** *Let  $f$  be a linear transformation on a finite dimensional vector space  $V$ . The characteristic polynomial  $c(x)$  of  $f$  is the polynomial  $\det(xI_V - f)$ , i.e.  $\det(xI - A)$  where  $A$  is any matrix representing  $f$ .*

Note that if two matrices  $A, B$  represent the same linear transformation then  $A, B$  will be similar, i.e.  $B = PAP^{-1}$  with  $P$  invertible. Then  $(xI - B) = P(xI - A)P^{-1}$ , so  $\det(xI - B) = \det(xI - A)$ . Hence the resulting characteristic polynomial will not depend on the choice of matrix representing the linear transformation. Note also that this characteristic polynomial is *monic* and that  $\det(xI - A) = \pm \det(A - xI)$ .

**Theorem 1.3.11 (Cayley-Hamilton Theorem).** *Let  $f$  be a linear transformation on a finite dimensional vector space. Then  $f$  satisfies its characteristic polynomial. That is, if  $c(x)$  is the characteristic polynomial then  $c(f)$  is the zero linear transformation.*

*Proof.* Choose a basis for  $V$  as described in Theorem 1.3.8. Thus the matrix  $A$  of  $f$  with respect to this basis will have the form

$$A = \begin{bmatrix} A_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A_k \end{bmatrix}$$

where each  $A_i$  is a triangular matrix with  $a_i$  (say) on the diagonal. Suppose that  $A_i$  is  $n_i \times n_i$ . Then an easy calculation tells us that the characteristic polynomial of  $A$  is

$$c(X) = \prod_{i=1}^k (X - a_i)^{n_i}.$$

But the minimal polynomial of  $f$  is

$$m(X) = \prod_{i=1}^k (X - a_i)^{m_i}$$

and, by Lemma 1.3.9,  $m_i \leq n_i$ . Thus  $m(f)$  divides  $c(f)$  and so  $c(f) = 0$  since  $m(f) = 0$ .  $\square$

### Examples:

- (1) Let  $f$  denote a reflection in  $\mathbb{R}^3$ . Then the characteristic polynomial of  $f$  is  $(X - 1)^2(X + 1)$  (for example, choose a basis consisting of two vectors in the ‘mirror’ and one perpendicular to the mirror). The minimal polynomial is  $X^2 - 1$ .
- (2) Let  $f$  be a linear transformation with matrix

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

Then the characteristic polynomial is  $(X - 2)^2(X - 1)$  whereas the minimal polynomial is  $(X - 2)(X - 1)$ .

Applications of the Cayley-Hamilton theorem include calculation of inverses and matrix powers  $A^k$  (see exercises 3 and 4).

**Remark:** We have been assuming in this subsection that the field  $F$  is algebraically closed and we have used this for our proof of the Cayley-Hamilton Theorem. But the Cayley-Hamilton Theorem is true over any field. Think of a matrix  $A$ , rather than a linear transformation, with coefficients over an

arbitrary field  $K$ . Then its characteristic polynomial has coefficients in  $K$ . There will be an algebraically closed field  $F$  containing  $K$  and we can regard  $A$  as a matrix over  $F$ . The Cayley-Hamilton theorem over  $F$  will then apply to show that  $A$  satisfies its characteristic polynomial.

### 1.3.4 Jordan normal form

Triangular form is a practical way to represent a linear transformation but is not the best possible way. For this we need the *Jordan normal form* (JNF) — also known as the *Jordan canonical form* (JCF). We shall describe this but we shall not attempt to prove the result or to show how to calculate Jordan normal form in general.

**Definition 1.3.7.** *The  $n \times n$  Jordan block matrix  $J(a, n)$  (for  $a \in F$ ) is the  $n \times n$  matrix*

$$A = \begin{bmatrix} a & 1 & 0 & \dots & 0 \\ 0 & a & 1 & \dots & 0 \\ \vdots & \dots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a & 1 \\ 0 & \dots & 0 & 0 & a \end{bmatrix}.$$

A straightforward calculation shows that the characteristic polynomial and the minimal polynomial of  $J(a, n)$  are both  $(X - a)^n$ .

**Examples:**

The matrices

$$J(2, 3) = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix} \quad J(0, 4) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad J(4, 1) = [4]$$

are all Jordan blocks.

Recall that we are assuming in this section that the field of scalars  $F$  is **algebraically closed** (e.g.  $F = \mathbb{C}$ .)

**Theorem 1.3.12 (Jordan Normal Form).** *Let  $f$  be a linear transformation on a finite dimensional vector space  $V$ . Then there is a basis of  $V$  with respect to which the matrix of  $f$  takes the form*

$$\begin{bmatrix} A_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & A_k \end{bmatrix}$$

where each  $A_i$  is a Jordan block matrix. This expression for the matrix of  $f$  is unique up to re-ordering the diagonal blocks  $A_1, \dots, A_k$ .

Equivalently, every matrix is similar to a matrix in Jordan normal form, and this Jordan form is unique up to re-ordering the Jordan blocks.

**Question:** How can we determine the Jordan normal form?

There is a fairly easy observation which gives us information about the JNF just by knowing the characteristic and minimal polynomials.

**Lemma 1.3.13.** *Suppose that  $a$  is an eigenvalue of  $f$  and let  $(X - a)^m$  be the highest power of  $X - a$  dividing the minimal polynomial and let  $(X - a)^n$  be the highest power of  $X - a$  dividing the characteristic polynomial. Then  $m$  is the size of the largest Jordan block  $J(a, l)$  that occurs in the JNF of  $f$  and  $n$  is the sum of the sizes of the Jordan blocks  $J(a, l)$  that occur in the JNF of  $f$ .*

*Proof.* The claim about the characteristic polynomial is clear. A little computation is needed to check the claim about the minimal polynomial.  $\square$

**Theorem 1.3.14.** *A linear transformation  $f$  of a finite dimensional vector space  $V$  can be represented by a diagonal matrix (with respect to a suitable basis) if and only if the minimal polynomial of  $f$  has no repeated roots. In particular, if  $\dim V = n$  and  $f$  has  $n$  distinct eigenvalues then  $f$  can be represented by a diagonal matrix.*

*Proof.* The first thing to appreciate is that the uniqueness of the JNF guarantees that if  $f$  can be represented by a diagonal matrix then that matrix is the JNF of  $f$ . A JNF is diagonal if and only if each Jordan block which occurs has size 1. By the previous lemma, this happens if and only if each eigenvalue occurs as a root of the minimal polynomial with multiplicity 1.  $\square$

**Examples:**

(1) The matrices

$$\begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix} \quad \begin{bmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

are all in JNF.

(2) The matrix

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

has characteristic polynomial  $(X-1)(X-2)^2(X-3)$ . Thus its minimal polynomial is either  $(X-1)(X-2)^2(X-3)$  or  $(X-1)(X-2)(X-3)$ . Direct computation shows that  $(A-I)(A-2I)(A-3I) = 0$ . Hence its minimal polynomial is  $(X-1)(X-2)(X-3)$  and so its JNF is

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

For matrices of small size, we can gain much information from the minimal and characteristic polynomials. For example, for a  $2 \times 2$  matrix the characteristic polynomial is  $(X-a)(X-b)$  for some  $a, b \in F$ . The minimal polynomial then divides  $(X-a)(X-b)$  and has the same roots. If  $a \neq b$  then the minimal polynomial is also  $(X-a)(X-b)$  and the JNF is  $\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$ . If  $a = b$  then the minimal polynomial is either  $(X-a)^2$  in which case the JNF is  $\begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}$  or  $(X-a)$  in which case the JNF is  $\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ .

We can do a similar analysis for  $3 \times 3$  matrices. In the following,  $a, b, c$  are assumed to be different elements of  $F$  and each row represents a different type of possibility for the characteristic polynomial, the minimal polynomial and the JNF.

Characteristic	Minimal	JNF
$(X - a)(X - b)(X - c)$	$(X - a)(X - b)(X - c)$	$\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$
$(X - a)^2(X - b)$	$(X - a)^2(X - b)$	$\begin{bmatrix} a & 1 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}$
$(X - a)^2(X - b)$	$(X - a)(X - b)$	$\begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & b \end{bmatrix}$
$(X - a)^3$	$(X - a)^3$	$\begin{bmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}$
$(X - a)^3$	$(X - a)^2$	$\begin{bmatrix} a & 1 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix}$
$(X - a)^3$	$(X - a)$	$\begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix}$

Knowing the *dimensions of eigenspaces* of a matrix  $A$  also helps in determining the Jordan normal form. It's easy to check that:

(1) For each Jordan block in the JNF there is exactly *one* linearly independent eigenvector.

(2) the dimension of the eigenspace for an eigenvalue  $\lambda$  is the number of Jordan blocks with  $\lambda$  on the diagonal.

**Example:** Find the Jordan normal form for

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

The characteristic polynomial is  $c(x) = (x - 1)^3$  so there is only one eigenvalue,  $\lambda = 1$ . Using row reduction, we find the corresponding eigenspace  $\text{Nullspace}(A - I)$  has dimension 2. Thus the Jordan normal form  $J$  has 2 blocks, hence

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

**1.3.5 Exercises**

- (1) Find the minimal polynomials of the matrices:

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

- (2) Show that the matrices

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

have the same minimal polynomial. Do they have the same characteristic polynomial?

- (3) Show that the matrix

$$A = \begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix}$$

has minimal polynomial  $X^2 - 2X - 8$ . Use this to determine the inverse of  $A$ .

- (4) Show that a linear transformation  $f$  is invertible if and only if its minimal polynomial has non-zero constant term. Assuming  $f$  is invertible, how can the inverse be calculated if the minimal polynomial is known?
- (5) Suppose that  $A$  is an  $n \times n$  upper triangular matrix with zeroes on the diagonal. Prove that  $A^n = 0$ .
- (6) Let  $f$  be a linear transformation on a vector space  $V$  with minimal polynomial  $X^2 - 1$  and suppose that  $2 \neq 0$  in the field of scalars. (Thus, for example,  $\mathbb{Z}_2$  is not allowed as the field of scalars.) Show directly that the subspaces  $\{v \in V : f(v) = v\}$  and  $\{v \in V : f(v) = -v\}$  are complementary subspaces of  $V$ . Find a diagonal matrix representing  $f$ .
- (7) Prove Lemma 1.3.3.
- (8) Prove Lemma 1.3.6.

- (9) Show that the linear transformation  $\mathcal{P}_n(\mathbb{R}) \rightarrow \mathcal{P}_n(\mathbb{R})$  given by differentiation cannot be represented by a diagonal matrix.
- (10) If  $f$  is a linear transformation on a finite dimensional vector space  $V$  satisfying  $f^2 = f$ , explain how to find a diagonal matrix representing  $f$ .
- (11) Suppose that linear transformations  $f$  and  $g$  on a vector space  $V$  commute; that is, that  $fg = gf$ . Show that an eigenspace of  $f$  will be  $g$ -invariant. If the field  $F$  of scalars is algebraically closed, deduce that  $f$  and  $g$  have a common eigenvector.
- (12) Find the Jordan normal form of the following matrices:

$$\begin{bmatrix} -1 & 1 \\ -1 & -3 \end{bmatrix}, \begin{bmatrix} -1 & 3 & 0 \\ 0 & 2 & 0 \\ 2 & 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix}.$$

- (13) For each of the following pairs of minimal and characteristic polynomials, find all possibilities for the Jordan normal form:

Minimal polynomial	Characteristic polynomial
$X^2(X+1)^2$	$X^2(X+1)^4$
$(X-3)^2$	$(X-3)^5$
$X^3$	$X^7$
$(X-1)^2(X+1)^2$	$(X-1)^4(X+1)^4$

- (14) Which of the following pairs of matrices (over  $\mathbb{C}$ ) are similar?

- (a)  $\begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix};$
- (b)  $\begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 5 \\ 0 & -1 \end{bmatrix};$
- (c)  $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$

- (15) Given a  $4 \times 4$  matrix  $A$  over  $\mathbb{C}$  and given the minimal and characteristic polynomials of  $A$ , describe the possibilities for the JNF of  $A$ . (There will be one case where there are two possibilities.)

- (16) Show that any JNF matrix  $J$  is a sum  $J = D + N$  where  $D$  is diagonal and  $N$  is nilpotent; that is  $N^k = 0$  for some  $k$ . Deduce that any linear transformation  $f$  of a finite dimensional complex vector space can be written in the form  $f = d + n$  where  $d$  is diagonalisable and  $n$  is nilpotent.
- (17) In the language of the previous question, show that  $JN = NJ$  and  $JD = DJ$ . Deduce that  $fd = df$  and  $fn = nf$ .
- (18) (Harder) Show that the Jordan normal form of a complex matrix  $A$  is completely determined by the dimensions of the nullspaces of  $(A - \lambda I)^i$ ,  $i = 1, 2, 3, \dots$  for all the eigenvalues  $\lambda$  of  $A$ .

## 1.4 Inner product spaces

Inner products are generalizations of the *dot product* in  $\mathbb{R}^3$  or  $\mathbb{R}^n$ . They give a way to introduce *geometry* in a vector space.

You have already seen some of the properties of inner products on real vector spaces. Inner products on complex vector spaces have many of the same properties but there are sufficient differences that we need to cover the topic from scratch.

In this chapter, the **field  $F$  will be either the real numbers or the complex numbers**. Each has an absolute value function  $F \rightarrow \mathbb{R}_+$  and each has a ‘conjugate function’  $F \rightarrow F$ . For the complex numbers, this is the usual complex conjugation and for the real numbers it is simply the identity function.

### 1.4.1 Complex inner products

**Definition 1.4.1.** An inner product in a vector space  $V$  over  $F$  is a function  $V \times V \rightarrow F$  satisfying the following (we write the value of the function as  $(v, w)$  where  $v, w \in V$  and  $(v, w) \in F$ ):

- (1)  $(v, w) = \overline{(w, v)}$  for all  $v, w \in V$ ;
- (2)  $(au + bv, w) = a(u, w) + b(v, w)$  for all  $u, v, w \in V$  and  $a, b \in F$ ;
- (3)  $(v, v) \geq 0$  for all  $v \in V$  and  $(v, v) = 0$  if and only if  $v = 0$ .

(Here and later, when we say  $(v, v) \geq 0$  we will mean ‘ $(v, v)$  is real and non-negative’.)

A real inner product space is often called a Euclidean space and a complex inner product space is often called a unitary space.

**Note:** (1) Taking  $v = w$  in condition (1) gives  $(v, v) = \overline{(v, v)}$ ; hence  $(v, v)$  is always *real*.

(2) Conditions (1) and (2) imply that  $(w, au + bv) = \bar{a}(w, u) + \bar{b}(w, v)$ .

**Examples:**

(1) Set  $V = \mathbb{R}^n$  and define

$$((a_1, a_2, \dots, a_n), (b_1, b_2, \dots, b_n)) = a_1 b_1 + a_2 b_2 + \dots + a_n b_n.$$

This gives an inner product, the standard ‘dot product’.

(2) Set  $V = \mathbb{C}^n$  and define

$$((a_1, a_2, \dots, a_n), (b_1, b_2, \dots, b_n)) = a_1 \bar{b}_1 + a_2 \bar{b}_2 + \dots + a_n \bar{b}_n.$$

This again gives an inner product, the ‘complex dot product’.

(3) Let  $V$  be any  $n$ -dimensional real space and let  $\{v_1, \dots, v_n\}$  be a basis of  $V$ . Define an inner product on  $V$  by

$$(a_1 v_1 + \dots + a_n v_n, b_1 v_1 + \dots + b_n v_n) = a_1 b_1 + \dots + a_n b_n.$$

This gives an inner product; it is not hard to see that Example 1 is a special case of this.

(4) Let  $V$  be any  $n$ -dimensional complex space and let  $\{v_1, \dots, v_n\}$  be a basis of  $V$ . Define an inner product on  $V$  by

$$(a_1 v_1 + \dots + a_n v_n, b_1 v_1 + \dots + b_n v_n) = a_1 \bar{b}_1 + a_2 \bar{b}_2 + \dots + a_n \bar{b}_n.$$

This gives an inner product; again, Example 2 is a special case of this.

(5) Let  $V = M_{n \times n}(F)$ . Define an inner product by

$$(A, B) = \text{trace}(A\bar{B}^T)$$

where  $\text{trace}(C)$ , for a square matrix  $C$ , is the sum of the diagonal entries.

(6) Recall that  $\mathcal{P}(F)$  denotes the space of all polynomials with coefficients from  $F = \mathbb{R}$  or  $\mathbb{C}$ . We can define an inner product on  $\mathcal{P}(F)$  by

$$(p(x), q(x)) = \int_0^1 p(x) \overline{q(x)} dx.$$

- (7) Let  $V = C([a, b], F)$  be the vector space of all continuous functions  $f : [a, b] \rightarrow F$  where  $[a, b]$  is the closed interval  $\{t : a \leq t \leq b\}$ . Then we can define an inner product on  $V$  by

$$(f, g) = \int_a^b f(t)\overline{g(t)} dt.$$

**Definition 1.4.2.** *If  $V$  is an inner product space:*

- (1) *The length of a vector  $v \in V$  is  $\|v\| = \sqrt{(v, v)}$ .*
- (2) *Two elements  $v, w \in V$  are orthogonal if  $(v, w) = 0$ .*
- (3) *A subset  $S$  of  $V$  is said to be orthonormal if  $v, w \in S$  implies that*

$$(v, w) = \begin{cases} 0 & \text{if } v \neq w \\ 1 & \text{if } v = w \end{cases}$$

**Theorem 1.4.1.** (1) *Any orthonormal set is linearly independent.*

- (2) *Any orthonormal set can be extended to an orthonormal basis.*

*Proof.* We leave the proof of the first part as an exercise. The second part is the Gram-Schmidt orthogonalisation process which most of you will have seen for the space  $\mathbb{R}^n$ .

We sketch the argument. Let  $\mathcal{O} = \{v_1, \dots, v_m\}$  be an orthonormal set. By the first part it is linearly independent and so can be extended to a basis  $\mathcal{O} \cup \mathcal{B}$ . Let  $w \in \mathcal{B}$  and set  $w' = w - \sum_{i=1}^m (w, v_i)v_i$ . It is easily checked that  $(w', v_i) = 0$  for  $i = 1, \dots, m$ . Thus  $\mathcal{O} \cup \{w'/\|w'\|\}$  will be an orthonormal set strictly containing  $\mathcal{O}$ . Now repeat the process until you have a basis.  $\square$

**Examples:**

- (1) We can extend the orthonormal set  $\{(1/\sqrt{2})(1, 1, 0), (1/\sqrt{3})(1, -1, 1)\}$  of  $\mathbb{R}^3$  by adding the vector  $w = (0, 0, 1)$  to form a basis and then forming

$$\begin{aligned} w' &= (0, 0, 1) - 0 \cdot (1/\sqrt{2})(1, 1, 0) - 1 \cdot (1/\sqrt{3})(1, -1, 1) \\ &= (1/\sqrt{3})(-1, +1, \sqrt{3} - 1). \end{aligned}$$

So  $\{(1/\sqrt{2})(1, 1, 0), (1/\sqrt{3})(1, -1, 1), w'/\|w'\|\}$  will be an orthonormal basis of  $\mathbb{R}^3$ .

(2) The set  $\{1\}$  is an orthonormal set in  $\mathcal{P}_1(\mathbb{R})$ . If we extend it to a basis and apply the Gram-Schmidt orthogonalisation process, we obtain an orthonormal basis  $\{1, \sqrt{3}(2x - 1)\}$ .

(3) The set

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$$

is an orthonormal basis of  $M_{2 \times 2}(\mathbb{R})$ .

**Theorem 1.4.2 (Bessel's inequality).** *Let  $S = \{v_1, \dots, v_n\}$  be an orthonormal subset of an inner product space  $V$ . Let  $v \in V$  and set  $a_i = (v, v_i)$  for  $i = 1, \dots, n$ . Then*

$$\sum_{i=1}^n |a_i|^2 \leq \|v\|^2.$$

*The vector  $v - \sum_{i=1}^n (v, v_i)v_i$  is orthogonal to each  $v_j$ . In particular, if  $S$  is a basis of  $V$ , then*

$$v = \sum_{i=1}^n a_i v_i$$

and

$$\sum_{i=1}^n |a_i|^2 = \|v\|^2.$$

*Proof.* We have

$$\begin{aligned} 0 &\leq \left\| v - \sum_{i=1}^n a_i v_i \right\|^2 = \left( v - \sum_{i=1}^n a_i v_i, v - \sum_{i=1}^n a_i v_i \right) \\ &= (v, v) - \sum_{i=1}^n a_i (v_i, v) - \sum_{i=1}^n \bar{a}_i (v, v_i) + \sum_{i,j=1}^n a_i \bar{a}_j (v_i, v_j) \\ &= (v, v) - \sum_{i=1}^n |a_i|^2 - \sum_{i=1}^n |a_i|^2 + \sum_{i=1}^n |a_i|^2 \\ &= (v, v) - \sum_{i=1}^n |a_i|^2. \end{aligned}$$

which completes the proof of the first part.

For the second part, note that

$$\left( v - \sum_{i=1}^n a_i v_i, v_j \right) = (v, v_j) - \sum_{i=1}^n a_i (v_i, v_j) = (v, v_j) - a_j = 0.$$

Finally, if  $S$  is a basis of  $V$  then  $v - \sum_{i=1}^n a_i v_i$  will be orthogonal to every element of a basis and so must be zero. Further, equality holds in the first inequality.  $\square$

**Example:** Let  $\mathcal{F}$  be the subspace of  $\mathbb{R}^{\mathbb{R}}$  consisting of functions which are periodic of period  $2\pi$  and which are ‘sufficiently good to have a convergent Fourier series’. We could require the functions to be twice differentiable, for example, but weaker conditions will also do. Define an inner product on  $\mathcal{F}$  by

$$(f, g) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t)g(t)dt.$$

The functions  $\cos nt$  and  $\sin nt$  will certainly lie in  $\mathcal{F}$  and

$$\left\{ \frac{1}{\sqrt{2}}, \cos t, \sin t, \cos 2t, \sin 2t, \dots \right\}$$

is an infinite orthonormal set in  $\mathcal{F}$ .

It is **not** true that this orthonormal set is a basis of  $\mathcal{F}$  but the theory of Fourier series tells us that every function  $f$  in  $\mathcal{F}$  can be expressed in the form

$$\begin{aligned} f = (f, \frac{1}{\sqrt{2}}) \frac{1}{\sqrt{2}} + (f, \sin t) \sin t + (f, \cos t) \cos t + \dots + \\ + (f, \sin nt) \sin nt + (f, \cos nt) \cos nt + \dots \end{aligned} \quad (1.2)$$

This expression is an infinite series and so does not fit into our theory of vector spaces. But the expression here is very similar to that in the last sentence of the previous theorem and suggests that there is some generalisation of the current theory into which we could fit the theory of Fourier series.

**Lemma 1.4.3 (Schwarz’s inequality).** *If  $v, w$  are elements of an inner product space  $V$  then*

$$|(v, w)| \leq \|v\| \cdot \|w\|.$$

*Proof.* If  $w = 0$  then both sides are zero. If not, then  $\{w/\|w\|\}$  is an orthonormal set and the result follows directly from Bessel’s inequality.  $\square$

If  $V$  is a real inner product space, this allows us to define the *angle*  $\theta$  between two non-zero vectors  $v, w$  in  $V$  by

$$\cos \theta = \frac{(v, w)}{\|v\| \cdot \|w\|}, \text{ and } 0 \leq \theta \leq \pi.$$

**Examples:**

(1) If we take  $V = \mathbb{R}^n$  and the dot product, this becomes

$$\left| \sum_{i=1}^n a_i b_i \right| \leq \left( \sum_{i=1}^n a_i^2 \right)^{\frac{1}{2}} \left( \sum_{i=1}^n b_i^2 \right)^{\frac{1}{2}}$$

for any real numbers  $a_i, b_i$ .

(2) If we take  $V = \mathbb{C}^n$  and the complex dot product, we have

$$\left| \sum_{i=1}^n a_i \bar{b}_i \right| \leq \left( \sum_{i=1}^n |a_i|^2 \right)^{\frac{1}{2}} \left( \sum_{i=1}^n |b_i|^2 \right)^{\frac{1}{2}}$$

for any complex numbers  $a_i, b_i$ . Note that, by replacing  $b_i$  by  $\bar{b}_i$ , we can obtain a complex inequality identical to the previous example.

(3) If we take the inner product space of Example 7 above, then we have

$$\left| \int_a^b f(t) \overline{g(t)} dt \right| \leq \left( \int_a^b f(t)^2 dt \right)^{\frac{1}{2}} \left( \int_a^b g(t)^2 dt \right)^{\frac{1}{2}}.$$

Schwarz's inequality also allows us to define *distance* on an inner product space. If  $u, v \in V$  then we can define the distance between  $u$  and  $v$  to be  $\delta(u, v) = \|u - v\|$ . Clearly,  $\delta(u, u) \geq 0$  and  $\delta(u, v) = \delta(v, u)$ . The other major requirement of a 'distance function' is that  $\delta(u, v) \leq \delta(u, w) + \delta(w, v)$ , the 'triangle inequality'.

To see the latter, we will show that  $\|x + y\|^2 \leq (\|x\| + \|y\|)^2$  for any  $x, y \in V$ . Then if we take square roots and replace  $x$  by  $u - w$  and  $y$  by  $w - v$  we will have the triangle inequality.

So

$$\begin{aligned} \|x + y\|^2 &= (x + y, x + y) = (x, x) + (x, y) + (y, x) + (y, y) \\ &= \|x\|^2 + (x, y) + \overline{(x, y)} + \|y\|^2 \\ &= \|x\|^2 + 2\operatorname{Re}((x, y)) + \|y\|^2 \\ &\leq \|x\|^2 + 2|(x, y)| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2. \end{aligned}$$

### 1.4.2 Orthogonal complements

**Definition 1.4.3.** Let  $V$  be an inner product space and let  $W$  be a subspace of  $V$ . The orthogonal complement  $W^\perp$  of  $W$  in  $V$  is

$$W^\perp = \{v \in V : (v, w) = 0 \text{ for all } w \in W\}.$$

**Lemma 1.4.4.** Let  $V$  be an inner product space and let  $W$  be a subspace of  $V$ . Then  $W^\perp$  is a subspace of  $V$ ,  $W^\perp \cap W = \{0\}$  and  $V = W + W^\perp$ . Thus,  $V$  is a direct sum  $V = W \oplus W^\perp$  and  $\dim W^\perp = \dim V - \dim W$ .

*Proof.* It is an easy check that  $W^\perp$  is a subspace of  $V$ . If  $w \in W^\perp$  then  $(w, W) = 0$ . If also,  $w \in W$ , then  $(w, w) = 0$  and so  $w = 0$ ; that is,  $W^\perp \cap W = \{0\}$ .

Choose an orthonormal basis  $\mathcal{B}_1$  for  $W$  and extend this to an orthonormal basis  $\mathcal{B}_1 \cup \mathcal{B}_2$  for  $V$ . Because  $\mathcal{B}_1 \cup \mathcal{B}_2$  is orthonormal, each vector in  $\mathcal{B}_2$  is orthogonal to each vector in  $\mathcal{B}_1$  and so  $\mathcal{B}_2 \subseteq W^\perp$ . Every element  $v$  of  $V$  can be written as a sum  $v = v_1 + v_2$  with  $v_i \in \langle \mathcal{B}_i \rangle$ . But then  $v_1 \in W$  and  $v_2 \in W^\perp$ . Thus  $V = W + W^\perp$ .

The final sentence follows from Lemma 1.3.2. □

#### Examples:

- (1) The orthogonal complement to a plane through the origin in  $\mathbb{R}^3$  is the normal through the origin.
- (2) The orthogonal complement to a line through the origin in  $\mathbb{R}^3$  is the plane through the origin to which it is normal.
- (3) The orthogonal complement to the set of diagonal matrices in  $M_{n \times n}(\mathbb{R})$  is the set of matrices with zero entries on the diagonal.
- (4) If  $A$  is an  $m \times n$  matrix with real coefficients then the *row space* of  $A$  is the orthogonal complement of the *nullspace* of  $A$ . (See exercise 6.)

### 1.4.3 Adjoints, self-adjoint, Hermitian, normal

**Definition 1.4.4.** Let  $f : V \rightarrow V$  be a linear transformation on an inner product space  $V$ . The adjoint  $f^*$  of  $f$  is a linear transformation  $f^* : V \rightarrow V$  satisfying

$$(f(v), w) = (v, f^*(w)) \quad \text{for all } v, w \in V. \quad (*)$$

**Lemma 1.4.5.** If  $V$  is finite dimensional then the adjoint  $f^*$  exists and is unique.

*Idea of proof.* To see that  $(*)$  really does define a function  $f^*$  we need to show that, given  $w$  there is a unique  $w_1 \in V$  such that  $(f(v), w) = (v, w_1)$  for all  $v \in V$ ; we can then set  $f^*(w) = w_1$ . If  $V$  is finite-dimensional, this is always possible; the proof is left as Exercise 9. Once this is done, we know that  $f^*$  is a well-defined function. It is then not hard to see that  $f^*$  is a linear transformation; the proof is left as Exercise 10.  $\square$

**Note:** An adjoint  $f^*$  does not always exist if  $V$  is infinite dimensional (see Exercise 7).

The next lemma is often useful for working with adjoints.

**Lemma 1.4.6.** *If  $f, g : V \rightarrow V$  are linear transformations on an inner product space  $V$  satisfying*

$$(f(v), w) = (g(v), w) \text{ for all } v, w \in V$$

*then  $f = g$ .*

*Proof.* We have

$$(f(v) - g(v), w) = 0 \text{ for all } v, w \in V.$$

Taking  $w = f(v) - g(v)$  gives

$$(f(v) - g(v), f(v) - g(v)) = 0 \text{ for all } v \in V.$$

Hence  $f(v) - g(v) = 0$  for all  $v \in V$  by the positivity of inner products. So  $f(v) = g(v)$  for all  $v \in V$  and  $f = g$ .  $\square$

**Some properties of adjoints:** If  $f, g : V \rightarrow V$  are linear transformations on a finite dimensional inner product space  $V$  and  $\alpha \in \mathbb{C}$  then

- (1)  $(f + g)^* = f^* + g^*$ ,
- (2)  $(\alpha f)^* = \bar{\alpha} f^*$ ,
- (3)  $(fg)^* = g^* f^*$  (where  $fg$  denotes the composition of  $f \circ g$ ),
- (4)  $(f^*)^* = f$

These follow easily from the definition of adjoint and the previous lemma (see exercises 3,11).

If we have a matrix for  $f$ , then what is the matrix of  $f^*$ ? To get a nice answer we need to choose the matrix with respect to an orthonormal basis.

**Lemma 1.4.7.** *Let  $V$  be an inner product space with an orthonormal basis  $\mathcal{B} = \{v_1, \dots, v_n\}$ . Suppose that a linear transformation  $f$  has a matrix  $A$  with respect to  $\mathcal{B}$ . Then the matrix  $A^*$  of  $f^*$  with respect to  $\mathcal{B}$  is given by*

$$(A^*)_{ij} = \overline{A_{ji}};$$

that is,  $A^*$  is the ‘complex conjugate transpose’ of  $A$ .

*Proof.* Suppose that  $A = (a_{ij})$  and  $A^* = (b_{ij})$ . Then for any  $i, j$ ,

$$(f(v_i), v_j) = \left( \sum_k a_{ki} v_k, v_j \right) = \sum_k a_{ki} (v_k, v_j) = a_{ji}$$

and

$$(v_i, f^*(v_j)) = \left( v_i, \sum_k b_{kj} v_k \right) = \sum_k \overline{b_{kj}} (v_i, v_k) = \overline{b_{ji}}$$

and so  $b_{ij} = \overline{a_{ji}}$  as required.  $\square$

**Definition 1.4.5.** *If  $A$  is an  $n \times n$  matrix over  $F$  then  $A^*$  is the matrix defined by  $(A^*)_{ij} = \overline{A_{ji}}$ .*

We also call  $A^*$  the adjoint of  $A$ .

Next we look at some important kinds of linear transformations

**Definition 1.4.6.** *Let  $f$  be a linear transformation on an inner product space  $V$ .*

- (1) *We say that  $f$  is self-adjoint if  $f = f^*$ . If  $V$  is real, this is often called symmetric. If  $V$  is complex, it is called Hermitian.*
- (2) *We say that  $f$  is an isometry if  $f^*f = 1_V$  where  $1_V$  is the identity transformation on  $V$ . If  $V$  is real, this is often called orthogonal. If  $V$  is complex, it is called unitary.*
- (3) *We say that  $f$  is normal if  $ff^* = f^*f$ .*

Similar terminology applies to matrices.

**Examples:**

- (1) If a linear transformation is represented by a symmetric matrix with respect to an orthonormal basis, then it is self-adjoint. For example  $\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$  is self-adjoint.
- (2) A rotation of  $\mathbb{R}^n$  is orthogonal.
- (3) The matrix  $\begin{bmatrix} 1 & 2-i \\ 2+i & 3 \end{bmatrix}$  is self-adjoint (Hermitian).
- (4) Skew-symmetric real matrices (that is, matrices  $A$  that satisfy  $A = -A^*$ ) are normal but not self-adjoint.
- (5) The matrix  $\begin{bmatrix} 1 & 1 \\ i & 3+2i \end{bmatrix}$  is normal but not self-adjoint or skew-symmetric or orthogonal.

Isometries are so called because they are *distance preserving* transformations.

**Lemma 1.4.8.** *Let  $f$  be a linear transformation on an inner product space  $V$ . The following are equivalent.*

- (1)  $f^*f = 1_V$ ;
- (2)  $(f(u), f(v)) = (u, v)$  for all  $u, v \in V$ ;
- (3)  $\|f(v)\| = \|v\|$  for all  $v \in V$ .

**Note:** If  $V$  is finite dimensional, (1) means that  $f^* = f^{-1}$ .

*Proof.* If (1) is true and if  $u, v \in V$  then

$$(f(u), f(v)) = (u, f^*(f(v))) = (u, 1_V(v)) = (u, v)$$

and so (2) is true. Set  $u = v$  and take square roots to obtain (3) from (2). Assume (3) and consider the linear transformation  $g = f^*f - 1_V$ . Then

$$g^* = (f^*f - 1_V)^* = (f^*f)^* - 1_V = f^*f^{**} - 1_V = f^*f - 1_V = g$$

and so  $g$  is self-adjoint. For all  $v \in V$ ,

$$(g(v), v) = (f^*(f(v)) - v, v) = (f(v), f(v)) - (v, v) = \|f(v)\|^2 - \|v\|^2 = 0.$$

We leave it as Exercise 12 to show that this means that  $g$  is the zero linear transformation and so  $f^*f = 1_V$ .  $\square$

**Lemma 1.4.9.** *Let  $W$  be an  $f$ -invariant subspace of  $V$ . Then  $W^\perp$  is  $f^*$ -invariant.*

*Proof.* Suppose that  $u \in W^\perp$ . If  $w \in W$  it follows that  $f(w) \in W$  because  $W$  is  $f$ -invariant and so

$$(w, f^*(u)) = (f(w), u) = 0 \text{ as } u \in W^\perp.$$

Hence  $f^*(u) \in W^\perp$  and so  $W^\perp$  is  $f^*$ -invariant.  $\square$

**Lemma 1.4.10.** *Let  $f$  be a linear transformation over a real vector space  $V$ . Then  $V$  has an  $f$ -invariant subspace of dimension at most 2.*

*Proof.* Consider the minimal polynomial  $m(X)$  of  $f$ . Since the field of scalars is the real numbers, any irreducible factor  $p(X)$  of  $m(X)$  has degree at most 2. (If you have not seen this fact before it may require some thought; remember that the polynomial factors completely over the complex numbers and that complex roots of a real polynomial occur in complex conjugate pairs.) Set  $m(X) = p(X)q(X)$  and set  $W = \{q(f)(v) : v \in V\}$ . Then  $W$  is a non-zero subspace (otherwise we would have  $q(f) = 0$  on  $V$ ) and  $p(f)(w) = 0$  for any  $w \in W$ .

Choose any  $w \in W$  with  $w \neq 0$ . If  $p(X) = X - a$  for some  $a \in \mathbb{R}$  then  $f(w) = aw$  and so the subspace  $\langle w \rangle$  is 1-dimensional and invariant. If  $p(X) = X^2 + aX + b$  for some  $a, b \in \mathbb{R}$  then the subspace  $\langle w, f(w) \rangle$  is 2-dimensional and invariant.  $\square$

**Theorem 1.4.11.** *Let  $f$  be an orthogonal linear transformation over a real vector space  $V$ . Then there is an orthonormal basis of  $V$  of the form*

$$\{u_1, v_1, u_2, v_2, \dots, u_k, v_k, w_1, \dots, w_l\}$$

so that, for some  $\theta_1, \dots, \theta_k$ ,

$$f(u_i) = (\cos \theta_i)u_i + (\sin \theta_i)v_i \quad \text{and} \quad f(v_i) = -\sin(\theta_i)u_i + (\cos \theta_i)v_i$$

and  $f(w_i) = \pm w_i$ .

*Proof.* We shall prove this by induction on the dimension of  $V$ . By Lemma 1.4.10,  $V$  has an  $f$ -invariant subspace  $W$  which is 1 or 2 dimensional. By Lemma 1.4.9,  $W^\perp$  is  $f^*$ -invariant. Since  $f$  is orthogonal,  $f^* = f^{-1}$  and so  $W^\perp$  is  $f^{-1}$ -invariant; that is,  $f^{-1}(W^\perp) = W^\perp$  and so  $f(W^\perp) = W^\perp$ . That is,  $W^\perp$  is  $f$ -invariant.

We leave as Exercise 14 that the two linear transformations obtained by restriction,  $f_W$  and  $f_{W^\perp}$ , are still orthogonal. Since  $W$  has dimension either

1 or 2 then it has an orthonormal basis of the kind described (we leave this as Exercise 15). Since the dimension of  $W^\perp$  is less than that of  $V$ , we can apply the inductive hypothesis to deduce that  $W^\perp$  has a basis of the kind described. Combining these two bases (possibly with some re-ordering) we have a basis for  $V$  as required in the statement of the theorem.  $\square$

### Examples:

- (1) In dimension 2, the possibilities for orthogonal matrices *up to similarity* are

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

for some  $\theta$ . The first represents a reflection, the last represents a rotation through  $\theta$ .

- (2) In dimension 3, there must be a real eigenvalue because the characteristic polynomial has degree 3 and real polynomials of odd degree must have a real root. We can summarise the possibilities as follows:

three eigenvalues equal to 1: the identity;

two eigenvalues equal to 1 and one equal to -1: a reflection;

one eigenvalue equal to 1, the other two either both -1 or complex: a rotation

one eigenvalue equal to -1, the other two either both -1 or complex: an 'improper rotation', i.e. the product of a rotation and a reflection.

### 1.4.4 Exercises

- (1) Find the length of
- $(2 + i, 3 - 2i, -1)$  in the inner product space of Example 2.
  - $x^2 - 3x + 1$  in the inner product space of Example 6.
  - $\begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$  in the inner product space of Example 5.
- (2) An exercise (from an anonymous textbook) claims that, for all elements  $u, v$  of an inner product space,  $\|u + v\| + \|u - v\| = 2\|u\| + 2\|v\|$ . Prove that this is false. Can you guess what was intended?
- (3) If  $f$  and  $g$  are linear transformations on an inner product space  $V$ , show that

$$(f + g)^* = f^* + g^* \quad \text{and} \quad (fg)^* = g^* f^*.$$

- (4) If  $A$  is a transition matrix between orthonormal bases, show that  $A$  is isometric.
- (5) Suppose that  $f$  is a linear transformation on a finite dimensional inner product space  $V$ .
  - (a) If  $f$  is self-adjoint, show that the eigenvalues of  $f$  are real;
  - (b) if  $f$  is isometric, show that the eigenvalues of  $f$  have absolute value 1.
- (6) Suppose that  $f$  is a linear transformation on a finite dimensional inner product space  $V$ . Show that the range of  $f^*$  is the orthogonal complement of the nullspace of  $f$ . Deduce that the rank of  $f$  is equal to the rank of  $f^*$ . Deduce that the row-rank of a square matrix is equal to its column rank.
- (7) (Harder) Show that the function  $\delta$  of differentiation on the inner product space of Example 6 (after Definition 1.4.1) has no adjoint. (Hint: Try to find what  $\delta^*(1)$  should be.)
- (8) Show that a triangular matrix which is self-adjoint or unitary is diagonal.
- (9) Let  $V$  be a finite dimensional inner product space and  $f$  a linear transformation on  $V$ . Show that, given a vector  $w \in V$ , there exists a unique vector  $w_1 \in V$  such that  $(f(v), w) = (v, w_1)$  for all  $v \in V$ . (Hint: First show that it will be enough to consider only those  $v$  which lie in some fixed orthonormal basis of  $V$ .)
- (10) Deduce that the definition given in Definition 1.4.4 does define a linear transformation.
- (11) Let  $f$  be a linear transformation on a finite dimensional inner product space  $V$ . Show, without using matrices, that  $(f^*)^* = f$ .
- (12) Let  $g$  be a self-adjoint linear transformation on a finite dimensional inner product space  $V$ . Suppose that  $(g(v), v) = 0$  for all  $v \in V$ .
  - (a) Show that  $(g(u), w) + (g(w), u) = 0$  for all  $u, w \in V$  (replace  $v$  by  $u + w$ ).
  - (b) Deduce that  $g$  is the zero linear transformation if the space is a real space. (This is the time to use the fact that  $g$  is self-adjoint).

- (c) Assume now that the space is complex; deduce that  $(g(u), w)$  is imaginary for all  $u, w \in V$ .
- (d) Deduce that  $(g(iu), w)$  is imaginary for all  $u, w \in V$  and so that  $(g(u), w) = 0$  for all  $u, w \in V$ .
- (e) Deduce that  $g$  is zero in the complex case also.
- (13) Let  $f$  be a linear transformation on a finite dimensional inner product space  $V$ . Suppose that  $W$  is an  $f$ -invariant and  $f^*$ -invariant subspace of  $V$ . Show that  $(f_W)^* = (f^*)_W$ .
- (14) Let  $f$  be an isometry on a finite dimensional inner product space  $V$ . Suppose that  $W$  is an  $f$ -invariant subspace of  $V$ . Show that  $f_W$  is also an isometry.
- (15) Let  $V$  be a two dimensional real inner product space and let  $f$  be an isometry of  $V$ . Show that  $f$  can be represented by a matrix of the form:

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

where  $\epsilon = \pm 1$ .

## 1.5 The Spectral Theorem and applications

The main result of this section shows that for the important class of *normal* linear transformations we can find a basis made up of orthonormal eigenvectors.

### 1.5.1 The Theorem

**Theorem 1.5.1 (Spectral Theorem; first version).** *Let  $f$  be a normal linear transformation on a finite dimensional complex inner product space  $V$ . Then there is an orthonormal basis for  $V$  such that the matrix of  $f$  with respect to this basis is diagonal.*

We can also state this in terms of matrices: Let  $A$  be a normal matrix. Then there exists a unitary matrix  $U$  such that  $U^{-1}AU = U^*AU = D$  is diagonal.

Before we prove the spectral theorem we need the following preliminary result.

**Lemma 1.5.2.** *Let  $f$  be a normal linear transformation on a finite dimensional complex inner product space  $V$ . Then there is a non-zero element of  $V$  which is an eigenvector for both  $f$  and  $f^*$ . The two corresponding eigenvalues are complex conjugates.*

*Proof.* Firstly, let  $a$  be an eigenvalue of  $f$  and let  $V_a$  be the subspace of eigenvectors corresponding to  $a$ . It is easily checked that  $V_a$  is  $f$ -invariant. We claim that it is also  $f^*$ -invariant. If  $v \in V_a$  then  $f(f^*(v)) = f^*(f(v))$ , as  $f$  is normal, and so  $f(f^*(v)) = f^*(av) = af^*(v)$  as  $v \in V_a$ . Thus  $f^*(v)$  is also an  $a$ -eigenvector of  $f$  and so  $f^*(v) \in V_a$ . Hence  $V_a$  is  $f^*$ -invariant.

Thus we can consider  $f_{V_a}^*$ . This will have an eigenvector in  $V_a$ . Call this eigenvector  $w$  and suppose that  $b$  is the corresponding eigenvalue. Thus  $f^*(w) = bw$ . Also, as  $w \in V_a$ ,  $f(w) = aw$ . Note that

$$a(w, w) = (aw, w) = (f(w), w) = (w, f^*(w)) = (w, bw) = \bar{b}(w, w)$$

and so  $a = \bar{b}$ . □

More generally, let  $f, g : V \rightarrow V$  be two linear transformations on a complex vector space  $V$ . If  $f, g$  commute (i.e.  $fg = gf$ ), then they have a common eigenvector (see exercise 11, section 1.3.5).

*Proof of spectral theorem.* We shall prove the theorem by induction on  $\dim V$ . If  $\dim V = 1$ , the result is immediate. So we shall suppose that  $\dim V > 1$  and that the theorem is true for all spaces of dimension less than  $V$ .

By Lemma 1.5.2 we can choose an element of  $V$  which is a non-zero eigenvector for both  $f$  and  $f^*$ . Let  $W = \langle v \rangle^\perp$ . Since  $\langle v \rangle$  is both  $f$  and  $f^*$  invariant,  $W$  will be both  $f^*$  and  $f^{**} = f$  invariant, by Lemma 1.4.9.

Thus we can apply the inductive hypothesis to  $W$  and  $f_W$ . (Exercise: check that  $f_W$  is a normal linear transformation on  $W$ , using exercise 2 below). We obtain an orthonormal basis of  $W$  with respect to which the matrix of  $f_W$  is diagonal. Adding the vector  $v/\|v\|$  gives an orthonormal basis of  $V$  consisting of eigenvectors for  $f$ . □

The first version of the spectral theorem is a straightforward statement which is, hopefully, easy to understand. We are now going to give another version which is less easy to understand. But there is a reason for this obfuscation. The second version is one which generalises much more easily to infinite dimensional spaces of the right kind (Hilbert spaces) where matrices are not available (or, at least, much less useful). In Hilbert space, the statement will be similar but the summations will be replaced by integrals.

**Theorem 1.5.3 (Spectral Theorem; second version).** *Let  $f$  be a normal linear transformation on a finite dimensional complex inner product space  $V$ . There exist self-adjoint (Hermitian) linear transformations  $e_1, \dots, e_k$  and scalars  $a_1, \dots, a_k$  such that:*

- (1)  $a_i \neq a_j$  if  $i \neq j$ ;
- (2)  $e_i^2 = e_i$  and no  $e_i$  is zero;
- (3)  $\sum_{i=1}^k e_i = 1_V$ ;
- (4)  $\sum_{i=1}^k a_i e_i = f$ .

(In fact,  $e_i$  is the orthogonal projection onto the  $a_i$ -eigenspace.)

*Proof.* Find the orthonormal basis given by the first version of the theorem and let  $A$  be the corresponding diagonal matrix. Let  $a_1, \dots, a_k$  be the distinct eigenvalues of  $f$  and so the distinct diagonal entries of  $A$ . Let  $E_i$  be the diagonal matrix with an entry of 1 wherever  $a_i$  occurs in  $A$  and an entry of 0 elsewhere. Then  $\sum_{i=1}^k E_i$  is the identity matrix and  $\sum_{i=1}^k a_i E_i = A$ . Also  $E_i^2 = E_i$ . Let  $e_i$  be the linear transformation which has the matrix  $E_i$  with respect to the chosen orthonormal basis. The required properties of  $e_i$  are now easy to check.  $\square$

## 1.5.2 Polar form

There are strong similarities between complex numbers and complex matrices (or linear transformations over a complex vector space). Just as we have absolute values, real and imaginary parts and polar decomposition for complex numbers, we have similar ideas for normal matrices. We shall try to make most of our ideas eventually independent of the matrix representation of the linear transformation so they can be more easily generalised.

Firstly, some observations.

**Lemma 1.5.4.** *Let  $f$  be a linear transformation on a complex inner product space  $V$ .*

- (1) *If  $f$  is unitary then the eigenvalues of  $f$  are of absolute value 1.*
- (2) *If  $f$  is self-adjoint then the eigenvalues of  $f$  are real.*

*Proof.* (1) Suppose that  $f$  is unitary and that  $f(v) = av$  for some  $a \in \mathbb{C}$  and  $0 \neq v \in V$ . Then

$$a\bar{a}(v, v) = (av, av) = (f(v), f(v)) = (f^*(f(v)), v) = (v, v)$$

and so  $a\bar{a} = 1$ .

(2) Suppose that  $f$  is self-adjoint and that  $f(v) = av$  for some  $a \in \mathbb{C}$  and  $0 \neq v \in V$ . Then

$$a(v, v) = (av, v) = (f(v), v) = (v, f^*(v)) = (v, f(v)) = (v, av) = \bar{a}(v, v)$$

and so  $a = \bar{a}$ ; that is,  $a$  is real.  $\square$

It follows that a diagonal matrix for a self-adjoint linear transformation  $f$  has real entries. It is reasonable to define a self-adjoint matrix to be *non-negative* if it has non-negative (real) eigenvalues and *positive* if it has positive eigenvalues. There are some more convenient versions of the definition, however.

**Lemma 1.5.5.** *Let  $f$  be a linear transformation on a finite dimensional complex inner product space  $V$ . The following are equivalent:*

- (1)  $f$  is self-adjoint and all eigenvalues of  $f$  are non-negative
- (2)  $f = g^2$  for some self-adjoint  $g$ ;
- (3)  $f = hh^*$  for some  $h$ ;
- (4)  $f$  is self-adjoint and  $(f(v), v) \geq 0$  for all  $v \in V$ .

*Proof.* Suppose that (1) is true and let  $A$  be the diagonal matrix for  $f$  guaranteed by the Spectral Theorem. Suppose that  $A = \text{diag}(a_1, \dots, a_n)$ . Then  $a_i \geq 0$  so there are  $b_i \geq 0$  with  $b_i^2 = a_i$ . Let  $B = \text{diag}(b_1, \dots, b_n)$ . Then  $B^2 = A$ . Let  $g$  be the linear transformation corresponding to  $B$ .

If (2) is true then (3) is trivial; take  $h = g$ .

Suppose that (3) is true. Then  $f^* = (hh^*)^* = h^{**}h^* = hh^* = f$  and so  $f$  is self-adjoint. Also,

$$(f(v), v) = (hh^*(v), v) = (h^*(v), h^*(v)) \geq 0$$

and so (4) is true.

Finally, suppose that (4) is true. If  $a$  is an eigenvalue of  $f$  with associated eigenvector  $v$  then  $(f(v), v) = (av, v) = a(v, v) \geq 0$  and so  $a \geq 0$  as  $(v, v) > 0$ .

Thus we have proved that (1) implies (2) implies (3) implies (4) implies (1) and so the equivalence of all four.  $\square$

It is not difficult to show that, just as any complex number  $z$  can be expressed in the form  $z = |z| \exp^{i \arg z}$  with  $|z|$  real and non-negative and  $\exp^{i \arg z}$  of absolute value 1, we can write any normal linear transformation  $f$  as a product of a non-negative self-adjoint linear transformation with a unitary linear transformation. We sketch the argument. Find a diagonal matrix  $A$  for  $f$ ; say  $A = \text{diag}(z_1, \dots, z_n)$ . Write  $z_i = p_i u_i$  with  $p_i$  real and positive and  $u_i$  of absolute value 1. Set  $P = \text{diag}(p_1, \dots, p_n)$  and  $U = \text{diag}(u_1, \dots, u_n)$ . Then  $A = PU$ . Let  $p$  and  $u$  be the linear transformations corresponding to the matrices  $P$  and  $U$ . Then  $f = pu$  is the required decomposition.

In fact we do not need to assume that  $f$  is normal.

**Theorem 1.5.6.** *Let  $f$  be a linear transformation on a finite dimensional complex inner product space. Then there exists a non-negative linear transformation  $p$  and a unitary linear transformation  $u$  such that  $f = pu$ .*

*Proof.* We shall give the proof only in the case that  $f$  is invertible.

By Lemma 1.5.5,  $ff^*$  is non-negative (in fact positive if  $f$  is invertible) and so, by Lemma 1.5.5, we can find a non-negative  $p$  such that  $p^2 = ff^*$ . Since  $f$  is assumed invertible, so also is  $p$ . Then  $p^{-1}$  will also be self-adjoint. Consider  $p^{-1}f$ . We have

$$p^{-1}f(p^{-1}f)^* = p^{-1}(ff^*)(p^{-1})^* = p^{-1}(p^2)(p^{-1}) = 1_V.$$

Thus  $u = p^{-1}f$  is unitary and so  $f = pu$  as required.  $\square$

### 1.5.3 Commuting normal matrices

The spectral theorem gives us a tool to deal with matrices which commute and also with the problem of deciding which matrices commute with a given matrix.

**Theorem 1.5.7.** *Let  $f$  and  $g$  be normal linear transformations on a finite dimensional complex inner product space  $V$ . Suppose that  $fg = gf$ . Then there is an orthonormal basis for  $V$  such that the matrices of both  $f$  and  $g$  with respect to this basis are diagonal.*

*Proof.* Let  $V_1, \dots, V_m$  be the eigenspaces of  $f$  in  $V$  and suppose that the corresponding eigenvectors are  $a_1, \dots, a_m$ . We know from the Spectral Theorem (or by direct checking) that  $V_i$  and  $V_j$  are orthogonal if  $i \neq j$ .

We claim that each  $V_i$  is  $g$ -invariant. For, if  $v \in V_i$  and  $w = g(v)$ , then

$$f(w) = f(g(v)) = g(f(v)) = g(a_i v) = a_i g(v) = a_i w$$

and so  $g(v) = w \in V_i$ . Thus we can consider  $g_{V_i}$ . It is not difficult to check that  $g_{V_i}$  is still normal and so we can apply the Spectral Theorem to  $g_{V_i}$ . We then find an orthonormal basis  $\mathcal{B}_i$  of  $V_i$  which consists of eigenvectors of  $g_{V_i}$  and so of  $g$ . It clearly consists of eigenvectors of  $f$  since every element of  $V_i$  is an eigenvector of  $f$ .

The set  $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_m$  is an orthonormal basis of  $V$  and consists of vectors which are eigenvectors for both  $f$  and  $g$ . It follows that the matrices of  $f$  and  $g$  with respect to this basis are diagonal.  $\square$

**Theorem 1.5.8.** *Let  $f$  and  $g$  be normal linear transformations on a finite dimensional complex inner product space  $V$ . Then  $f$  and  $g$  commute if and only if there is a normal linear transformation  $h$  and polynomials  $p(X)$  and  $q(X)$  such that  $f = p(h)$  and  $g = q(h)$ .*

*Proof.* It is clear that two linear transformations of the form  $p(h)$  and  $q(h)$  must commute so we turn to the converse.

Suppose that  $f$  and  $g$  are normal linear transformations satisfying  $fg = gf$ . By the previous theorem, we can find an orthonormal basis of  $V$  so that the matrices  $A$  of  $f$  and  $B$  of  $g$ , with respect to this basis, are diagonal. Suppose that  $A = \text{diag}(a_1, \dots, a_n)$  and  $B = \text{diag}(b_1, \dots, b_n)$ .

Set  $C = \text{diag}(1, \dots, n)$ . There will be polynomials  $p$  and  $q$  so that  $p(i) = a_i$  and  $q(i) = b_i$ . (This requires the theory of ‘interpolation’ of polynomials. If you haven’t seen it before, try to do it from first principles for small values of  $n$ .) Thus  $p(C) = A$  and  $q(C) = B$ .

Let  $h$  be the linear transformation corresponding to  $C$ . Then  $p(h) = f$  and  $q(h) = g$ , as required.  $\square$

### 1.5.4 Exercises

- (1) Show that if  $A = UDU^*$  where  $D$  is a diagonal matrix and  $U$  is unitary, then  $A$  is a normal matrix. (The spectral theorem implies that the converse is true).
- (2) Show that a linear transformation  $f : V \rightarrow V$  on a complex inner product space  $V$  is normal if and only if  $(f(u), f(v)) = (f^*(u), f^*(v))$  for all  $u, v \in V$ .
- (3) Show that every normal matrix  $A$  has a square root; that is, a matrix  $B$  so that  $B^2 = A$ .
- (4) Must every complex square matrix have a square root?

- (5) Two linear transformations  $f$  and  $g$  on a finite dimensional complex inner product space are *unitarily equivalent* if there is a unitary linear transformation  $u$  such that  $g = u^{-1}fu$ . Two matrices are *unitarily equivalent* if their linear transformations, with respect to some fixed orthonormal basis, are *unitarily equivalent*.

Decide whether the following matrices are unitarily equivalent.

(a)

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

(b)

$$\begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

(c)

$$\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{bmatrix}.$$

- (6) Are  $f$  and  $f^*$  always unitarily equivalent?
- (7) If  $f$  is a normal linear transformation on a finite dimensional inner product space, and if  $f^2 = f^3$ , show that  $f = f^2$ . Show also that  $f$  is self-adjoint.
- (8) If  $f$  is a normal linear transformation on a finite dimensional inner product space show that  $f^* = p(f)$  for some polynomial  $p$ .
- (9) If  $f$  and  $g$  are normal linear transformations on a finite dimensional inner product space and  $fg = gf$ , show that  $f^*g = gf^*$ . (Harder) Prove that the same result holds assuming only that  $f$  is normal.
- (10) Let  $f$  be a linear transformation on a finite dimensional inner product space. Suppose that  $f$  commutes with  $f^*f$ ; that is, that  $f(f^*f) = (f^*f)f$ . We aim to show that  $f$  is normal.
- (a) Show that  $f^*f$  is normal.
- (b) Choose an orthonormal basis so that the matrix of  $f^*f$  takes the block diagonal form  $\text{diag}(A_1, \dots, A_m)$  where  $A_i = \lambda_i I_{m_i}$  and  $\lambda_i = \lambda_j$  only if  $i = j$ .

- (c) Show that  $f$  has matrix, with respect to this basis, of the block diagonal form  $\text{diag}(B_1, \dots, B_m)$  for some  $m_i \times m_i$  matrices  $B_i$ .
- (d) Deduce that  $B_i^* B_i = A_i$  and so that  $B_i^* B_i = B_i B_i^*$ .
- (e) Deduce that  $f$  is normal.
- (11) The following is a question (unedited) submitted to an Internet news group:

Hello,

I have a question hopefully any of you can help.

As you all know :

If we have a square matrix  $A$ , we can always find another square matrix  $X$  such that

$$X^{-1} * A * X = J$$

where  $J$  is the matrix with Jordan normal form. Column vectors of  $X$  are called principal vectors of  $A$ .

(If  $J$  is a diagonal matrix, then the diagonal members are the eigenvalues and column vectors of  $X$  are eigenvectors.)

It is also known that if  $A$  is real and symmetric matrix, then we can find  $X$  such that  $X$  is "orthogonal" and  $J$  is diagonal.

The question :

Are there any less strict conditions of  $A$  so that we can guarantee  $X$  orthogonal, with  $J$  not necessarily a diagonal ?

I would appreciate any answers and/or pointers to any references.

Can you help?