LINEAR ALGEBRA 2

VLADIMIR V. KISIL

MODULE SUMMARY. This lecture notes are based on a course at the University of Leeds.

COURSE OUTLINE

1. VECTOR SPACES AND SUBSPACES

Vector spaces have two built-in concepts.

(1) Vectors: these can be added, subtracted from each other.
(2) Scalars: these can be added, subtracted, multiplied, divided (except by zero).

Also, we can multiply a vector by a scalar (and get a vector). The set of scalars in this course is usualy $\mathbb{R}$ (the real numbers) or $\mathbb{C}$ (the complex numbers), but in general there are many other possibilities.

Example 1.1. Here are some vector spaces. The term 'vector space' really refers to the set of vectors.

(1) $\mathbb{R}^3$ (or more generally, $\mathbb{R}^n$, for any fixed $n \geqslant 1$). Here the set of scalars is $\mathbb{R}$. Observe

$$\begin{aligned}(x_1,x_2,x_3)+(y_1,y_2,y_3) &= (x_1+y_1,x_2+y_2,x_3+y_3)\\ (x_1,x_2,x_3)-(y_1,y_2,y_3) &= (x_1-y_1,x_2-y_2,x_3-y_3),\\ a(x_1,x_2,x_3) &= (ax_1,ax_2,ax_3),\end{aligned}$$

for any x_i, y_i, $a \in \mathbb{R}$.

(2) $\mathbb{C}^n$. This is similar, but here the scalars are complex numbers.

(3) $M_{m,n}(\mathbb{R})$, the set of all $m \times n$ (i.e. m rows, n columns) matrices over $\mathbb{R}$ (the set of scalars). Here the vectors ARE matrices, and vector addition is matrix addition.
$M_{m,n}(\mathbb{C})$, the set of all $m \times n$-matrices over $\mathbb{C}$ (and the set of scalars is $\mathbb{C}$).

(4) $P_n(t)$, the set of all polynomials in variable t of degree at most n, i.e.

$$P_n(t) := \{a_0 + a_1 t + \ldots + a_n t^n : a_i \in \mathbb{R}\}.$$

The scalars are $\mathbb{R}$.

(5) Fix a set X. The set $\mathbb{R}^X$ is the collection of all functions from X into $\mathbb{R}$. For $f, g \in \mathbb{R}^X$and $x \in X$, we have

$$(f+g)(x) = f(x)+g(x)$$
$$(af)(x) = af(x).$$

Likewise, $\mathbb{C}^X$ (functions from X to $\mathbb{C}$) is a vector space with scalars $\mathbb{C}$.

1.1. **Axioms for a vector space.** I'll generally use u, v, w, ...etc. for vectors, and a, b, c, ...for scalars. Let K denote the set of scalars, which in this module is always $\mathbb{R}$ or $\mathbb{C}$. (In general, K can be any *field*, but I am not defining this; K can even be finite.) Let V be a non-empty set of *vectors*, and suppose there are rules of addition of vectors, and scalar multiplication, so that if $u, v \in V$ then the sum $u+v \in V$, and if $a \in K$, $v \in V$, then $av \in V$. Then V is called a *vector space* over K if the following axioms hold.

(1) V is an abelian group under $+$, i.e.
   (a) $\forall u, v \in V$ $(u+v \in V)$ (V is closed under vector addition).
   (b) $(u+v = v+u)$, $\forall u, v \in V$ ($+$ is *commutative*, i.e. V is an *abelian* group with operation $+$).
   (c) $(u+v)+w = u+(v+w)$ $\forall u, v, w \in V$, ($+$ is *associative*);
   (d) there is a zero vector $0 \in V$ such that $\forall v \in V$ $(0+v = v+0 = v)$;
   (e) for all $v \in V$ there is an 'inverse vector' $-v$ in V such that $v+(-v) = (-v)+v = 0$;

(2) Scalar multlication satisfies (for all $u, v \in V$ and $a, b \in K$):
   (a) $\forall u \in V$, $\forall a \in \mathbb{K}$ $(av \in V)$ (V is closed under scalar multiplication).
   (b) $a(u+v) = au+av$;
   (c) $(a+b)u = au+bu$;
   (d) $(ab)u = a(bu)$;
   (e) $1u = u$.

In practice, we don't usually have to check all these, to see that something is a vector space.

Remark 1.2. Note that sometimes 0 denotes the zero vector, sometimes the zero scalar – this should be clear from context.

The next theorem collects facts really obvious for $\mathbb{R}^n$ or $\mathbb{C}^n$. We need to know them for *all* vector spaces, i.e. that they follow from the axioms for vector spaces.

Theorem 1.3. *Let V be a vector space over a field K. Then*

(1) *if $a \in K$ and $0 \in V$ then $a \cdot 0 = 0$;*
(2) *if $0 \in K$ and $v \in V$ then $0 \cdot v = 0$;*
(3) *if $av = 0$ then either $a = 0$ or $v = 0$;*
(4) $(-a)v = a(-v) = -(av)$.

Proof. I'll just do (i), (iii).

(i) $a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$. Now add $-(a \cdot 0)$ to both sides. We find $0 = a \cdot 0 + 0 = 0 + a \cdot 0$.
(iii) Suppose that $av = 0$ and $a \neq 0$. Then there is $a^{-1} \in K$, so $a^{-1}av = a^{-1} \cdot 0 = 0$. But also, $a^{-1}av = 1 \cdot v = v$. Hence $v = 0$.

□

Example 1.4. To demonstrate that choice of axioms is a delicate matter we will consider examples which violet these "obvious" properties.

On the set $\mathbb{R}^2$ of ordered pairs (x, y) define $+$ as usual, i.e. $(x, y) + (u, v) = (x + u, y + v)$, but define the multiplication by scalar as $\lambda \cdot (x, y) = (x, 0)$.

Since the addition is defined in the usual way we could easily check all corresponding axioms:

(1) $\mathbf{a} + \mathbf{b} \in V$.
(2) $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$.
(3) $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$.
(4) $\mathbf{a} + \mathbf{0} = \mathbf{a}$, where $\mathbf{0} = (0, 0)$.
(5) For any vector $\mathbf{a} = (x, y)$ there is $-\mathbf{a} = (-x, -y)$ such that $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$.

However for our strange multiplication $\lambda \cdot (x, y) = (x, 0)$ we should be more careful:

(1) $\lambda \cdot \mathbf{a} \in V$.
(2) $\lambda \cdot (\mathbf{a} + \mathbf{b}) = \lambda \cdot \mathbf{a} + \lambda \cdot \mathbf{b}$, because $\lambda \cdot ((x, y) + (u, v)) = \lambda \cdot (x + u, y + v) = (x + u, 0)$ $\lambda \cdot (x, y) + \lambda \cdot (u, v) = (x, 0) + (u, 0) = (x + u, 0)$.
(3) $(\lambda + \mu) \cdot \mathbf{a} \neq \lambda \cdot \mathbf{a} + \mu \cdot \mathbf{a}$, because $(\lambda + \mu) \cdot (x, y) = (x, 0)$ however $\lambda \cdot (x, y) + \mu \cdot (x, y) = (x, 0) + (x, 0) = (2x, 0)$.
(4) $(\lambda\mu) \cdot \mathbf{a} = \lambda(\mu \cdot \mathbf{a}) = \mu(\lambda \cdot \mathbf{a})$, i.e. the *associative law* holds (why?)
(5) $1 \cdot \mathbf{a} \neq \mathbf{a}$ (why?)

This demonstrate that the failing axioms could not be derived from the rest, i.e they are independent.

1.2. **Subspaces.** Let W be a subset of a vector space V over K. Then W is a *subspace* of V if W is also a vector space over K. This looks boring, with lots of axioms to check, but it's easy to check in practice, because of the following.

Theorem 1.5. *$W \subseteq V$ is a subspace of V if and only if*

(1) $0 \in W$,
(2) *W is closed under addition of vectors, i.e. if $v, w \in W$ then $v + w \in W$;*
(3) *W is closed under scalar multiplication, i.e. if $w \in W$ and $a \in K$ then $aw \in W$.*

Proof. Clearly, if W is a subspace then 1–3 hold. We show the converse, so assume 1–3 above. Then W is a group: the inverse of w is $-w = -(1 \cdot w) = (-1)w$, so lies in W.

The other properties all hold in W since they hold in V, e.g. $w + 0 = 0 + w = w$ holds if $w \in W$, since $W \subseteq V$ so $w \in V$, and the above rule holds in V. □

Corollary 1.6. *$W \subseteq V$ is a subspace of V if and only if*

(1) $0 \in W$;
(2) $av + bw \in W$ *whenever* $a, b \in K$ *and* $v, w \in W$.

Example 1.7. (1) Always, $\{0\}$ and V are subspaces of V;
(2) In the vector space $\mathbb{R}^2$, the set of points on a line through the origin (eg $2x+3y = 0$) is a subspace, but $2x + 3y = 1$ is not.
(3) In $M_{n,n}(\mathbb{R})$, consider the set X of *symmetric* $n \times n$ matrices. Here $A = (a_{ij})$ is *symmetric* if $a_{ij} = a_{ji}$ for all i, j, i.e. if $A = A^T$. An example of a symmetric matrix is $\begin{pmatrix} 1 & 2 & 4 \\ 2 & 3 & 6 \\ 4 & 6 & 5 \end{pmatrix}$. Then X is a subspace of $M_{n,n}(\mathbb{R})$.
(4) The solution space of any system of homogeneous linear equations in n variables (coefficients in $\mathbb{R}$) is a subspace of $\mathbb{R}^n$.

Remark 1.8. (1) If U_1, U_2 are subspaces of V then $U_1 \cap U_2$ is a subspace of V. For $0 \in U_1 \cap U_2$, and if $u, v \in U_1 \cap U_2$, and $a, b \in K$, then $au + bv \in U_1$, $au + bv \in U_2$, so $au + bv \in U_1 \cap U_2$, so Corollary 1.6 applies.
(2) In general, the *union* of two subspaces is not a subspace; for if $u_1 \in U_1$ and $u_2 \in U_2$, then possibly $u_1 + u_2 \notin U_1 \cup U_2$. For example, in $\mathbb{R}^2$, let

$$\begin{aligned} U_1 &:= \{(x, 0) : x \in \mathbb{R}\}, \\ U_2 &:= \{(0, y) : y \in \mathbb{R}\}. \end{aligned}$$

Then $(1, 0), (0, 1) \in U_1 \cup U_2$, but $(1, 1) \notin U_1 \cup U_2$.

Definition 1.9. Given a set S of vectors in a vector space V, a vector v is a *linear combination*of elements of S if $v = a_1 s_1 + \ldots + a_k s_k$ for some $s_1, \ldots, s_k \in S$ and $a_1, \ldots, a_k \in K$. We denote by $\mathrm{span}(S)$ the set of linear combinatioms of elements of S.

Theorem 1.10. *Let S be a non-empty subset of V. Then*

(1) $\mathrm{span}(S)$ *is a subspace of* V;
(2) *any subspace of V which contains S also contains* $\mathrm{span}(S)$.

Proof. As usual, we apply Corollary 1.6.

(1) Clearly $0 \in \mathrm{span}(S)$, e.g. as $0 = 0 \cdot s_1$ for any $s_1 \in S$. Let $v, w \in \mathrm{span}(S)$, say $v = a_1 s_1 + \ldots + a_k s_k$, $w = b_1 t_1 + \ldots + b_k t_k$ (so $a_i, b_i \in K$ and $s_i, t_i \in S$). Suppose $a, b \in K$. Then $av + bw =$

$$a(a_1 s_1 + \ldots + a_k s_k) + b(b_1 t_1 + \ldots + b_k t_k) = (aa_1)s_1 + \ldots + (aa_k)s_k + (bb_1)t_1 + \ldots + (bb_k)t_k.$$

(2) This is clear, (formally, by induction).

□

Example 1.11. In $\mathbb{R}^2$, the smallest subspace containing $(1,1),(2,3)$ is all of $\mathbb{R}^2$, as any vector (x,y) can be written as $a(1,1)+b(2,3)$, as we can always solve

$$\begin{aligned} a+2b &= x, \\ a+3b &= y. \end{aligned}$$

Thus, $\text{span}\{(1,1),(2,3)\} = \mathbb{R}^2$.

1.3. **Direct Sums.**

Definition 1.12. The *sum* of two subspaces U_1, U_2 of V is the set

$$U_1 + U_2 := \{u_1 + u_2 : u_1 \in U_1, u_2 \in U_2\}.$$

Theorem 1.13. *If U_1, U_2 are subspaces of V, then $U_1 + U_2$ is a subspace of V, and is the smallest subspace containing U_1, U_2.*

Proof. (1) $U_1 + U_2$ is a subspace. Indeed, $0 = 0 + 0 \in U_1 + U_2$, and if $u_1, u_1' \in U_1$, $u_2, u_2' \in U_2$, then

$$a(u_1+u_2) + b(u_1'+u_2') = (au_1 + bu_1') + (au_2 + bu_2') \in U_1 + U_2.$$

(2) Any subspace of V which contains U_1, U_2 must contain all the vectors $u_1 + u_2$, so contains $U_1 + U_2$.

□

Remark 1.14. We have $U_1 + U_2 = \text{span}(U_1 \cup U_2)$.

Example 1.15. Work in $\mathbb{R}^3$. Let

$$\begin{aligned} V_0 &:= \{(x,y,z) : x+y+z = 0\}, \\ V_1 &:= \{a(1,0,0) : a \in \mathbb{R}\}, \\ V_2 &:= \{a(1,0,-1) : a \in \mathbb{R}\}. \end{aligned}$$

Then $V_0 + V_1 = \mathbb{R}^3$, since for $a, b, c \in \mathbb{R}$,

$$(a,b,c) = (-b-c, b, c) + (a+b+c, 0, 0) \in V_0 + V_1.$$

Also $V_0 + V_2 = V_0$, since $V_2 \subset V_0$ (so anything in $V_0 + V_2$ already lies in V_0). The subspace $V_1 + V_2$ is a plane.

Definition 1.16. If U_1, U_2 are subspaces of V, and $U_1 \cap U_2 = \{0\}$, then $U_1 + U_2$ is called a *direct sum* of U_1, U_2, written $U_1 \oplus U_2$.

Remark 1.17. Any vector in $U_1 \oplus U_2$ has the form $u_1 + u_2$ (where $u_i \in U_i$) in a *unique* way. For suppose $u_1 + u_2 = u_1' + u_2'$ (where $u_i, u_i' \in U_i$). Then $u_1 - u_1' = u_2 - u_2'$. But the left hand side lies in U_1, and the right hand side lies in U_2, so $u_1 - u_1' \in U_1 \cap U_2 = \{0\}$, so $u_1 - u_1' = 0$, so $u_1 = u_1'$, and likewise $u_2 = u_2'$.

Example 1.18. (1) In Example 1.15, $\mathbb{R}^3 = V_0 \oplus V_1$. For if $(a,0,0) \in V_0$, then $a+0+0 = 0$, so $a = 0$.

(2) Let $V := P_2(t)$ be the set of all polynomials over $\mathbb{R}$ of degree at most 2 (see Example 1.1). Let

$$\begin{aligned} V_0 &:= \{at^2 : a \in \mathbb{R}\}, \\ V_1 &:= \{bt + c : b, c \in \mathbb{R}\}, \\ V_2 &:= \{a(t^2+1) : a \in \mathbb{R}\}. \end{aligned}$$

Then V_0, V_1, V_2 are subspaces of V (check this!). And $V = V_0 \oplus V_1 = V_2 \oplus V_1$. To see that $V = V_2 \oplus V_1$, observe that $at^2 + bt + c = a(t^2 + 1) + (bt + (c - a))$; it is trivial that $V_1 \cap V_2 = \{0\}$.

Definition 1.19. If $U_1 \oplus U_2 = V$, then U_2 is a *complement* of U_1 in V.

Remark 1.20. The complement is not necessarily unique; e.g., in Example 1.18, V_0 and V_2 are both complements of V_1.

2. Bases and dimension

The material in this section (up to and including 2.8) was done for subspaces of $\mathbb{R}^n$ in MATH1011.

Definition 2.1. A set S *spans* the vector space V if $\text{span}(S) = V$, i.e., any $v \in V$ can be written $v = a_1 s_1 + \ldots + a_n s_n$ for some $a_i \in K$, $s_i \in S$.

The set S is *linearly independent* if the only linear combination of members of S giving 0 is the trivial one, i.e., if $0 = a_1 s_1 + \ldots + a_n s_n$ (where $s_i \in S$ are distinct), then $a_1 = \ldots = a_n = 0$.

Example 2.2. Some examples in $V = \mathbb{R}^2$.

(1) $(1,1)$, $(2,3)$ span $\mathbb{R}^2$ and are linearly independent.
(2) $(1,1)$, $(2,2)$ span $\mathbb{R}^2$ but are not linearly independent.
(3) $(1,1)$, $(2,2)$, $(2,3)$ span $\mathbb{R}^2$ but are not linearly independent.
(4) $(1,1)$ is linearly independent but does not span $\mathbb{R}^2$.

Exercise 2.3.
(1) Any set including the zero vector is linearly independent.
(2) Any subset of a linearly independent subset is linearly independent as well.
(3) Any set, which includes a spanning set, is spanning as well.
(4) A spanning set expanded by a single vector become linearly dependent.
(5) A linearly independent set with any single vector removed cannot be spanning.

To check whether a set of r vectors in $\mathbb{R}^n$ is linearly independent, form the $r \times n$ matrix whose rows are these vectors, and put it into row reduced form. The original set is linearly independent if and only if the row reduced form has r non-zero rows.

Definition 2.4. A *basis* of V is a set of vectors in V which is both linearly independent and spans V.

Remark 2.5. The set $\{s_1, \ldots, s_n\}$ is a basis of V if and only if every vector $v \in V$ can be written as $v = a_1 s_1 + \ldots + a_n s_n$ ($a_i \in K$, s_i distinct members of S) in a *unique* way (up to order). To see this uniqueness, note that if $\{s_1, \ldots, s_n\}$ is linearly independent and $a_1 s_1 + \ldots + a_n s_n = b_1 s_1 + \ldots + b_n s_n$, then $0 = (a_1 - b_1)s_1 + \ldots + (a_n - b_n)s_n$, so $a_i = b_i = 0$ for all i, so the expressions are the same.

If $\{s_1, \ldots, s_n\}$ is a basis for V then V 'looks like' K^n; for each $v \in V$ can be written uniquely as $v = a_1 s_1 + \ldots + a_n s_n$, so we can identify v with $(a_1, \ldots, a_n) \in K^n$.

Theorem 2.6. *Suppose that V is a vector space over K, and $S \subseteq V$ is linearly independent. Then there is a basis B of V with $S \subseteq B$.*

Proof. We'll do the special case when S is finite, and there are $v_1, \ldots, v_m \in V$ with $V = \text{span}(v_1, \ldots, v_m)$. Suppose $S = \{s_1, \ldots, s_n\}$.

Case 1 $\mathrm{span}(S) = V$. Now S is a basis of V, so put $B := S$.

Case 2 There is some $v_i \in \{v_1, \ldots, v_m\}$ not in $\mathrm{span}(S)$. Now $S \cup \{v_i\}$ is linearly independent. For if $a_1 s_1 + \ldots + a_n s_n + b v_i = 0$, then $b = 0$, for otherwise $v_i = -b^{-1}(a_1 s_1 + \ldots + a_n s_n) \in \mathrm{span}(S)$. Hence, as S is linearly independent, also $a_1 = \ldots = a_n = 0$. Now repeat the above argument with $S \cup \{v_i\}$ in place of S. If we do not stop before then, we eventually obtain a linearly independent set $S \cup \{v_1, \ldots, v_m\}$, which certainly spans V so is a basis. If we stop before then, we have a basis.

□

Corollary 2.7. *Suppose V is a vector space over K, $S \subseteq V$ is linearly independent, and $A \subseteq V$ satisfies $\mathrm{span}(A) = V$. Then there is a basis B of V with $S \subseteq B \subseteq S \cup A$.*

Proof. Our proof of Theorem 2.6 gave this, under the assumption that S, A are finite (put $A = \{v_1, \ldots, v_m\}$ in the theorem). We omit the proof in general. □

Note that the proof of Theorem 2.6 give you a *method* for answering questions like Problem Sheet 2 Q1(a).

Corollary 2.8. *Suppose S is a vector space over K and $A \subseteq V$ spans V. Then A contains a basis of V.*

Proof. Apply Corollary 2.7 with $S = \emptyset$. □

The next lemma is the key to showing any two bases of a vector space have the same size (so we can define *dimension*).

Lemma 2.9 (Exchange Lemma). (1) *Suppose $u_1, \ldots, u_n, v$ are vectors in a vector space V, and $v \in \mathrm{span}(u_1, \ldots, u_{n-1}, u_n)$, but $v \notin \mathrm{span}(u_1, \ldots, u_{n-1})$. Then $u_n \in \mathrm{span}(u_1, \ldots, u_n$*

(2) *Under the same assumptions, if $u_1, \ldots, u_{n-1}, u_n$ are linearly independent, so are $u_1, \ldots, u_{n-}$*

Proof. (1) Since $v \in \mathrm{span}(u_1, \ldots, u_n)$, there are scalars b_i such that

$$v = b_1 u_1 + \ldots + b_{n-1} u_{n-1} + b_n u_n. \tag{1}$$

Now $b_n \neq 0$, as otherwise $v \in \mathrm{span}(u_1, \ldots, u_{n-1})$. So

$$u_n = b_n^{-1}(v - b_1 u_1 - \ldots - b_{n-1} u_{n-1}) = b_n^{-1} v - b_n^{-1} b_1 u_1 - \ldots - b_n^{-1} b_{n-1} u_{n-1},$$

so $u_n \in \mathrm{span}(u_1, \ldots, u_{n-1}, v)$.

(2) Suppose that $u_1, \ldots, u_n$ are linearly independent, and that

$$a_1 u_1 + \ldots + a_{n-1} u_{n-1} + a_n v = 0.$$

Substituting from (1),

$$a_1 u_1 + \ldots + a_{n-1} u_{n-1} + a_n (b_1 u_1 + \ldots b_{n-1} u_{n-1} + b_n u_n) = 0$$

so

$$(a_1 + a_n b_1) u_1 + \ldots + (a_{n-1} + a_n b_{n-1}) u_{n-1} + a_n b_n u_n = 0.$$

As $u_1, \ldots, u_n$ are linearly independent, all the coefficients are zero, so $a_n b_n = 0$. As $b_n \neq 0$, $a_n = 0$, so $a_1 u_1 + \ldots a_{n-1} u_{n-1} = 0$. But $u_1, \ldots, u_{n-1}$ are linearly independent, so this forces $a_1 = \ldots = a_{n-1} = 0$.

□

Theorem 2.10. *Suppose that A, B are both bases of the vector space V. Then A and B have the same number of elements.*

Proof. For convenience, we assume that A, B are finite, and that A is at least as big as B. Let $B = \{v_1, \dots, v_n\}$, and choose distinct $u_1, \dots, u_n \in A$. We shall show that $A = \{u_1, \dots, u_n\}$.

Now $u_1 \in V = \operatorname{span}(v_1, \dots, v_n)$, so $u_1 = a_1 v_1 + \dots + a_n v_n$ for some scalars a_i. Not all the a_i are zero, as otherwise $u_1 = 0$, contradicting that A is linearly independent. So, without loss, suppose $a_1 \neq 0$. Now $u_1 \notin \operatorname{span}(v_2, \dots, v_n)$, as otherwise we'd have $u_1 = 0 \cdot v_1 + b_2 v_2 + \dots + b_n v_n = a_1 v_1 + \dots + a_n v_n$, contradicting that B is linearly independent.

So by Lemma 2.9, $v_1 \in \operatorname{span}(u_1, v_2, \dots, v_n)$, and (using 2.9(ii), $\{u_1, v_2, \dots, v_n\}$ is a basis of V. Next, we apply the above argument to u_2 (in place of u_1), then to u_3, and so on. Eventually, we find that $\{u_1, \dots, u_n\}$ is a basis of V, and so $A = \{u_1, \dots, u_n\}$. □

Definition 2.11. The number of elements of a basis of V (which by Theorem 2.10 does not depend on the choice of basis) is called the *dimension* of V, denoted $\dim(V)$.

Corollary 2.12. *Let V be a vector space and $n = \dim(V)$. If a set S has exactly n vectors then the following are equivalent:*

(1) *S is linearly independent;*
(2) *S spans V;*
(3) *S is basis of V.*

Example 2.13.
(1) $\mathbb{C}^n$, a vector space over $\mathbb{C}$, has a *standard basis* $(1, 0, \dots, 0), (0, 1, 0, \dots, 0)$ $(0, \dots, 0, 1)$, so has dimension n.
(2) $P_n(t)$ has basis $1, t, t^2, \dots, t^n$, so has dimension $n + 1$.
(3) $M_{m,n}(\mathbb{R})$ has a basis $\{E_{ij} : 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n\}$, so has dimension mn. Here, E_{ij} is the $m \times n$ matrix with a 1 in the (i, j)-entry, zeros elsewhere.
(4) The set of points on the line $x + 2y = 0$ can be described as $\{a(1, \frac{-1}{2}) : a \in \mathbb{R}\}$, so has basis $(1, \frac{-1}{2})$, so dimension 1.
(5) The plane $x + 2y + 3z = 0$ in $\mathbb{R}^3$ has solution set

$$\{(-2a - 3b, a, b) : a, b \in \mathbb{R}\} = \{a(-1, 1, 0) + b(-3, 0, 1) : a, b \in \mathbb{R}\},$$

so basis $(-2, 1, 0), (-3, 0, 1)$, and dimension 2. Suppose we want to expand this set to a baiss of $\mathbb{R}^3$. Try to add in the vectors $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, and the first time the dimension goes up to 3, we have a basis. In fact, $(-2, 1, 0), (-3, 0, 1), (1, 0, 0)$ is a basis of $\mathbb{R}^3$.

Finally, an important theorem.

Theorem 2.14. *Let U, V be subspaces of the finite dimensional vector space W. Then*

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

Proof. Let $w_1, \dots, w_m$ be a basis for $U \cap V$. Using Theorem 2.6, we can extend this set to a basis $w_1, \dots, w_m, u_1, \dots, u_r$ for U, and to a basis $w_1, \dots, w_m, v_1, \dots, v_s$ for V. Now $\dim(U \cap V) = m$, $\dim(U) = m + r$, $\dim(V) = m + s$.

Lemma 2.15. *$w_1, \dots, w_m, u_1, \dots, u_r, v_1, \dots, v_s$ is a basis for $U + V$.*

Assuming the above Lemma, the theorem is proved, for then $\dim(U + V) = m + r + s$, so

$$\dim(U) + \dim(V) = (m + r) + (m + s) = m + (m + r + s) = \dim(U \cap V) + \dim(U + V).$$

□

Proof of Lemma 2.15. First, spanning. So let $z \in U + V$. Then $z = u + v$ for some $u \in U$, $v \in V$. We can write

$$\begin{aligned} u &= a_1 w_1 + \ldots + a_m w_m + b_1 u_1 + \ldots + b_r u_r, \\ v &= c_1 w_1 + \ldots + c_m w_m + d_1 v_1 + \ldots + d_s v_s, \end{aligned}$$

so

$$z = u + v = (a_1 + c_1)w_1 + \ldots + (a_m + c_m)w_m + b_1 u_1 + \ldots + b_r u_r + d_1 v_1 + \ldots + d_s v_s,$$

so our set spans $U + V$.

Next, linear independence. So suppose

$$a_1 w_1 + \ldots + a_m w_m + b_1 u_1 + \ldots + b_r u_r + c_1 v_1 + \ldots + c_c v_s = 0 \tag{2}$$

and put $w := \sum a_i w_i$, $u = \sum b_i u_i$ and $v = \sum c_i v_i$. Now $v = -w - u \in U$, so $v \in U \cap V$, so $v = d_1 w_1 + \ldots + d_m w_m$ for some d_i. Thus:

$$v = c_1 v_1 + \ldots + c_s v_s = d_1 w_1 + \ldots + d_m w_m.$$

Hence $c_1 v_1 + \ldots + c_s v_s - d_1 w_1 - \ldots - d_m w_m = 0$. But $w_1, \ldots, w_m, v_1, \ldots, v_s$ are linearly independent, so $c_1 = \ldots = c_s = d_1 = \ldots = d_m = 0$, so $v = 0$. Thus, (2) gives

$$a_1 w_1 + \ldots + a_m w_m + b_1 u_1 + \ldots + b_b u_r = 0,$$

so (as $w_1, \ldots, w_m, u_1, \ldots, u_r$ are linearly independent), $a_1 = \ldots = a_m = b_1 = \ldots = b_r = 0$ as required. □

Corollary 2.16. *If $V = V_1 \oplus V_2$, then* $\dim(V) = \dim(V_1) + \dim(V_2)$.

Proof. This follows from the theorem, as in this case $\dim(V_1 \cap V_2) = 0$. □

Example 2.17. Inside $V = \mathbb{R}^4$, let

$$\begin{aligned} V_1 &:= \{(a, b, c, 0) : a, b, c \in \mathbb{R}\}, \\ V_2 &:= \{(a, a, b, b) : a, b \in \mathbb{R}\}. \end{aligned}$$

Then $V_1 \cap V_2 := \{(a, a, 0, 0) : a \in \mathbb{R}\} := \{a(1, 1, 0, 0) : a \in \mathbb{R}\}$, so $\dim(V_1 \cap V_2) = 1$. Clearly $\dim(V_1) = 3$, and also $\dim(V_2) = 2$ (V_2 has basis $(1, 1, 0, 0), (0, 0, 1, 1)$). Now

$$\dim(V_1 + V_2) = \dim(V_1) + \dim(V_2) - \dim(V_1 \cap V_2) = 3 + 2 - 1 = 4.$$

Thus, $V_1 + V_2 = \mathbb{R}^4$ (but it is not a direct sum).

3. Linear Transformations

Definition 3.1. Let U, V be vector spaces over K. A mapping $\alpha : U \to V$ is a *linear transformation* (or linear mapping) if

$$\begin{aligned} \alpha(u_1 + u_2) &= \alpha(u_1) + \alpha(u_2) \text{ and} \\ \alpha(au) &= a\alpha(u) \end{aligned}$$

for all $u_1, u_2, u \in U$, $a \in K$.

Equivalently, α is a linear transformation if $\alpha(a_1 u_1 + a_2 u_2) = a_1 \alpha(u_1) + a_2 \alpha(u_2)$ for all $a_1, a_2 \in K$ and $u_1, u_2 \in U$.

Exercise 3.2. Show that under linear transformation:

(1) The null vector of U is mapped to the null vector of V.

(2) Any straight line in U is mapped to a straight line in V.

Theorem 3.3. *Any linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ has the form*

$$\alpha(x_1,\dots,x_m)^T = A(x_1,\dots,x_m)^T,$$

where $A = (a_{ij})$ is an $n \times m$ matrix. The basis vectors $(1,0,\dots,0)^T$, …, $(0,\dots,0,1)^T$ map under α to the columns of A.

Remark 3.4. Here, to save on trees, we write $(x_1,\dots,x_m)^T$ (the transpose) for the column vector $\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$.

Example 3.5. Consider linear mappings $\mathbb{R}^2 \to \mathbb{R}^2$. As examples, we have the identity transformation (matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$), expansions (with matrix $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$), rotations (matrix $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ for rotation anticlockwise by θ) and a reflection in the x-axis (matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$).

Before proving the theorem, we fix the following notation, used throughout the course. We let $e_1 := (1,0,\dots,0)^T$, $e_2 := (0,1,0,\dots,0)^T, \dots, e_m := (0,\dots,0,1)^T$ be the *standard basis* of $\mathbb{R}^m$.

Proof of Theorem 3.3. By linearity of α,

$$\alpha(x_1,\dots,x_m)^T = \alpha(\sum_{j=1}^{m} x_j e_j) = \sum_{j=1}^{m} x_j \alpha(e_j) = \sum_{j=1}^{m} x_j f_j,$$

where $f_1 := \alpha(e_1), \dots, f_m := \alpha(e_m)$. Write $f_j = (a_{1j},\dots,a_{nj})^T$, the jth column of A. Then $\alpha(x_1,\dots,x_m)^T = \sum_{j=1}^{m} x_j f_j =$

$$= x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ \vdots \\ a_{n2} \end{pmatrix} + \dots + x_m \begin{pmatrix} a_{1m} \\ \vdots \\ a_{nm} \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix},$$

as claimed in the theorem. □

Remark 3.6. If we choose any $f_1,\dots f_m \in \mathbb{R}^n$, then there is a *unique* linear mapping $\beta : \mathbb{R}^m \to \mathbb{R}^n$ with $\beta(e_i) = f_i$ for each $j = 1,\dots,m$. For we must have $\beta(\sum_{j=1}^{m} x_j e_j) = \sum_{j=1}^{m} x_j f_j$, by linearity.

Now we do something similar for linear transformations between *arbitrary* vector spaces.

Theorem 3.7. *Let U, V be vector spaces over K of dimensions m, n respectively, and let $u_1,\dots,u_m$ be a basis of U, $v_1,\dots v_n$ a basis of V. Let $\alpha : U \to V$ be a linear transformation. For each $j = 1,\dots,m$, put $\alpha(u_j) = \sum_{i=1}^{n} a_{ij} v_i$, and let $A = (a_{ij})$, an $n \times m$ matrix. Then for any $c_1,\dots,c_m \in K$, $\alpha(\sum_{j=1}^{m} c_j u_j) = \sum_{i=1}^{n} d_i v_i$, where $(d_1,\dots,d_n)^T = A(c_1,\dots,c_m)^T$.*

Proof. We have

$$\begin{aligned}
\alpha\left(\sum_{j=1}^{m} c_j u_j\right) &= \sum_{j=1}^{m} c_j \alpha(u_j) \qquad \text{(by linearity)} \\
&= \sum_{j=1}^{m} c_j \sum_{i=1}^{n} a_{ij} v_i \\
&= \sum_{j=1}^{m} \sum_{i=1}^{n} a_{ij} c_j v_i \\
&= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_{ij} c_j\right) v_i \\
&= \sum_{i=1}^{n} d_i v_i.
\end{aligned}$$

□

Remark 3.8. The j-th column of A gives the coordinates of $\alpha(u_j)$ with respect to the basis $v_1, \ldots, v_n$ of V.

Much of the aim of the rest of the module is to choose bases carefully so that the matrix for A (which depends on the choice of bases) has a nice shape.

Definition 3.9. Let $\alpha : U \to V$ be a linear transformation. Then the *image* of α is $\mathrm{Im}(\alpha) := \{\alpha(u) : u \in U\}$, and the *kernel* of α is $\mathrm{Ker}(\alpha) := \{u \in U : \alpha(u) = 0\}$. The kernel is also called the *null space* of α.

Theorem 3.10. *Let $\alpha : U \to V$ be a linear transformation. Then $\mathrm{Im}(\alpha)$ is a subspace of V and $\mathrm{Ker}(\alpha)$ is a subspace of U.*

Proof. First, observe that $\alpha(0) = \alpha(0 \cdot u) = 0 \cdot \alpha(u) = 0$, so $0_V \in \mathrm{Im}(\alpha)$, and $0_U \in \mathrm{Ker}(\alpha)$.

To see $\mathrm{Im}(\alpha)$ is a subspace: if $v_1, v_2 \in \mathrm{Im}(\alpha)$ then there are $u_1, u_2 \in U$ with $\alpha(u_1) = v_1$, $\alpha(u_2) = v_2$. Now if $a_1, a_2 \in K$, then $a_1 v_1 + a_2 v_2 = \alpha(a_1 u_1 + a_2 u_2) \in \mathrm{Im}(\alpha)$.

To see $\mathrm{Ker}(\alpha)$ is a subspace, let $u_1, u_2 \in \mathrm{Ker}(\alpha)$. Then $\alpha(a_1 u_1 + a_2 u_2) = a_1 \alpha(u_1) + a_2 \alpha(u_2) = a_1 \cdot 0 + a_2 \cdot 0 = 0$, so $a_1 u_1 + a_2 u_2 \in \mathrm{Ker}(\alpha)$. □

Example 3.11. Let $\alpha : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear map

$$(x, y, z)^T \mapsto (x + y - 2z, x - z, x - y)^T.$$

We shall find the matrix for α (with respect to the standard basis of $\mathbb{R}^3$) and then find bases for $\mathrm{Im}(\alpha)$ and $\mathrm{Ker}(\alpha)$. So we see how α acts on the standard basis. Now $\alpha(1,0,0)^T = (1,1,1)^T$, $\alpha(0,1,0)^T = (1,0,-1)^T$, and $\alpha(0,0,1)^T = (-2,-1,0)^T$. Hence the matrix for α is

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

Now $\mathrm{Ker}(\alpha)$ consists of all solutions to the simultaneous equations $x+y-2z = 0$, $x-z = 0$ and $x - y = 0$. By row operations, this solution space is $\{a(1,1,1) : a \in \mathbb{R}\}$ so $\{(1,1,1)\}$ is a

basis for $\mathrm{Ker}(\alpha)$. Also, $\mathrm{Im}(\alpha)$ is spanned by the columns of A. To get a basis, write these as rows (ie form A^T), and do row operations. We can reduce A^T to

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

so the non-zero rows $\{(1,1,1),(0,-1,-2)\}$ form a basis for $\mathrm{Im}(\alpha)$.

Definition 3.12. If $\alpha : U \to V$ is a linear transformation, then $r(\alpha) := \dim(\mathrm{Im}(\alpha))$ is the *rank* of α, and $n(\alpha) = \dim(\mathrm{Ker}(\alpha))$ is the *nullity* of α.

Theorem 3.13 (Rank and Nullity Theorem)**.** *Let $\alpha : U \to V$ be a linear transformation. Then*

$$r(\alpha) + n(\alpha) = \dim(U).$$

Proof. As usual, we shall assume $\dim(U)$ is finite.

Let $v_1, \ldots, v_r$ be a basis for $\mathrm{Im}(\alpha)$. Then there are $u_1, \ldots, u_r \in U$ with $\alpha(u_i) = v_i$ for each i. Also, choose a basis $t_1, \ldots, t_k$ for $\mathrm{Ker}(\alpha)$. The theorem follows from the following claim.

Lemma 3.14. *$u_1, \ldots, u_r, t_1, \ldots, t_k$ is a basis for U.*

Proof of Lemma. Spanning. If $u \in U$, then $\alpha(u) \in \mathrm{Im}(\alpha)$, so we can write $\alpha(u) = a_1v_1 + \ldots + a_rv_r = a_1\alpha(u_1) + \ldots + a_r\alpha(u_r)$. Then

$$\alpha(u - (a_1u_1 + \ldots + a_ru_r)) = \alpha(u) - (a_1\alpha(u_1) + \ldots + a_r\alpha(u_r)) = 0.$$

Hence $u - (a_1u_1 + \ldots + a_ru_r) \in \mathrm{Ker}(\alpha)$, so has form $b_1t_1 + \ldots + b_kt_k$ for some scalars b_i. Then $u = a_1u_1 + \ldots + a_ru_r + b_1t_1 + \ldots + b_kt_k$.

Linear Independence. Suppose $a_1u_1 + \ldots + a_ru_r + b_1t_1 + \ldots + b_kt_k = 0$. Then

$$0 = \alpha(a_1u_1 + \ldots + a_ru_r + b_1t_1 + \ldots + b_kt_k) = a_1v_1 + \ldots + a_rv_r.$$

Hence $a_1 = \ldots = a_r = 0$, as $v_1, \ldots, v_r$ are linearly independent. Then $b_1t_1 + \ldots + b_kt_k = 0$. But $t_1, \ldots, t_k$ are linearly independent, so also $b_1 = \ldots = b_k = 0$. □

□

Remark 3.15. In Example 3.11, $\dim(U) = \dim(\mathbb{R}^3) = 3$, $n(\alpha) = 1$, and $r(\alpha) = 2$. The proof of Theorem 3.13 shows that $(1,1,1)^T$, $\alpha^{-1}(1,1,1)^T$, $\alpha^{-1}(0,-1,-2)^T$ is a basis for $\mathbb{R}^3$.

3.1. **Discussion of Simultaneous Equations.** Suppose we have a system of n linear simultaneous equations

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1m}x_m &= b_1, \\ a_{21}x_1 + \ldots + a_{2m}x_m &= b_2 \\ \ldots \quad &\ldots \quad \ldots \\ a_{n1}x_1 + \ldots + a_{nm}x_m &= b_n, \end{aligned}$$

so of the form $Ax = b$, where $A = (a_{ij})$ is an $n \times m$ matrix, $x = (x_1, \ldots, x_m)^T$, and $b = (b_1 \ldots, b_n)^T \in \mathbb{R}^n$.

Multiplication by A gives a linear map $\alpha : \mathbb{R}^m \to \mathbb{R}^n$. Now $r(\alpha) = n \Leftrightarrow \mathrm{Im}(\alpha) = \mathbb{R}^n$, which holds if and only if we can solve the above system for *every* $b \in \mathbb{R}^n$.

Also, $n(\alpha) = 0$ if and only if the homogeneous system $Ax = (0,\dots,0)^T$ has the zero vector in $\mathbb{R}^m$ as the only solution. This means also that if $\alpha(x) = \alpha(y) = b$, then $\alpha(x-y) = 0$, so $x - y = 0$; that is, $n(\alpha) = 0$ means that solutions to $Ax = b$, if they exist, are *unique*.

Now $r(\alpha) + n(\alpha) = m$, so if $m < n$, then $r(\alpha) < n$ and we can't always solve the system. If $m > n$, then $m > r(\alpha)$ so $n(\alpha) > 0$, wo we cannot solve *uniquely*. If $m = n$, then either $r(\alpha) = n$ and $n(\alpha) = 0$ (so for all $b \in \mathbb{R}^n$ there is a *unique* solution of $Ax = b$), or $r(\alpha) < n$ and $n(\alpha) > n$ (so for some b there is no solution, but if there is a solution there are infinitely many).

3.2. **Composition of Mappings.**

Lemma 3.16. *Let U, V, W be vector spaces, and $\alpha : U \to V$, $\beta : V \to W$ be linear transformations. Then the function $\beta\alpha : U \to W$ defined by $\beta\alpha(u) = \beta(\alpha(u))$ is linear.*

Proof. We have

$$\begin{aligned}(\beta\alpha)(a_1u_1 + a_2u_2) &= \beta(\alpha(a_1u_1 + a_2u_2)) \\ &= \beta(a_1\alpha(u_1) + a_2\alpha(u_2)) \qquad (\text{as } \alpha \text{ is linear}) \\ &= a_1\beta(\alpha(u_1)) + a_2\beta(\alpha(u_2)) \qquad (\text{as } \beta \text{ is linear}) \\ &= a_1(\beta\alpha)(u_1) + a_2(\beta\alpha)(u_2).\end{aligned}$$

□

Theorem 3.17. *Suppose that $\dim(U) = m$, with basis $u_1,\dots,u_m$, $\dim(V) = n$, with basis $v_1,\dots,v_n$, $\dim(W) = p$, with basis $w_1,\dots,w_p$. Suppose $\alpha : U \to V$ and $\beta : V \to W$ are linear transformations with matrices A, B with respect to these bases. Then $\beta\alpha$ has matrix BA with respect to the bases $u_1,\dots,u_m$ of U and $w_1,\dots,w_p$ of W.*

Proof. By Theorem 3.3,

$$\begin{aligned}\alpha(u_j) &= \sum_{i=1}^{n} a_{ij}v_i \text{ for } j = 1,\dots,m \\ \beta(v_i) &= \sum_{k=1}^{p} b_{ki}w_k \text{ for } i = 1,\dots,n,\end{aligned}$$

where $A = (a_{ij})$ and $B = (b_{ij})$. Put $C = BA = (c_{ij})$. Now,

$$\begin{aligned}(\beta\alpha)(u_j) &= \beta(\sum_{i=1}^{n} a_{ij}v_i)\\ &= \sum_{i=1}^{n} a_{ij}\beta(v_i)\\ &= \sum_{i=1}^{n}\sum_{k=1}^{p} b_{ki}a_{ij}w_k\\ &= \sum_{k=1}^{p}(\sum_{i=1}^{n} b_{ki}a_{ij})w_k\\ &= \sum_{k=1}^{p} c_{kj}w_k.\end{aligned}$$

Thus, $\beta\alpha$ has matrix C. □

3.3. **Inverses.** We adopt the following convention. If $\alpha : U \to U$, then we use the same basis for U on both sides.

Definition 3.18. If $\alpha : U \to U$ is linear, then α is *invertible* if α is 1-1 and onto. In this case, for each $b \in U$, the equation $\alpha(u) = b$ has a unique solution, which we write as $u = \alpha^{-1}(b)$. We call α^{-1} the *inverse* of α.

Proposition 3.19. *Suppose $\alpha : U \to U$ is invertible. Then*

(1) $r(\alpha) = \dim(U)$, $n(\alpha) = 0$;
(2) $\alpha^{-1} : U \to U$ *is linear.*
(3) *if α has matrix A with respect to basis $u_1, \dots, u_n$, the α^{-1} has matrix A^{-1} with respect to this basis.*

Proof. First, since α is onto, $\mathrm{Im}(\alpha) = U$, so $r(\alpha) = \dim(U)$. Since $\alpha(u) = b$ can be solved *uniquely*, $\ker(\alpha) = \{0\}$, so $n(\alpha) = 0$.

To see that α^{-1} is linear, suppose $\alpha^{-1}(v_1) = u_1$ and $\alpha^{-1}(v_2) = u_2$, and let a_1, a_2 be scalars. Then $\alpha(a_1u_1 + a_2u_2) = a_1v_1 + a_2v_2$, so

$$\alpha^{-1}(a_1v_1 + a_2v_2) = a_1u_1 + a_2u_2 = a_1\alpha^{-1}(v_1) + a_2\alpha^{-1}(v_2).$$

Finally, if α^{-1} has matrix B, then since $\alpha\alpha^{-1} = \alpha^{-1}\alpha = \mathrm{id}$ which has matrix I_n, by Theorem 3.17 $AB = BA = I$, so $B = A^{-1}$. □

Theorem 3.20 (The $AP = PB$ Theorem)**.** *Let $\alpha : U \to U$ be a linear transformation, and $u_1, \dots, u_n$ and $v_1, \dots, v_n$ be two bases of U. Suppose that α is respresented by the matrix A with respect to $u_1, \dots, u_n$, and by B with respect to basis $v_1, \dots, v_n$. Then there is a nonsingular matrix P such that $AP = PB$, and $P = (p_{ij})$ is given by $v_j = \sum_{i=1}^{n} p_{ij}u_i$ (for each j).*

Remark 3.21. (1) The conclusion says $B = P^{-1}AP$.
(2) P is a 'change of basis' matrix. It's jth column is $(p_{1j}, \dots, p_{nj})^T$ – the coordinates of v_j with respect to the basis $u_1, \dots, u_n$.

Proof of Theorem 3.20. As α is represented by B in the v-basis,

$$\begin{aligned}\alpha(v_j) &= \sum_{i=1}^{n} b_{ij} v_i \\ &= \sum_{i=1}^{n} b_{ij} \sum_{k=1}^{n} p_{ki} u_k \\ &= \sum_{k=1}^{n} \sum_{i=1}^{n} p_{ki} b_{ij} u_k \\ &= \sum_{k=1}^{n} (PB)_{kj} u_k.\end{aligned}$$

Also,

$$\begin{aligned}\alpha(v_j) &= \alpha(\sum_{i=1}^{n} p_{ij} u_i) \\ &= \sum_{i=1}^{n} p_{ij} \alpha(u_i) \\ &= \sum_{i=1}^{n} \sum_{k=1}^{n} p_{ij} a_{ki} u_k \qquad \text{as } A \text{ represents } \alpha \text{ with respect to } u\text{-basis} \\ &= \sum_{k=1}^{n} \sum_{i=1}^{n} a_{ki} p_{ij} u_k \\ &= \sum_{k=1}^{n} (AP)_{kj} u_k.\end{aligned}$$

Thus, $(PB)_{kj} = (AP)_{kj}$ for all k, j, so $PB = AP$.

Finally, we show P is invertible. Now the jth column of P is v_j written in the u-basis. Thus, working in the u-basis, multiplication by P takes each u_i to v_i. This is invertible, with the inverse transformation taking v_i to u_i. □

Example 3.22. Consider $\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\alpha \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}$. The u_1, u_2 basis will be $u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $u_2 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$, and the v_1, v_2 basis will be $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $v_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. First, we find A, the matrix of α for the u-basis. Now

$$\begin{aligned}\alpha(u_1) &= \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 \cdot u_1 + 0 \cdot u_2, \qquad \text{and} \\ \alpha(u_2) &= \alpha \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \end{pmatrix} = 0 \cdot u_1 - 1 \cdot u_2.\end{aligned}$$

Hence $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Next, we find B, the matrix of α for the v-basis. Now

$$\begin{aligned}\alpha(v_1) &= \alpha\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1\\-1\end{pmatrix} = 3v_1 - 2v_2 \qquad \text{and}\\ \alpha(v_2) &= \alpha\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}1\\-2\end{pmatrix} = 4v_1 - 3v_2.\end{aligned}$$

Hence, $B = \begin{pmatrix}3 & 4\\ -2 & -3\end{pmatrix}$.

To find P: its jth column is v_j written in the u-basis.

$$\begin{aligned}v_1 &= \begin{pmatrix}1\\1\end{pmatrix} = 1\cdot\begin{pmatrix}1\\0\end{pmatrix} + \frac{1}{2}\cdot\begin{pmatrix}0\\2\end{pmatrix}, \qquad \text{and}\\ v_2 &= \begin{pmatrix}1\\2\end{pmatrix} = 1\cdot\begin{pmatrix}1\\0\end{pmatrix} + 1\cdot\begin{pmatrix}0\\2\end{pmatrix},\end{aligned}$$

so $P = \begin{pmatrix}1 & 1\\ \frac{1}{2} & 1\end{pmatrix}$. Finally, a calculation shows $AP = PB$.

4. Diagonalisation of matrices

From now on, $\alpha : V \to V$ is a linear transformation, where V is an n-dimensional real or complex vector space. If α is represented by the $n \times n$ matrix A with respect to one fixed basis, then with respect to any other basis it has matrix $P^{-1}AP$, for some non-singular P (by Theorem 3.20).

Definition 4.1. Two matrices A, B are said to be *similar* if there is an invertible matrix P such that $P^{-1}AP = B$; that is, if A, B represent the same linear transformation with respect to (different) bases.

Exercise 4.2. Show that the relation "matrices A and B are similar" is an equivalence.

To investigate similarity of matrices we need the following notions.

Definition 4.3. Let $\alpha : V \to V$ be linear. A scalar λ is said to be an *eigenvalue* of α if there is a **non-zero** vector $v \in V$ such that $\alpha(v) = \lambda v$. The vector v is called an *eigenvector* corresponding to the eigenvalue λ.

Remark 4.4.

(1) λ is an eigenvalue of $\alpha \Leftrightarrow$ there is $v \neq 0$ such that $(\alpha - \lambda I)v = 0$
$\Leftrightarrow$ $\mathrm{Ker}(\alpha - \lambda I) \neq \{0\}$
$\Leftrightarrow$ $n(\alpha - \lambda I) \neq 0$
$\Leftrightarrow$ $r(\alpha - \lambda I) \neq n$ by Rank and Nullity Theorem
$\Leftrightarrow$ $\alpha - \lambda I$ is not invertible.

(2) Some α have no eigenvalues. For example, suppose $K = \mathbb{R}$, and $\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}$ defines a linear transformation $\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ (with respect to the standard basis). Now $\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = \lambda\begin{pmatrix}x\\y\end{pmatrix}$ means that $y = \lambda x$ and $-x = \lambda y$, so either $x = y = 0$, or $-yx = \lambda^2 yx$, so $\lambda^2 = -1$, which has no solutions in $\mathbb{R}$.

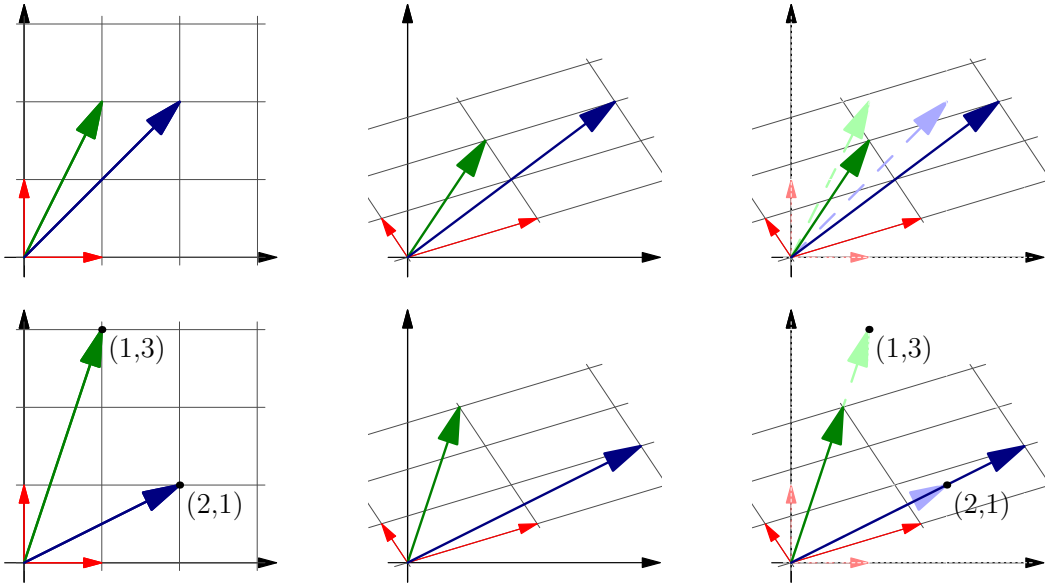

FIGURE 1. Eigenvalues and Eigenvectors

Theorem 4.5. *Let $\alpha : V \to V$ be a linear transformation, where $\dim(V) = n$. Then the eigenvalues of α are the roots of the* characteristic equation $\det(tI - A) = 0$, *where A is any matrix representing α with respect to some basis. The equation is independent of the choice of basis, so can be written $\det(tI - \alpha) = 0$. It is a polynomial in t of degree n.*

Proof. First, we verify that the characteristic equation is independent of the choice of basis. So suppose that A represents α with respect to one basis, and B represents α with respect to another. Then there is a non-singular matrix P such that $B = P^{-1}AP$ (Theorem 3.20). Now

$$\begin{aligned}
\det(tI - B) &= \det(tI - P^{-1}AP) \\
&= \det(P^{-1}(tI)P - P^{-1}AP) \\
&= \det(P^{-1}(tI - A)P) \\
&= \det(P^{-1})\det(tI - A)\det(P) \\
&= \det(tI - A),
\end{aligned}$$

the last step using that $\det(P^{-1})\det(P) = \det(P^{-1}P) = \det(I) = 1$.

Now, λ is an eigenvalue of $\alpha \Leftrightarrow \alpha - \lambda I$ is not invertible

$$\Leftrightarrow \det(A - \lambda I) = 0 \Leftrightarrow \det(\lambda I - A) = 0.$$

This clearly is a polynomial of degree n in t, with leading term t^n. □

Remark 4.6. There is some confusion as to whether $\det(tI - A)$ or $\det(A - tI)$ is the characteristic equation. I go for $\det(tI - A)$, so the coefficient of t^n is 1. But the two have the same roots. Indeed, $\det(tI - A) = (-1)^n \det(A - tI)$.

We often write $\chi(t)$ for the characteristic equation (also called the *characteristic polynomial*). By Theorem 4.5, this can be regarded equally as the characteristic equation of A, or of α (where A represents α with respect to some basis).

Definition 4.7. (1) The linear transformation $\alpha : V \to V$ is *diagonalisable* if there is a basis of V such that α is represented by a diagonal matrix with respect to this basis.
(2) We denote by $\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ the diagonal matrix whose diagonal entry (in the (i,i) position, for each i) is λ_i.

Remark 4.8. If α is represented by A with respect to some basis, then α is diagonalisable if and only if there is a non-singular matrix P such that $P^{-1}AP$ is diagonal, that is, A is similar to a diagonal matrix.

Theorem 4.9. *Let $\alpha : V \to V$ be linear. Then α is diagonalisable if and only if there is a basis of V consisting of eigenvectors of α.*

Proof. $\Leftarrow$ If $v_1, \ldots, v_n$ is a basis of V consisting of eigenvectors of α, with $\alpha(v_j) = \lambda_j v_j$ for each j, then the matrix for α with respect to the basis $v_1, \ldots, v_n$ is $\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$.

$\Rightarrow$ Conversely, suppose α has matrix $\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ with respect to basis $v_1, \ldots, v_n$. Then $\alpha v_i = \lambda_i v_i$ for each i, so the v_i are eigenvectors. □

Example 4.10. Let $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. First, consider it as giving a linear transformation $\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ (with respect to the standard basis). The characteristic equation is $\det \begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix} = \lambda^2 + 1$, which has no real roots. So A has no eigenvalues in $\mathbb{R}$, so no eigenvectors, so is not diagonalisable. Next, work over *complex* scalars, and consider A as giving a linear transformation $\alpha : \mathbb{C}^2 \to \mathbb{C}^2$, again with respect to the standard basis. Now $\lambda^2+1 = (\lambda+i)(\lambda-i)$, with roots $i, -i$. We find the eigenvectors.

For $\lambda = i$, the equation $(iI - A)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is

$$\begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which gives the single equation $ix - y = 0$, with solution set $\{a(1,i) : a \in \mathbb{C}\}$, so $(1,i)$ is an eigenvector.

For $\lambda = -i$, we solve $(-iI - A)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, that is

$$\begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This gives the equation $-ix - y = 0$, with solutions $\{a(i,1) : a \in \mathbb{C}\}$, so $(i,1)$ is an eigenvector.

The vectors $(1,i)$ and $(i,1)$ are linearly independent, so form a basis of eigenvectors of α, so by Theorem 4.9, α is dagonalisable. Pur $P = \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$ (so the *columns*) of P are

the eigenvectors, treated as column vectors). Then $P^{-1} = \frac{1}{2}\begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix}$, and a calculation shows that $P^{-1}AP = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$.

4.1. **More on eigenvalues, eigenvectors.** In the discussion which follows, we shall work over $\mathbb{C}$. The key fact is that *every* polynomial $b_0 + b_1 t + \ldots + b_m t^m$ (with $b_i \in \mathbb{C}$) factorises completely into linear factors over $\mathbb{C}$. So we shall for a while assume $K = \mathbb{C}$. This avoids problems like that in the last example (working over $\mathbb{R}$).

Definition 4.11. Let $\alpha : V \to V$ be linear, and λ be an eigenvalue of α (so a root of the characteristic polynomial $\chi(t)$). Then the *algebraic multiplicity* $m_a(\lambda)$ is defined to be its multiplicity as a root of $\chi(t)$. Its *geometric multiplicity* $m_g(\lambda)$ is $\dim(\ker(tI-\alpha)) = n(tI-\alpha)$.

Given an $n \times n$ matrix A, we can regard A as giving a linear transformation $\mathbb{C}^n \to \mathbb{C}^n$, so can talk of the algebraic and geometric multiplicities of eigenvalues of A (in fact, these do not depend on the choice of basis).

Proposition 4.12. *Let $\alpha : V \to V$ be linear, where V is an n-dimensional vector space.*

(1) *For each eigenvalue λ of α, we have*

$$1 \leqslant m_g(\lambda) \leqslant m_a(\lambda) \leqslant n$$

(that is, 'geometric multiplicity $\leqslant$ algebraic multplicity').

(2) *If $\lambda_1, \ldots, \lambda_r$ are the* distinct *eigenvalues of α then*

$$m_a(\lambda_1) + m_a(\lambda_2) + \ldots + m_a(\lambda_r) = n.$$

(3) *If $\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues, and v_i is an eigenvector of λ_i for each i, then $\{v_1, \ldots, v_r\}$ is linearly independent.*

Proof. (1) By the definition of *eigenvalue* $m_g(\lambda) \geqslant 1$; and as $\chi(t)$ has degree n, $m_a(\lambda) \leqslant n$. So we must show that $m_g(\lambda) \leqslant m_a(\lambda)$.

Suppose that $m_g(\lambda) = p$, take a basis $v_1, \ldots, v_p$ for $\mathrm{Ker}(\lambda I - \alpha)$, and extend it to a basis $v_1, \ldots, v_n$ for V. With respect to the basis $v_1, \ldots, v_n$, α has matrix

$$\begin{pmatrix} \mathrm{Diag}(\lambda, \ldots, \lambda) & S \\ Z & T \end{pmatrix},$$

where $\mathrm{Diag}(\lambda, \ldots, \lambda)$ is $p \times p$, S is some $p \times (n-p)$ matrix, Z is the all-zero $(n-p) \times p$ matrix, and T is some $(n-p) \times (n-p)$ matrix. Now $\det(tI-\alpha) = (t-\lambda)^p \det(tI-T)$, so $m_a(\lambda) \geqslant p$.

(2) We have $\det(tI-\alpha) = (t-\lambda_1)^{m_a(\lambda_1)}(t-\lambda_2)^{m_a(\lambda_2)} \ldots (t-\lambda_r)^{m_a(\lambda_r)}$, and $\det(tI-\alpha)$ has degree n.

(3) We show by induction on m that if $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues with eigenvectors $v_1, \ldots, v_m$, then $v_1, \ldots, v_m$ are linearly independent.

If $m = 1$, this is trivial, as eigenvectors are non-zero. So suppose

$$c_1 v_1 + \ldots + c_m v_m = 0. \tag{3}$$

Multiplying (3) by λ_m,

$$c_1 \lambda_m v_1 + \ldots + c_m \lambda_m v_m = 0.$$

Also, by applying α to (3), we have

$$c_1\lambda_1 v_1 + \ldots + c_m\lambda_m v_m = 0.$$

Hence, subtracting, $c_1(\lambda_1 - \lambda_m)v_1 + \ldots + c_{m-1}(\lambda_1 - \lambda_m)v_m = 0$. As $v_1, \ldots, v_{m-1}$ are linearly independent (by induction), this gives

$$c_1(\lambda_1 - \lambda_m) = \ldots = c_{m-1}(\lambda_{m-1} - \lambda_m) = 0.$$

As the λ_j are distinct, this forces $c_1 = \ldots = c_{m-1} = 0$, and from (3) we then also get $c_m = 0$.

□

Definition 4.13. If $V_1, \ldots, V_r$ are subspaces of the vector space V, then V is a *direct sum* $V = V_1 \oplus \ldots \oplus V_r$ if each $v \in V$ is *uniquely* expressible as $v = v_1 + \ldots + v_r$ (for $v_i \in V_i$).

Remark 4.14. The last definition extends Definition 1.16 to the situation where there are more than two subspaces.

Let $V_1, \ldots, V_r$ be subspaces of V, and suppose that whenever $v_1 \in V_1, \ldots, v_r \in V_r$ and $v_1 + \ldots + v_r = 0$, we have $v_1 = \ldots = v_r = 0$. Then $V_1 \oplus \ldots \oplus V_r$ is a direct sum, and its dimension is $\dim(V_1) + \ldots + \dim(V_r)$. We omit the proof, but it follows easily by induction from Corollary 2.16.

Proposition 4.15. *The following are equivalent.*

(1) α *is diagonalisable;*
(2) $m_g(\lambda) = m_a(\lambda)$ *for each eigenvalue* λ*;*
(3) $V = \mathrm{Ker}(\lambda_1 I - \alpha) \oplus \ldots \oplus \mathrm{Ker}(\lambda_r I - \alpha)$.

Proof. For each $i = 1, \ldots, r$, put $V_i := \mathrm{Ker}(\lambda_i I - \alpha)$.

(i)⇒ (ii) We use Theorem 4.9 and its proof. So suppose α is diagonalisable. This means that with respect to a basis of eigenvectors, say $u_1, \ldots, u_n$, α has matrix

$$A = \mathrm{Diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_r, \ldots, \lambda_r).$$

For each j there are $m_a(\lambda_j)$ occurrences of λ_j in this matrix (take its characteristic polynomial). Hence, there are $m_a(\lambda_j)$ basis vectors from $u_1, \ldots, u_n$ which are eigenvectors of λ_j (cf. Proof of Theorem 4.9). Hence, $m_g(\lambda_j) \geqslant m_a(\lambda_j)$. Thus, by (i), $m_g(\lambda_j) = m_a(\lambda_j)$ for each j.

(ii) ⇒ (iii) Suppose (ii), i.e. that $m_g(\lambda) = m_a(\lambda)$ for each λ. First, using (iii) and Remark 4.14, observe that the sum $V_1 + \ldots + V_r$ is direct. Thus, it suffices to show that it equals V, i.e. that the dimensions sum to $n = \dim(V)$. Since $\dim(V_i) = m_g(\lambda_i)$, assuming (b) we have

$$\sum_{i=1}^{r} \dim(V_i) = \sum_{i=1}^{r} m_g(\lambda_i) = \sum_{i=1}^{r} m_a(\lambda_i) = n,$$

as required.

(iii) ⇒ (i) Suppose $V = V_1 \oplus \ldots \oplus V_r$. Then, by Remark 4.14, $\dim(V) = \dim(V_1) + \ldots + \dim(V_r)$. Let B_i be a basis for V_i for each i. Observe that B_i consists of eigenvectors with eigenvalue λ_i. Then, using the definition of direct sum, $B := B_1 \cup \ldots \cup B_r$ is linearly independent, and as it has size $\dim(V)$, B is a basis of V. Thus, V has a basis of eigenvectors of α, so by Theorem 4.9, α is diagonalisable.

□

Example 4.16. Let $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. In both cases, $m_a(3) = 1$ and $m_a(2) = 2$, since each have characteristic equation $(\lambda - 2)^2(\lambda - 3)$. Since $1 \leqslant m_g(3) \leqslant m_a(3)$, we must in each case have $m_g(3) = 1$. Thus, in each case, the matrix is diagonalisable if and only if $m_g(2) = 2$.

For A, we find $2I - A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, and the row reduced for of this has one non-zero row. That is, $r(2I - A) = 1$ and $n(2I - A) = 2$ (regarding these as linear transformations with respect to standard basis), so for A we find $m_g(2) = 2$, and so A is diagonalisable.

For B, on the other hand, $2I - B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ which has rank 2 and nullity 1, so $m_g(2) = 1$ for B, so B is not diagonalisable.

To diagonalise A, find a basis of eigenvectors. For $\lambda = 2$, $v_1 = (1, 0, 0)^T$ and $v_2 = (0, 1, 0)^T$ are eigenvectors. For $\lambda = 3$, $v_3 = (0, 1, 1)^T$ is an eigenvector. Let P have v_1, v_2, v_3 as columns, so $P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. Then $P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$, and $P^{-1}AP = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$, which is diagonal with the eigenvalues 2,2,3 listed in the order corresponding to eigenvectors v_1, v_2, v_3.

4.2. **Polynomials.** We shall continue to work with vector spaces over $\mathbb{C}$, to ensure that any polynomial factorises into linear factors. So V is an n-dimensional vector space over $\mathbb{C}$.

We have $\alpha : V \to V$. If A represents α with respect to the basis $v_1, \ldots, v_n$, then by Theorem 3.17, A^2 represents the composition α^2, A^3 represents α^3, and so on. Also, if A represents α, and B represents β, then $A+B$ represents $\alpha+\beta$, where $\alpha+\beta$ is defined so that $(\alpha+\beta)(v) = \alpha(v)+\beta(v)$. More generally, given the polynomial $p(t) = b_0+b_1t+\ldots+b_mt^m$, we can form the linear transformation $p(\alpha) = b_0 + b_1\alpha + \ldots + b_m\alpha^m$, which has matrix $p(A) := b_0I + b_1A + \ldots + b_mA^m$ with respect to the basis $v_1, \ldots, v_n$.

Lemma 4.17. (1) *Let $\alpha : V \to V$ be a non-zero linear transformation. Then there is a polynomial $p(t)$ of degree at most n^2 with $p(\alpha) = 0$ (that is, $p(\alpha)$ is the zero linear transformation, so it maps each vector to the zero vector).*

(2) *All polynomials p such that $p(\alpha) = 0$ are multiples of a unique polynomial of minimal degree which is* monic, *i.e. has leading coefficient 1.*

Definition 4.18. The monic polynomial of minimal degree described in (ii) is called the *minimal polynomial* of α, and is usually denoted $\mu(t)$.

Proof of Lemma 4.17. (1) Let $\mathrm{Lin}(V, V)$ be the set of all linear transformations $V \to V$. Then $\mathrm{Lin}(V, V)$ is a vector space over $\mathbb{C}$ (compare Example 1.1(e).) If we work over a fixed basis of V, then $\mathrm{Lin}(V, V)$ is essentially the same as $M_{n,n}(\mathbb{C})$ (just replace each linear transformation by the matrix representing it in our given basis.) Now the dimension of $M_{n,n}(\mathbb{C})$ is n^2, since we may form a basis $\{E_{i,j} : 1 \leqslant i, j \leqslant n\}$,

where $E_{i,j}$ has a 1 in the (i,j)-entry and zeros elsewhere. Hence $\dim(\mathrm{Lin}(V,V)) = n^2$. It follows that the set $I, \alpha, \alpha^2, \ldots, \alpha^{n^2}$ is linearly *dependent* (as it has size n^2+1). Hence, there are constants $c_0, c_1, \ldots, c_{n^2} \in \mathbb{C}$ such that $c_0 + c_1\alpha + \ldots c_{n^2}\alpha^{n^2} = 0$. Thus, if $p(t) = c_0 + c_1 t + \ldots + c_{n^2}t^{n^2}$, then $p(\alpha) = 0$.

(2) Let $\mu(t)$ be of minimal degree and of form $\mu(t) = t^k + d_{k-1}t^{k-1} + \ldots + d_1 t + d_0$ – we can put it in this (monic) form by multiplying by a non-zero constant.

If $p(t)$ is another polynomial with $p(\alpha) = 0$, then by division for polynomials, we have $p(t) = q(t)\mu(t) + r(t)$ where $\deg(r(t)) < \deg(\mu(t))$ or $r(t) = 0$. (If you don't know this fact about division of polynomials, don't worry about it! The idea is that $q(t)$ is the quotient, $r(t)$ is the remainder.) Now $r(\alpha) = p(\alpha) - q(\alpha)\mu(\alpha) = 0$, so by the minimality of the degree of $\mu(t)$, $r(t) = 0$, so $p(t) = q(t)\mu(t)$ is a multiple of $\mu(t)$.

□

We now state the main theorem about the minimal polynomial, the Cayley-Hamilton Theorem. The proof of the Cayley-Hamilton Theorem is given in these notes but not in the lectures, and for the exam, you will be expected to know the statement but not the proof of Cayley-Hamilton, and you will be expected to know the proof of Corollary 4.20.

Theorem 4.19 (Cayley-Hamilton Theorem)**.** *The characteristic polynomial $\chi(t)$ of α satisfies $\chi(\alpha) = 0$.*

Corollary 4.20. *If $\mu(t)$ and $\chi(t)$ are respectively the minimal and characteristic polynomials of t, then $\mu(t)$ divides $\chi(t)$.*

Proof. By the Cayley-Hamilton Theorem, $\chi(\alpha) = 0$, so by Lemma 2, $\chi(t)$ is a multiple of $\mu(t)$. □

Before proving the Cayley-Hamilton Theorem, recall that if A is an $n \times n$ matrix then $\mathrm{adj}(A)$ is the transposed 'matrix of cofactors'. That is, $\mathrm{adj}(A) = B^T$, where $B = (b_{ij})$ and b_{ij} is + or - the determinant of the $(n-1) \times (n-1)$-matrix obtained from A by removing the ith row and jth column. For example, if $A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ then $B = \begin{pmatrix} 1 & 1 & -1 \\ -2 & 0 & 2 \\ 1 & -1 & 1 \end{pmatrix}$, so $\mathrm{adj}(A) = B^T = \begin{pmatrix} 1 & -2 & 1 \\ 1 & 0 & -1 \\ -1 & 2 & 1 \end{pmatrix}$.

Now $A \cdot \mathrm{adj}(A) = \det(A) \cdot I_n$ (a fact about matrices; you may recall that $A^{-1} = \mathrm{adj}(A)/\det(A)$ if A is invertible).

Proof of Theorem 4.19. Let A be the matrix of α with respect to some basis. By the last remark,

$$(tI - \alpha)\,\mathrm{adj}(tI - \alpha) = \chi(t)I.$$

Put $B := \mathrm{adj}(tI - \alpha)$. Then B is a matrix such that each entry is a polynomial in t. So $B = B_0 + tB_1 + t^2B_2 + \ldots$, where each B_j is a matrix over $\mathbb{C}$. (For example, we could write

$$\begin{pmatrix} t^2+t+1 & t-1 \\ t^2 & t \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} t + \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} t^2.$$

Thus,

$$(tI - A)(B_0 + tB_1 + t^2B_2 + \ldots) = \chi(t)I,$$

for any t. Now put $t = A$. Then the left hand side is zero, so $\chi(A)I = 0$. This forces $\chi(A) = 0$, so $\chi(\alpha) = 0$. □

Corollary 4.21. *Suppose $\chi(t) = \prod_{i=1}^{m}(t-\lambda_i)^{n_i}$ is the characteristic polynomial of α. Then $\mu(t)$ has the form $\prod_{i=1}^{m}(t-\lambda_i)^{p_i}$ where $1 \leqslant p_i \leqslant n_i$ for each i.*

Proof. Since $\mu(t)$ divides $\chi(t)$, we just need to check that each λ_i is a root of $\mu(t)$, i.e. that $1 \leqslant p_i$. So let λ be an eigenvalue of α, with eigenvector v. Then $\alpha v = \lambda v$, so $\alpha^2 v = \lambda^2 v$, $\alpha^3(v) = \lambda^3(v)$, etc. Thus, $\mu(\alpha)v = \mu(\lambda)v$. Since $\mu(\alpha) = 0$, and $v \neq 0$, this forces $\mu(\lambda) = 0$, so λ is a root of $\mu(t)$. □

Finally, we obtain our last criterion for diagonalisability of a linear transformation.

Proposition 4.22. *Let $\alpha : V \to V$ be a linear transformation, where V is a vector space over $\mathbb{C}$. Then α is diagonalisable if and only if $\mu(t)$ is a product of distinct linear factors.*

Proof. $\Rightarrow$ Suppose that α is diagonalisable, represented by the matrix

$$A = \mathrm{Diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_r, \ldots, \lambda_r).$$

Then you can check that

$$(\lambda_1 I - A)(\lambda_2 I - A)\ldots(\lambda_r I - A) = 0,$$

so $\mu(t) = (t-\lambda_1)(t-\lambda_2)\ldots(t-\lambda_r)$ (a product of *distinct* linear factors).

$\Leftarrow$ Suppose $\mu(t) = (t-\lambda_1)(t-\lambda_2)\ldots(t-\lambda_r)$. Let $\mu_i(t) := \mu(t)/(t-\lambda_i)$. Now, expanding by partial fractions,

$$\frac{1}{\mu(t)} = \frac{1}{(t-\lambda_1)(t-\lambda_2)\ldots(t-\lambda_r)} = \frac{c_1}{t-\lambda_1} + \frac{c_2}{t-\lambda_2} + \ldots + \frac{c_r}{t-\lambda_r}.$$

Hence, multiplying by $\mu(t)$,

$$1 = \frac{\mu(t)}{\mu(t)} = c_1\mu_1(t) + c_2\mu_2(t) + \ldots + c_r\mu_r(t).$$

Let $v \in V$. Then, from the last equation,

$$v = c_1\mu_1(\alpha)v + c_2\mu_2(\alpha)v + \ldots + c_r\mu_r(\alpha)v. \tag{4}$$

It follows that for each i, $c_i\mu_i(\alpha)v \in \mathrm{Ker}(\lambda_i I - \alpha)$. Indeed,

$$(\lambda_i I - \alpha)c_i\mu_i(\alpha)v = (\lambda_i I - \alpha)c_i\frac{\mu(\alpha)}{\lambda_i I - \alpha}v = c_i\mu(\alpha)v = 0.$$

Thus, each $c_i\mu_i(\alpha)$ is an eigenvalue of α, so by (4), the eigenvectors of α span V, and so α is diagonalisable by Theorem 4.9.

□

Example 4.23. Return to matrices A and B from the Example 4.16, which both have the characteristic polynomial $(\lambda-2)^2(\lambda-3)$. Since A is diagonalisable its minimal polynomial is $(\lambda-2)(\lambda-3)$, and since B is not diagonalisable its minimal polynomial coincides with the characteristic one. We also can conclude that A and B are not similar.

5. Inner Product Spaces

We now return to working with vector spaces over $\mathbb{R}$ (so from now on, $K = \mathbb{R}$), but the theory below could be developed for vector spaces over $\mathbb{C}$.

Definition 5.1. Let V be a vector space over $\mathbb{R}$, and suppose that for each pair of vectors $u, v \in V$ there is defined a real number written $\langle u, v\rangle$ (sometimes denoted (u, v), or $u \cdot v$). This is called a (real) *inner product* on V if it satisfies the following, for $a, b \in \mathbb{R}$, and $u_1, u_2, u, v \in V$.

(1) (Linearity) $\langle au_1 + bu_2, v\rangle = a\langle u_1, v\rangle + b\langle u_2, v\rangle$.
(2) (Symmetry) $\langle u, v\rangle = \langle v, u\rangle$.
(3) (Positive Definiteness) $\langle u, u\rangle \geqslant 0$, and $\langle u, u\rangle = 0 \Leftrightarrow u = 0$.

The vector space V with $\langle -, -\rangle$ satisfying 1–3 is called an *inner product space*.

Remark 5.2. We can deduce from these axioms the following properties.

(1) $\langle u, av_1 + bv_2\rangle = \langle av_1 + bv_2, u\rangle = a\langle v_1, u\rangle + b\langle v_2, u\rangle = a\langle u, v_1\rangle + b\langle u, v_2\rangle$ – that is, the inner product is linear also in the second variable.
(2) For all $v \in V$, $\langle 0, v\rangle = \langle 0v, v\rangle = 0\langle v, v\rangle = 0$, and also $\langle v, 0\rangle = \langle 0, v\rangle = 0$.
(3) We have the following generalisation of 1

$$\langle a_1u_1 + a_2u_2 + \ldots + a_nu_n, b_1v_1 + b_2v_2 + \ldots + b_mv_m\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_j\langle u_i, v_j\rangle.$$

This can be proved by induction. First use induction on n to get

$$\langle a_1u_1 + \ldots + a_nu_n, b_1v_1\rangle = a_1b_1\langle u_1, v_1\rangle + \ldots + a_nb_1\langle u_n, v_1\rangle,$$

and then use induction on m.

Example 5.3.
(1) Take the usual scalar (or 'dot') product on $V = \mathbb{R}^n$. Here, if $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, then $\langle x, y\rangle = \sum_{i=1}^{n} x_iy_i$. This is 'Euclidean n-space'.
(2) Let V be the vector space of all continuous functions on the closed interval $[a, b]$ (or restrict to the subspace of polynomial functions on this interval). Define

$$\langle f, g\rangle = \int_a^b f(t)g(t)dt.$$

This is an inner product, by the basic rules for integration.
(3) As a slight adjustment of (2), let $w(t)$ be a continuous and strictly positive function on $[a, b]$ (a 'weight function'), and let

$$\langle f, g\rangle = \int_a^b f(t)g(t)w(t)dt.$$

Theorem 5.4 (Cauchy-Schwarz Inequality)**.** *If V is an inner product space, and $u, v \in V$, then*

$$\langle u, v\rangle^2 \leqslant \langle u, u\rangle\langle v, v\rangle.$$

Proof. We can assume $\langle u, u\rangle \neq 0$, as otherwise $u = 0$ and both sides vanish. Let $t \in \mathbb{R}$. Then $\langle tu - v, tu - v\rangle \geqslant 0$, and (by linearity) equals

$$t^2\langle u, u\rangle - 2t\langle u, v\rangle + \langle v, v\rangle.$$

Put $t = \frac{\langle u,v\rangle}{\langle u,u\rangle}$. Then we obtain

$$\frac{\langle u,v\rangle^2}{\langle u,u\rangle} - \frac{2\langle u,v\rangle^2}{\langle u,u\rangle} + \langle v,v\rangle \geqslant 0.$$

Multiplying out, we get $\langle v,v\rangle\langle u,u\rangle \geqslant \langle u,v\rangle^2$. □

Example 5.5. In $\mathbb{R}^n$, with the usual scalar product, the Cauchy-Schwarz Inequality says

$$\sum_{i=1}^{n} x_i y_i \leqslant \left(\sum x_i^2\right)^{\frac{1}{2}} \left(\sum y_i^2\right)^{\frac{1}{2}},$$

or $x \cdot y \leqslant |x|\,|y|$ for vectors in $\mathbb{R}^n$. This is true in $\mathbb{R}^3$, as $x \cdot y = |x|\,|y|\cos\theta$, where θ is the angle between the vectors x and y.

Definition 5.6. If $v \in V$ then the *norm* of v, also called the *length* of v, is defined to be $\sqrt{\langle v,v\rangle}$, and denoted $\|v\|$.

Theorem 5.7. *Let V be an inner product space. Then the norm on V satisfies*

(1) $\|v\| \geqslant 0$, *and* $\|v\| = 0$ *if and only if* $v = 0$,
(2) $\|kv\| = |k| \cdot \|v\|$ *for any* $k \in \mathbb{R}$,
(3) $\|u+v\| \leqslant \|u\| + \|v\|$ *(the triangle inequality).*

Proof. (1) This is just an axiom for inner products.
(2) $\langle kv, kv\rangle = k^2\langle v,v\rangle$. Now take square roots, noting both sides are positive.
(3) We have

$$\begin{aligned}\langle u+v, u+v\rangle &= \langle u,u\rangle + 2\langle u,v\rangle + \langle v,v\rangle \\ &\leqslant \langle u,u\rangle + 2\langle u,u\rangle^{\frac{1}{2}}\langle v,v\rangle^{\frac{1}{2}} + \langle v,v\rangle \qquad \text{by Theorem 5.4} \\ &= \|u\|^2 + 2\,\|u\| \cdot \|v\| + \|v\|^2 = (\|u\| + \|v\|)^2.\end{aligned}$$

Now take square roots.

□

Definition 5.8. We say that vectors $u, v \in V$ (a real inner product space) are *orthogonal*, written $u \perp v$, if $\langle u,v\rangle = 0$.

Theorem 5.9 (Pythagoras). *If $u \perp v$ then $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.*

Proof. $\langle u+v, u+v\rangle = \langle u,u\rangle + 2\langle u,v\rangle + \langle v,v\rangle = \|u\|^2 + \|v\|^2$. □

Definition 5.10. A set $\{e_1, \ldots, e_m\}$ in an inner product space V is said to be *orthonormal* if $\|e_i\| = 1$ for each i and $\langle e_i, e_j\rangle = 0$ for all distinct i, j. An *orthonormal basis* for a subspace W of V is a basis of W which is an orthonormal set.

Proposition 5.11. *Suppose that $e_1, \ldots, e_m$ is an orthonormal set.*

(1) *If $v = \sum_{i=1}^m a_i e_i$ and $w = \sum_{i=1}^m b_i e_i$, then $\langle v,w\rangle = \sum_{i=1}^m a_i b_i$.*
(2) $\|\sum_{i=1}^m a_i e_i\|^2 = (\sum_{i=1}^m a_i^2)$, *so $e_1, \ldots, e_m$ is linearly independent;*
(3) *If $v = a_1 e_1 + \ldots + a_m e_m$, then $a_i = \langle v, e_i\rangle$ for each i.*

Proof. (1) Apply Remark 3 after Definition 5.1.

(2)
$$\langle \sum_{i=1}^{m} a_i e_i, \sum_{j=1}^{m} a_j e_j \rangle = \sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \langle e_i, e_j \rangle = \sum_{i=1}^{m} a_i^2.$$
In particular, is $\sum_{i=1}^{m} a_i e_i = 0$, then $\sum_{i=1}^{m} a_i^2 = 0$, which implies that $a_i = 0$ for all i (as squares are non-negative).

(3) Observe that $\langle v, e_i \rangle = \langle a_1 e_1 + \ldots + a_m e_m, e_i \rangle = a_i$.

□

Example 5.12. (1) In $\mathbb{R}^n$, the standard basis is an orthonormal basis.

(2) There are many other orthonormal bases in $\mathbb{R}^n$, for example, in $\mathbb{R}^3$
$$\begin{aligned} u_1 &= \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \\ u_2 &= \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0\right), \\ u_3 &= \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}\right) \end{aligned}$$
is an orthonormal basis.

(3) Here is a set of orthonormal functions on the interval $[0, 2\pi]$, with respect to the inner product $\langle f, g \rangle = \int_0^{2\pi} f(t)g(t)dt$ (see Example 5.3):
$$\frac{1}{\sqrt{2\pi}}, \frac{\cos t}{\sqrt{\pi}}, \frac{\sin t}{\sqrt{\pi}}.$$

5.1. **Gram–Schmidt Orthogonalisation Process.** This is a procedure for finding an orthonormal basis. Let $v_1, \ldots, v_m$ be a linearly independent set of vectors in the real inner product space V. We construct an orthonormal set $e_1, \ldots, e_m$ such that for each $j = 1, \ldots, m$,
$$\operatorname{span}(e_1, \ldots, e_j) = \operatorname{span}(v_1, \ldots, v_j).$$

The construction is as follows. First, put $e_1 = v_1 / \|v_1\|$ (that is, we *normalise* v_1, to ensure length 1).

To find each subsequent e_j, we first subtract off the components of v_j in the directions of the previous e_k, and then normalise.

More formally, put $w_2 = v_2 - \langle v_2, e_1 \rangle e_1$. Now
$$\langle w_2, e_1 \rangle = \langle v_2, e_1 \rangle - \langle v_2, e_1 \rangle \langle e_1, e_1 \rangle = 0,$$
and $\operatorname{span}(w_2, e_1) = \operatorname{span}(v_2, e_1) = \operatorname{span}(v_2, v_1)$, so $w_2 \neq 0$. Now, put $e_2 = w_2 / \|w_2\|$ (i.e., normalise w_2).

In general, let
$$w_j = v_j - \sum_{k=1}^{j-1} \langle v_j, e_k \rangle e_k.$$
Then, for $l < j$,
$$\langle w_j, e_l \rangle = \langle v_j, e_l \rangle - \sum_{k=1}^{j-1} \langle v_j, e_k \rangle \langle e_k, e_l \rangle = 0, \text{ and}$$

$$\operatorname{span}(w_j, e_1, \dots, e_{j-1}) = \operatorname{span}(v_j, e_1, \dots, e_{j-1}) = \operatorname{span}(v_1, \dots, v_j),$$

so $w_j \neq 0$. Now put $e_j = w_j / \|w_j\|$.

Example 5.13. The plane $x + y + z = 0$ in $\mathbb{R}^3$ has basis $v_1 = (1, 0, -1)$, $v_2 = (0, 1, -1)$. We find an orthonormal basis.

First, $e_1 = (1, 0, -1)/\|(1, 0, -1)\| = (\frac{1}{\sqrt{2}}, 0, \frac{-1}{\sqrt{2}})$.

Next,

$$w_2 = v_2 - \langle v_2, e_1\rangle e_1 = (0, 1, -1) - \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}, 0, \frac{-1}{\sqrt{2}}\right) = \left(\frac{-1}{2}, 1, \frac{-1}{2}\right).$$

Hence,

$$e_2 = w_2/\|w_2\| = \frac{\sqrt{2}}{\sqrt{3}}\left(\frac{-1}{2}, 1, \frac{-1}{2}\right) = \left(\frac{-1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{-1}{\sqrt{6}}\right).$$

We'll now extend e_1, e_2 to an orthonormal basis of $\mathbb{R}^3$. Put $v_3 = (0, 0, 1)$, so v_1, v_2, v_3 is a basis of $\mathbb{R}^3$. Then

$$\begin{aligned} w_3 &= v_3 - \langle v_3, e_1\rangle e_1 - \langle v_3, e_2\rangle e_2 \\ &= (0, 0, 1) - \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}, 0, \frac{-1}{\sqrt{2}}\right) + \frac{1}{\sqrt{6}}\left(\frac{-1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{-1}{\sqrt{6}}\right) \\ &= (0, 0, 1) + \left(\frac{1}{2}, 0, \frac{-1}{2}\right) + \left(\frac{-1}{6}, \frac{2}{6}, \frac{-1}{6}\right) \\ &= \left(\frac{2}{6}, \frac{2}{6}, \frac{2}{6}\right) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right). \end{aligned}$$

Hence, $e_3 = \sqrt{3}w_3 = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$. Geometrically, e_3 is a unit vector perpendicular to the plane $x + y + z = 0$.

5.2. **Orthogonal Complements.** Recall from Definition 1.19: If $W_1 \subset V$ are vector spaces, then W_2 is a *complement* of W_1 if and only if $V = W_1 \oplus W_2$ (direct sum), or equivalently, $W_1 + W_2 = V$ and $W_1 \cap W_2 = \{0\}$, or equivalently again, each vector $v \in V$ is uniquely expressible as $v = w_1 + w_2$, for $w_i \in W_i$.

In general, complements to a subspace are not unique, see Remark 1.20. However, given an inner product we have a particularly nice complement.

Definition 5.14. Let V be an inner product space, and W be a subspace of V. The *orthogonal complement* $W^\perp$ of W is defined by

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

Theorem 5.15. *Let W be a subspace of a finite-dimensional inner product space V. Then*

(1) *$W^\perp$ is a subspace of V;*
(2) *$V = W \oplus W^\perp$, so $\dim V = \dim W + \dim(W^\perp)$;*
(3) *$(W^\perp)^\perp = W$.*

Proof. (1) Clearly, $0 \in W^\perp$. If $v_1, v_2 \in W^\perp$ and $a_1, a_2 \in \mathbb{R}$, and $w \in W$, then

$$\langle a_1 v_1 + a_2 v_2, w\rangle = a_1\langle v_1, w\rangle + a_2\langle v_2, w\rangle = 0 + 0 = 0.$$

Hence $a_1 v_1 + a_2 v_2 \in W^\perp$, so $W^\perp$ is a subspace.

(2) Take a basis $w_1, \dots, w_m$ for W, and extend it to a basis $w_1, \dots, w_m, v_{m+1}, \dots, v_n$ of V. Now apply Gram–Schmidt: We get an orthonormal basis $e_1, \dots, e_n$ of V, such that $e_1, \dots, e_m$ is an orthonormal basis of W. It suffices to prove the following claim, which yields that $V = W + W^\perp$ and $W \cap W^\perp = \{0\}$.

Lemma 5.16. $W^\perp = \operatorname{span}(e_{m+1}, \dots, e_n)$.

Proof of the Lemma 5.16. First, $W^\perp \subseteq \operatorname{span}(e_{m+1}, \dots, e_n)$. For let $v \in W^\perp$, and suppose $v = \sum_{i=1}^n a_i e_i$ (as the e_i are a basis of V). Then $a_j = \langle v, e_j \rangle = 0$ for $j \leqslant m$, since $e_j \in W$ and $v \in W^\perp$. Hence, $v = a_{m+1}e_{m+1} + \dots + a_n e_n \in \operatorname{span}(e_{m+1}, \dots, e_n)$.

Also $\operatorname{span}(e_{m+1}, \dots, e_n) \subseteq W^\perp$. For suppose that $v \in \operatorname{span}(e_{m+1}, \dots, e_n)$ and $w \in W$, say $v = \sum_{i=m+1}^n a_i e_i$ and $w \in \sum_{i=1}^m b_i e_i$. Then

$$\langle v, w \rangle = \langle a_{m+1}e_{m+1} + \dots + a_n e_n, b_1 e_1 + \dots + b_m e_m \rangle = 0.$$

□

(3) First, we show $W \subseteq (W^\perp)^\perp$. To see this, let $w \in W$ and $v \in W^\perp$. Then $\langle w, v \rangle = 0$, so $w \in (W^\perp)^\perp$.

Thus, it remains to show that $\dim(W) = \dim(W^\perp)$. But

$$\dim(W) + \dim(W^\perp) = n = \dim(W^\perp)^\perp + \dim(W^\perp)$$

(as $V = W^\perp + (W^\perp)^\perp$ by 2), so $\dim(W) = \dim(W^\perp)^\perp$. Hence $W = (W^\perp)^\perp$.

□

As a brief example, recall Example 5.13. There, $V = \mathbb{R}^3$, and W is the plane $x + y + z = 0$. Now

$$W^\perp = \operatorname{span}(e_3) = \operatorname{span}\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right).$$

6. Real Symmetric Matrices

We show here that any real symmetric matrix can be diagonalised by an "orthogonal" matrix P. This gives applications to quadratic forms in the final section.

First, a quick observation about inner products. Let $\langle \cdot, \cdot \rangle$ be an inner product on the real vector space V, with an orthonormal basis $e_1, \dots, e_n$ for V. We shall write vectors $v \in V$ as column vectors with respect to the e_i basis (but for reasons of space, these will be denoted as transposes of row vectors). So if $v = \sum_{i=1}^n a_i e_i$, we write $v = (a_1, \dots, a_n)^T$. Now observe that $\langle v, w \rangle = w^T v$ (a matrix product of a row vector by a column vector). For if $v = a_1 e_1 + \dots + a_n e_n$ and $w = b_1 e_1 + \dots + b_n e_n$, then

$$\langle v, w \rangle = \sum_{i=1}^n a_i b_i = (b_1, \dots, b_n) \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = w^T v.$$

Recall that a matrix A is *symmetric* if $A = A^T$, that is, if $A = (a_{ij})$, then $a_{ij} = a_{ji}$ for all i, j.

Recall also that if matrix A represents the linear transformation $\alpha : V \to V$ with respect to the basis $e_1, \dots, e_n$, then $\alpha(\sum_{j=1}^n x_j e_j) = \sum_{i=1}^n y_i e_i$, where $A(x_1, \dots, x_n)^T = (y_1, \dots, y_n)^T$, so $y_i = \sum_{j=1}^n a_{ij} x_j$ for each $i = 1, \dots, n$.

Definition 6.1. The linear transformation $\alpha : V \to V$ is called *self-adjoint* if, for all $u, v \in V$, we have

$$\langle \alpha u, v \rangle = \langle u, \alpha v \rangle.$$

Theorem 6.2. *Let $\alpha : V \to V$ be linear, represented by A with respect to the orthonormal basis $e_1, \ldots, e_n$ for V. Then A is symmetric if and only if α is self-adjoint.*

Proof. $\Leftarrow$ Suppose $\langle \alpha u, v \rangle = \langle u, \alpha v \rangle$ for al $u, v \in V$. Put $u = e_i$ and $v = e_j$. Then, (with the subscripts i, j indicating which vector entry is 1),

$$\langle \alpha u, v \rangle = v^T \cdot \alpha u = (0, \ldots, 0, 1_j, 0, \ldots, 0) A (0, \ldots, 0, 1_i, 0, \ldots 0)^T$$

$$= (0, \ldots, 0, 1_j, 0, \ldots, 0) \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix} = a_{ji}.$$

Also,

$$\langle u, \alpha v \rangle = (a_{1j}, \ldots, a_{nj})(0, \ldots, 0, 1_i, 0, \ldots, 0)^T = a_{ij}.$$

Hence, $a_{ij} = a_{ji}$.

$\Rightarrow$ Suppose A is symmetric. Let $v = (v_1, \ldots, v_n)^T$ and $u = (u_1, \ldots, u_n)^T$. Then $\langle \alpha u, v \rangle = v^T A u$, and $\langle u, \alpha v \rangle = (Av)^T u = v^T A^T u = v^T A u$ (as A is symmetric). Hence, $\langle \alpha u, v \rangle = \langle u, \alpha v \rangle$.

□

Theorem 6.3. *Let $\alpha : V \to V$ be self-adjoint, represented by the real symmetric matrix A with respect to the orthonormal basis $e_1, \ldots, e_n$ of V. Then*

(1) *The eigenvalues of A (and hence α) are all real;*
(2) *eigenvalues corresponding to distinct eigenvectors are orthogonal;*
(3) *V has an orthonormal basis of eigenvectors of A.*

In particular, α is diagonalisable.

Proof. (1) Suppose that λ is an eigenvalue of A, with eigenvector $x_1 e_1 + \ldots + x_n e_n$. Put $x := (x_1, \ldots, x_n)^T$. Write $\bar{\lambda}$ for the complex conjugate of λ, and work over $\mathbb{C}$. (So if $\lambda = a + bi$, then $\bar{\lambda} = a - bi$.) Also, for the vector x, write $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)^T$.

Now $Ax = \lambda x$, so $A\bar{x} = \overline{\lambda x} = \bar{\lambda}\bar{x}$. Taking transposes, it follows that $\bar{x}^T A^T = \bar{\lambda}\bar{x}^T$. Hence, multiplying on the right by the $n \times 1$ matrix x,

$$\bar{x}^T A^T x = \bar{\lambda}\bar{x}^T x.$$

Also,

$$\lambda \bar{x}^T x = \bar{x}^T (\lambda x) = \bar{x}^T (Ax) = \bar{x}^T A^T x,$$

the last step as $A = A^T$. Thus, $\bar{\lambda}\bar{x}^T x = \lambda \bar{x}^T x$. As $x \neq 0$, it follows that $\lambda = \bar{\lambda}$, that is, λ is real.

(2) Suppose $\lambda \neq \mu$ are eigenvalues, with eigenvectors v, w respectively. So $\alpha x = \lambda x$ and $\alpha y = \mu y$. Then

$$\lambda \langle x, y \rangle = \langle \lambda x, y \rangle = \langle \alpha x, y \rangle = \langle x, \alpha y \rangle = \langle x, \mu y \rangle = \mu \langle x, y \rangle.$$

As $\lambda \neq \mu$, it follows that $\langle x, y \rangle = 0$.

(3) We use induction on $n = \dim V$. If $n = 1$ the result is obvious.

By 1, there is a real eigenvalue λ_1 of α, with eigenvector x_1. We may suppose that $\|x_1\| = 1$ (normalise). Put $W := (\mathrm{span}(x_1))^\perp < V$. Then $\dim(W) = n - 1$, by Theorem 5.15(ii).

We claim that $\alpha(W) \subseteq W$. To see this, observe that if $w \in W$, then

$$\langle \alpha w, x_1 \rangle = \langle w, \alpha x_1 \rangle = \langle w, \lambda_1 x_1 \rangle = \lambda_1 \langle w, x_1 \rangle = 0,$$

so $\alpha w \in W$.

It follows that α induces a self-adjoint mapping $W \to W$. Since $\dim(W) = n - 1$, there is an orthonormal basis $w_2, \ldots, w_n$ of eigenvectors of α in W. Now $x_1, w_2, \ldots, w_n$ is an orthonormal basis of eigenvectors of α in V.

□

Definition 6.4. An $n \times n$ matrix P is *orthogonal* if $P^T = P^{-1}$, that is $P^T P = I$.

The above definition esentially says that P is orthogonal if and only if its columns form an orthonormal basis of $\mathbb{R}^n$. In the 2×2 case, an example is $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

Lemma 6.5. *Let $\alpha : V \to V$ be represented by a matrix A with respect to some orthonormal basis $e_1, \ldots, e_n$ of V. Then if we change to another orthonormal basis $f_1, \ldots, f_n$ of V, the matrix for α in this basis has form $P^{-1}AP$, where P is orthogonal.*

Proof. We have $f_j = \sum_{i=1}^n p_{ij} e_i$, where $P = (p_{ij})$ (see Theorem 3.20). Now

$$\langle f_j, f_k \rangle = \langle \sum_{i=1}^n p_{ij} e_i, \sum_{i=1}^n p_{ik} e_i \rangle = \sum_{i=1}^n p_{ij} p_{ik}.$$

Hence, $\sum_{i=1}^n p_{ij} p_{ik}$ takes value 1 if $j = k$, and zero otherwise. This just says $P^T P = I$. □

The following corollary is really the point of this section. It says in particular than any real symmetric matrix is diagonalisable, but is much stronger than this (because P is orthogonal).

Corollary 6.6. *If A is a real symmetric matrix, then there is an orthogonal matrix P with $P^{-1}AP = P^T AP$ equal to a diagonal matrix D.*

Proof. Take $e_1, \ldots, e_n$ as the standard basis of $\mathbb{R}^n$, and $f_1, \ldots, f_n$ as an orthonormal basis of eigenvectors of A (given by Theorem 6.3). Then the matrix D of A with respect to $f_1, \ldots, f_n$ is diagonal (as it is a basis of eigenvectors). Also, by Lemma 6.5, there is orthogonal P with $P^{-1}AP = D$. □

Example 6.7. Let $\alpha : \mathbb{R}^3 \to \mathbb{R}^3$ be given by the matrix $A = \begin{pmatrix} 0 & 2 & 0 \\ 2 & 2 & 2 \\ 0 & 2 & 0 \end{pmatrix}$ with respect to the standard basis. Note that A is symmetric, so there is an orthonormal basis of $\mathbb{R}^3$ consisting of eigenvectors of A. First, calculating $\det(tI - A)$ along the first row, we find $\chi(t) = t(t(t-2) - 4) + 2(-2t) = t(t^2 - 2t - 8) = t(t-4)(t+2)$, so the eigenvalues are $0, 4, -2$.

For $\lambda = 0$, solving

$$\begin{pmatrix} 0 & -2 & 0 \\ -2 & -2 & -2 \\ 0 & -2 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

gives eigenvector $(1, 0, -1)^T$. For $\lambda = 4$, we find eigenvector $(1, 2, 1)^T$. For $\lambda = -2$, an eigenvector is $(1, -1, 1)^T$. Normalising these, we find

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{3}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}.$$

Then

$$P^{-1}AP = P^T AP = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

7. Quadratic Forms

Let V be an n-dimensional real vector space, with basis $v_1, \ldots, v_n$. So any vector $x \in V$ can be written in the form $x = x_1 v_1 + \ldots + x_n v_n$.

Definition 7.1. In this setting, a *quadratic form* is a function $q : V \to \mathbb{R}$ of the form

$$q(x) = q(x_1 v_1 + \ldots + x_n v_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_i x_j,$$

where each $c_{ij} \in \mathbb{R}$.

Example 7.2.

(1) The standard norm of vector produced by the inner product $\langle x, x \rangle = x_1^2 + x_2^2 + \ldots + x_n^2$ in $\mathbb{R}^n$ is a quadratic form. Moreover a norm from any inner product is a quadratic form as well.

(2) Working in the standard basis of $\mathbb{R}^2$, with $x = (x_1, x_2)$, the function $q(x) = 3x_1^2 + 4x_1x_2 - x_2^2$ is a quadratic form. We can write $q(x)$ as a matrix product

$$(x_1 x_2) \begin{pmatrix} 3 & 4 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1 x_2) \begin{pmatrix} 3x_1 + 4x_2 \\ -x_2 \end{pmatrix}.$$

It can also be written as

$$(x_1 x_2) \begin{pmatrix} 3 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

The second matrix expression above suggests there is a connection between quadratic forms and real symmetric matrices. We now formalise this.

Proposition 7.3. *Working with respect to a fixed basis of V, any quadratic form $q(x)$ can be written in the form*

$$q(x) = (x_1 \ldots x_n) A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

where A is a real symmetric *matrix. Furthermore, such an expression is unique, and every $n \times n$ real symmetric matrix gives rise to a quadratic form on V.*

Proof. We just prove the first assertion. Suppose our form is

$$q(x) = q(x_1v_1 + \ldots + x_nv_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_ix_j.$$

Put $a_{ij} := \frac{c_{ij}+c_{ji}}{2}$. Then $A = (a_{ij})$ is symmetric, and

$$q(x) = \sum_{i=1}^{b}\sum_{j=1}^{n} a_{ij}x_ix_j = (x_1 \ldots x_n)A\left(x_1 \ldots x_n\right)^T.$$

□

The last proposition says that, if we write $x = x_1v_1 + \ldots + x_nv_n$ as the column vector $(x_1, \ldots, x_n)^T$, then $q(x) = x^TAx$, where A is real symmetric. We would like to change the basis of V, to ensure that the matrix A is diagonal.

Proposition 7.4. *Suppose that $w_1, \ldots, w_n$ is another basis of V, with $w_j = \sum_{i=1}^{n} m_{ij}v_i$, for each j. Put $M = (m_{ij})$. Then*

(1) $\sum_{i=1}^{n} x_iv_i = \sum_{j=1}^{n} y_jw_j$, *where* $(x_1 \ldots x_n)^T = M(y_1 \ldots y_n)^T$.
(2) *In the w_i basis, the quadratic form $q(x) = x^TAx$ becomes*

$$q(x) = (y_1 \ldots y_n)(M^TAM)\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

and the matrix M^TAM is symmetric. (Here, $(y_1, \ldots, y_n)$ is the sequence of coordinates of the vector x with respect to the w basis, that is, $x = y_1w_1 + \ldots + y_nw_n$.)

Proof. (1) We indeed have:

$$\sum_{j=1}^{n} y_jw_j = \sum_{j=1}^{n} y_j \sum_{i=1}^{n} m_{ij}v_i = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} m_{ij}y_j\right)v_i = \sum_{i=1}^{n} x_iv_i.$$

(2) Now $q(x) = x^TAx = (My)^TA(My) = y^T(M^TAM)y$.

□

Definition 7.5. We say that square matrices A, B are *congruent* if there is non-singular M with $B = M^TAM$.

Remark 7.6. (1) In Proposition 7.4, the jth column of M is the tuple of coordinates of w_j written in the v-basis.
(2) By Corollary 6.6, any real symmetric matrix is congruent to a diagonal matrix.
(3) In the first year Geometry module you used orthogonal transformations of $\mathbb{R}^2$ (which are rotations) to diagonalise quadratic forms in order to obtain canonical equations of ellipses, parabolas and hyperbolas, cf. Example 7.13.

Theorem 7.7. *For any quadratic form q on V, there is a choice of basis with redspect to which q is diagonal, that is, q is given as*

$$q(y) = y^TDy = y^T\,\mathrm{Diag}(d_1, \ldots, d_n)y$$

where $D = \mathrm{Diag}(d_1, \ldots d_n)$ is a diagonal matrix. Equivalently, $q(y) = d_1y_1^2 + d_2y_2^2 + \ldots + d_ny_n^2$.

Proof. We must show that if A is a real symmetric matrix, then there is a non-singular matrix M such that M^TAM is diagonal. By Corollary 6.6, there is an orthogonal matrix P such that $P^{-1}AP$ is diagonal, and as P is orthogonal, we have $P^{-1} = P^T$. □

The proof of Theorem 7.7 gives a method of diagonalising a quadratic form q:

(1) Find the real symmetric matrix A and its characteristic polynomial $\chi(t)$;
(2) Factorise $\chi(t)$ to find the eigenvalues
(3) Find a basis of eigenvectors, let M have these as columns, and calculate M^TAM.

This process is slow – for example, it may be difficult to factorise $\chi(t)$. Since we do not require that $M^T = M^{-1}$, there are much easier methods. The main point is that premultiplying by a non-singular matrix M^T has the effect of doing a sequence of row operations to A, and postmultiplying by M has the effect of doing the corresponding sequence of column operations. So we aim to diagonalise A by doing a sequence of row operations and the corresponding sequence of column operations. Here's an easy example.

Example 7.8. We diagonalise the quadratic form $q = x^2 + 4xy + 2yz$. (I have switched from having variables $x_1, \ldots, x_n$ to variables x, y, z – you will see both conventions in past papers.) The matrix here is $\begin{pmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$. To diagonalise it, we first aim to make the 2's in the (1,2) and (2,1) entries into 0. So do $r_2' = r_2 - 2r_1$ and then $c_2' = c_2 - 2c_1$, to get

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

This is still symmetric. To get rid of the 1's in the $(3,2)$ and $(2,3)$ entries, do $r_3' = r_3 + r_2/4$ and then $c_3' = c_3 + c_2/4$. We get $\begin{pmatrix} 1 & 0 & 0 \\ 0 & -4 & 1 \\ 0 & 0 & \frac{1}{4} \end{pmatrix}$ and then $\begin{pmatrix} 1 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix}$. Thus, with respect to a different basis, the quadratic form can be written in the diagonal form $q = u^2 - 4v^2 + \frac{1}{4}w^2$.

In general, the algorithm for diagonalising a quadratic form is slightly more complicated. For example, in the last example, how would we have got rid of the 2's in the $(1,2)$ and $(2,1)$ entries if the $(1,1)$ entry had been 0? Here is a sketch of the general procedure. It ensures that there is no non-zero entry in the first row or column except possibly in the $(1,1)$ entry. After we have achieved that, then we apply the same process to the $(n-1) \times (n-1)$ matrix obtained by deleting the first row and column.

(1) If $a_{11} \neq 0$, then for each $i > 1$ do operations $r_i' = r_i + ar_1$ and then $c_i' = c_i + ac_1$ for appropriate a.
(2) If $a_{11} = 0$ but $a_{ii} \neq 0$ for some $i > 1$, interchange rows r_1 and r_i, and then interchange columns c_1 and c_i, to get a matrix with a_{ii} in the $(1,1)$ entry. Then apply (i) above.
(3) Suppose all the diagonal entries a_{ii} are zero. Find some j with $a_{ij} \neq 0$, and then do $r_i' = r_i + r_j$ and then $c_i' = c_i + c_j$. The (i,i) entry is now non-zero, so proceed as in (ii).

Example 7.9. Consider the quadratic form $q = 4yz$ (a function of x, y, z). This has matrix $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix}$. There is no non-zero entry in first row or column, and no non-zero diagonal entry, so use method (iii). Do $r_2' = r_2 + r_3$ then $c_2' = c_2 + c_3$ to get $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 0 \end{pmatrix}$. Now do $r_3' = r_3 - \frac{1}{2}r_2$ and then $c_3' = c_3 - \frac{1}{2}c_2$, to get $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -1 \end{pmatrix}$. Thus, a diagonalised form is $q = 4v^2 - w^2$.

A given quadratic form can usually be diagonalised in many different ways, i.e., there is no *unique* answer. However, the following definition and theorem give a partial uniqueness.

Definition 7.10. Let $q(y) = d_1y_1^2 + \ldots + d_ny_n^2$ be a quadratic form in diagonal form. Let P be the number of strictly positive d_i, and N be the number of strictly negative d_i. Then the *rank* of q is defined to be $\operatorname{rank}(q) = P + N$, and the *signature* of q is defined to be $\operatorname{signature}(q) = P - N$.

Theorem 7.11 (Sylvester's Law of Inertia)**.** *Let $q(x) = x^TAx$ be a quadratic form, and suppose it has two diagonalised forms $x = My$ and $x = Nz$, so $M^TAM = \operatorname{Diag}(c_1, \ldots, c_n)$ and $N^TAN = \operatorname{Diag}(d_1, \ldots, d_n)$. Then the ranks of $c_1y_1^2 + \ldots + c_ny_n^2$ and $d_1y_1^2 + \ldots + d_ny_n^2$ are equal, and their signatures are also equal.*

Proof. This is omitted. □

From the last theorem, it follows that we can define the rank of a quadratic form to be the rank of *any* quadratic form of it. Likewise for signature.

Example 7.12. In Example 7.8 we found a diagonal form $u^2 - 4v^2 + \frac{1}{4}w^2$. Thus, the rank is $2 + 1 = 3$, and the signature is $2 - 1 = 1$.

In Example 7.9, a diagonal form was $4v^2 - w^2$. The rank is $1 + 1 = 2$, and the signature is $1 - 1 = 0$.

Rank and signature have certain geometrical meaning.

Example 7.13. Consider the equation $cx^2 + dy^2 = 1$.

- If $c, d > 0$, this is an *ellipse*.
- If $c > 0$ and $d < 0$ then it is a *hyperbola*.
- If $c > 0$ and $d = 0$, it is the union of 2 lines (a *degenerate* case).

INDEX

SCHOOL OF MATHEMATICS, UNIVERSITY OF LEEDS, LEEDS LS2 9JT, UK
Email address: kisilv@maths.leeds.ac.uk
URL: http://www.maths.leeds.ac.uk/~kisilv/