Draft 1.4, Version of 13 November 2007

Notes on Linear Algebra

by Peter M Neumann (Queen's College, Oxford)

Preface

These notes are intended as a rough guide to the course *Further Linear Algebra* which is a part of the Oxford 2nd year undergraduate course in mathematics. Please do not expect a polished account. They are lecture notes, not a carefully checked text-book. Nevertheless, I hope they may be of some help.

The course is designed to build on 1st year Linear Algebra. The syllabus for that course includes matrices, row reduction of matrices to echelon form, rank of a matrix, solution of simultaneous linear equations, vector spaces and their subspaces, linear dependence and independence, bases and dimension of a vector space, linear transformations and their matrices with respect to given bases, and eigenvalues and eigenvectors of a square matrix or of a linear transformation of a vector space to itself. In that first-year work the coefficients of our matrices or linear equations are almost exclusively real numbers; accordingly, the field over which our vector spaces are defined is $\mathbb{R}$.

A significant change of attitude occurs now. Our first task is to examine what happens when $\mathbb{R}$ is replaced by an arbitrary field F of coefficients. We find that the basic theory of matrices, linear equations, vector spaces and their subspaces, linear transformations and their matrices is unchanged. The theory of eigenvalues and eigenvectors does depend on the field F, however. And when we come to the "metric" or "geometric" theory associated with inner products, orthogonality, and the like, we find that it is natural to return to vector spaces over $\mathbb{R}$ or $\mathbb{C}$.

As a consequence the lecture course, and therefore also this set of notes, naturally divides into four parts: the first is a study of vector spaces over arbitrary fields; then we study linear transformations of a vector space to itself; third, a treatment of real or complex inner product spaces; and finally the theory of adjoints of linear transformations on inner product spaces.

It is a pleasure to acknowledge with warm thanks that these notes have benefitted from comments and suggestions by Dr Jan Grabowski. Any remaining errors, infelicities and obscurities are of course my own responsibility—I would welcome feedback.

ΠMN: Queen's: 13.xi.2007

CONTENTS

Part I: Fields and Vector Spaces

Part I: Fields and Vector Spaces

As has been indicated in the preface, our first aim is to re-work the linear algebra presented in the Mods course, generalising from $\mathbb{R}$ to an arbitrary field as domain from which coefficients are to be taken. Fields are treated in the companion course, *Rings and Arithmetic*, but to make this course and these notes self-contained we begin with a definition of what we mean by a field.

Axioms for Fields

A *field* is a set F with distinguished elements 0, 1, with a unary operation $-$, and with two binary operations $+$ and $\times$ satisfying the axioms below (the 'axioms of arithmetic'). Conventionally, for the image of (a,b) under the function $+ : F \times F \to F$ we write $a+b$; for the image of (a,b) under the function $\times : F \times F \to F$ we write $a\,b$; and $x-y$ means $x+(-y)$, $x+y\,z$ means $x+(y\,z)$.

The axioms of arithmetic

(1) $a+(b+c)=(a+b)+c$ [+ is associative]

(2) $a+b=b+a$ [+ is commutative]

(3) $a+0=a$

(4) $a+(-a)=0$

(5) $a\,(b\,c)=(a\,b)\,c$ [× is associative]

(6) $a\,b=b\,a$ [× is commutative]

(7) $a\,1=a$

(8) $a\neq 0 \Rightarrow \exists\, x\in F : a\,x=1$

(9) $a\,(b+c)=a\,b+a\,c$ [× distributes over +]

(10) $0\neq 1$

Note 1: All axioms are understood to have $\forall\, a,b,\ldots \in F$ in front.

Note 2: See my *Notes on Rings and Arithmetic* for more discussion.

Note 3: Examples are $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}_2$ or more generally $\mathbb{Z}_p$, the field of integers modulo p for any prime number p.

Vector Spaces

Let F be a field. A *vector space over F* is a set V with distinguished element 0, with a unary operation $-$, with a binary operation $+$, and with a function $\times : F\times V\to V$ satisfying the axioms below. Conventionally: for the image of (a,b) under the function $+ : V\times V\to V$ we write $a+b$; for $a+(-b)$ we write $a-b$; and for the image of (α,v) under the function $\times : F\times V\to V$ we write $\alpha\,v$.

Vector space axioms

(1)	$u + (v + w) = (u + v) + w$	[+ is associative]
(2)	$u + v = v + u$	[+ is commutative]
(3)	$u + 0 = u$	
(4)	$u + (-u) = 0$	
(5)	$\alpha\,(\beta\,v) = (\alpha\,\beta)\,v$	
(6)	$\alpha\,(u + v) = \alpha\,u + \alpha\,v$	
(7)	$(\alpha + \beta)\,v = \alpha\,v + \beta\,v$	
(8)	$1\,v = v$	

NOTE: All axioms are understood to have appropriate quantifiers $\forall\, u, v, \ldots \in V$ and/or $\forall\, \alpha, \beta, \ldots \in F$ in front.

EXAMPLES:

- F^n is a vector space over F;
- the polynomial ring $F[x]$ (see *Notes on Rings and Arithmetic*) is a vector space over F;
- $\mathrm{M}_{m\times n}(F)$ is a vector space over F;
- if K is a field and F a subfield then K is a vector space over F;
- etc., etc., etc.

EXERCISE 1. Let X be any set, F any field. Define F^X to be the set of all functions $X \to F$ with the usual point-wise addition and multiplication by scalars (elements of F). Show that F^X is a vector space over F.

Exactly as for rings, or for vector spaces over $\mathbb{R}$, one can prove important "trivialities". Of course, if we could not prove them then we would add further axioms until we had captured all properties that we require, or at least expect, of the algebra of vectors. But the fact is that this set of axioms, feeble though it seems, is enough. For example:

PROPOSITION: *Let V be a vector space over the field F. For any $v \in V$ and any $\alpha \in F$ we have*

(i) $0\,v = 0$*;*

(ii) $\alpha\,0 = 0$*;*

(iii) *if $\alpha\,v = 0$ then $\alpha = 0$ or $v = 0$;*

(iv) $\alpha\,(-v) = -(\alpha\,v) = (-\alpha)\,v$*;*

Proof. For $v \in V$, from Field Axiom (3) and Vector Space Axiom (7) we have

$$0\,v = (0 + 0)\,v = 0\,v + 0\,v.$$

Then adding $-(0\,v)$ to both sides of this equation and using Vector Space Axioms (4) on the left and (1), (4), (3) on the right, we get that $0 = 0\,v$, as required for (i). The reader is invited to give a proof of (ii).

For (iii), suppose that $\alpha v = 0$ and $\alpha \neq 0$: our task is to show that $v = 0$. By Field Axioms (8) and (6) there exists $\beta \in F$ such that $\beta \alpha = 1$. Then

$$v = 1 v = (\beta \alpha) v = \beta (\alpha v) = \beta 0$$

by Vector Space Axioms (8) and (5). But $\beta 0 = 0$ by (ii), and so $v = 0$, as required. Clause (iv), like Clause (ii), is offered as an exercise:

Exercise 2: Prove Clauses (ii) and (iv) of the above proposition.

Subspaces

Let V be a vector space over a field F. A *subspace* of V is a subset U such that

(1) $0 \in U$ and $u + v \in U$ whenever $u, v \in U$ and $-u \in U$ whenever $u \in U$;

(2) if $u \in U$, $\alpha \in F$ then $\alpha u \in U$.

Note 1: Condition (1) says that U is an additive subgroup of V. Condition (2) is closure under multiplication by scalars.

Note 2: We write $U \leqslant V$ to mean that U is a subspace of V.

Note 3: Always $\{0\}$ is a subspace; if $U \neq \{0\}$ then we say that U is *non-zero* or *non-trivial*. Likewise, V is a subspace; if $U \neq V$ then we say that U is a *proper* subspace.

Note 4: A subset U of V is a subspace if and only if $U \neq \emptyset$ and U is closed under + (that is, $u, v \in U \Rightarrow u + v \in U$) and under multiplication by scalars. The proof is offered as an exercise:

Exercise 3: Suppose that $U \subseteq V$. Show that $U \leqslant V$ if and only if

(1′) $U \neq \emptyset$ and $u + v \in U$ whenever $u, v \in U$;

(2) if $u \in U$, $\alpha \in F$ then $\alpha u \in U$.

Examples: (1) Let $L_1, \ldots, L_m$ be homogeneous linear expressions $\sum c_{ij} x_j$ with coefficients $c_{ij} \in F$, and let

$$U := \{(x_1, \ldots, x_n) \in F^n \mid L_1 = 0, \ \ldots, \ L_m = 0\}.$$

Then $U \leqslant F^n$.

(2) Let $F^{[n]}[x] := \{f \in F[x] \mid f = 0 \text{ or } \deg f \leqslant n\}$. Then $F^{[n]}[x] \leqslant F[x]$.

(3) Upper triangular matrices form a subspace of $\mathrm{M}_{n\times n}(F)$.

Exercise 4. With F^X as in Exercise 1, for $f \in F^X$ define the *support* of f by $\mathrm{supp}(f) := \{x \in X \mid f(x) \neq 0\}$. Define $U := \{f \in F^X \mid \mathrm{supp}(f) \text{ is finite}\}$. Show that U is a subspace of F^X.

Exercise 5. Let U_1, U_2, … be *proper* subspaces of a vector space V over a field F (recall that the subspace U is said to be proper if $U \neq V$).

(i) Show that $V \neq U_1 \cup U_2$. [*Hint:* what happens if $U_1 \subseteq U_2$ or $U_2 \subseteq U_1$? Otherwise, take $u_1 \in U_1 \setminus U_2$, $u_2 \in U_2 \setminus U_1$, and show that $u_1 + u_2 \notin U_1 \cup U_2$.]

(ii) Show that if $V = U_1 \cup U_2 \cup U_3$ then F must be the field $\mathbb{Z}_2$ with just 2 elements. [*Hint:* show first that we cannot have $U_1 \subseteq U_2 \cup U_3$, nor $U_2 \subseteq U_1 \cup U_3$; choose $u_1 \in U_1 \setminus (U_2 \cup U_3)$ and $u_2 \in U_2 \setminus (U_1 \cup U_3)$; observe that if $\lambda \in F \setminus \{0\}$ then $u_1 + \lambda u_2$ must lie in U_3 and exploit this fact.]

(iii) Show that if F is infinite (indeed, if $|F| > n-1$) then $V \neq U_1 \cup U_2 \cup \cdots \cup U_n$.

Quotient spaces

Suppose that $U \leqslant V$ where V is a vector space over a field F. Define the *quotient space* V/U as follows:

set $:= \{x + U \mid x \in V\}$ [additive cosets]

$0 := U$

additive inverse: $-(x+U) := (-x) + U$

addition: $(x+U) + (y+U) := (x+y) + U$

multiplication by scalars: $\alpha(x+U) := \alpha x + U$

Exercise 6 (worth doing carefully once in one's life, but not more than once—unless an examiner offers marks for it): Check that $-$, $+$ and multiplication by scalars are well defined, and that the vector space axioms hold in V/U.

Note: The notion of quotient space is closely analogous with the notion of quotient of a group by a normal subgroup or of a ring by an ideal. It is not in the Part A syllabus, nor will it play a large part in this course. Nevertheless, it is an important and useful construct which is well worth becoming familiar with.

Dimension (Revision)

Although technically new, the following ideas and results translate so simply from the case of vector spaces over $\mathbb{R}$ to vector spaces over an arbitrary field F that I propose simply to list headers for revision:

(1) spanning sets; linear dependence and independence; bases;

(2) dimension;

(3) $\dim V = d \;\Rightarrow\; V \cong F^d$;

(4) any linearly independent set may be extended (usually in many ways) to a basis;

(5) intersection $U \cap W$ of subspaces; sum $U + W$;

(6) $\dim(U+W) = \dim U + \dim W - \dim(U \cap W)$;

Exercise 7. Prove that if V is finite-dimensional and $U \leqslant V$ then

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

Exercise 8. With F^X as in Exercise 1, show that F^X is finite-dimensional if and only if X is finite, and then $\dim(F^X) = |X|$.

Linear Transformations (Revision)

Let F be a field, V_1, V_2 vector spaces over F. A map $T : V_1 \to V_2$ is said to be *linear* if

$$T\,0 = 0, \quad T(-x) = -T\,x, \quad T(x+y) = T\,x + T\,y, \quad \text{and } T(\lambda\,x) = \lambda\,(T\,x)$$

for all $x, y \in V$ and all $\lambda \in F$.

Note 1: The definition is couched in terms which are intended to emphasise that what should be required of a linear transformation (homomorphism of vector spaces) is that it preserves all the ingredients, 0, $+$, $-$ and multiplication by scalars, that go into the making of a vector space. What we use in practice is the fact that $T : V_1 \to V_2$ is linear if and only if $T(\alpha\,x + \beta\,y) = \alpha\,T\,x + \beta\,T\,y$ for all $x, y \in V$ and all $\alpha, \beta \in F$. The proof is an easy exercise and is left to the reader.

Note 2: The identity $I : V \to V$ is linear; if $T : V_1 \to V_2$ and $S : V_2 \to V_3$ are linear then $S \circ T : V_1 \to V_3$ is linear.

For a linear transformation $T : V \to W$ we define the *kernel* or *null-space* by $\operatorname{Ker} T := \{x \in V \mid T\,x = 0\}$. We prove that $\operatorname{Ker} T \leqslant V$, that $\operatorname{Im} T \leqslant W$ and we define

$$\operatorname{nullity} T := \dim \operatorname{Ker} T, \qquad \operatorname{rank} T := \dim \operatorname{Im} T.$$

Rank-Nullity Theorem: $\operatorname{nullity} T + \operatorname{rank} T = \dim V$

Note: The Rank-Nullity Theorem is a version of the First Isomorphism Theorem for vector spaces, which states that $\operatorname{Im} T \cong V/\operatorname{Ker} T$.

Exercise 9. Let V be a finite dimensional vector space over a field F and let $T : V \to V$ be a linear transformation. For $\lambda \in F$ define $E_\lambda := \{v \in V \mid Tv = \lambda v\}$.

(i) Check that E_λ is a subspace of V.

(ii) Suppose that $\lambda_1, \ldots, \lambda_m$ are distinct. For $i = 1, \ldots, m$ let $v_i \in E_{\lambda_i} \setminus \{0\}$. Show that $v_1, \ldots, v_m$ are linearly independent.

(iii) Suppose further that $S : V \to V$ is a linear transformation such that $ST = TS$. Show that $S(E_\lambda) \subseteq E_\lambda$ for each $\lambda \in F$.

Direct sums and projection operators

The vector space V is said to be the *direct sum* of its subspaces U and W, and we write $V = U \oplus W$, if $V = U + W$ and $U \cap W = \{0\}$.

Lemma: *Let U, W be subspaces of V. Then $V = U \oplus W$ if and only if for every $v \in V$ there exist unique vectors $u \in U$ and $w \in W$ such that $v = u + w$.*

Proof. Suppose first that for every $v \in V$ there exist unique vectors $u \in U$ and $w \in W$ such that $v = u + w$. Certainly then $V = U + W$ and what we need to prove is that $U \cap W = \{0\}$. So let $x \in U \cap W$. Then $x = x + 0$ with $x \in U$ and $0 \in W$. Equally, $x = 0 + x$ with $0 \in U$ and $x \in W$. But the expression $x = u + w$ with $u \in U$ and $w \in W$ is, by assumption, unique, and it follows that $x = 0$. Thus $U \cap W = \{0\}$, as required.

Now suppose that $V = U \oplus W$. If $v \in V$ then since $V = U + W$ there certainly exist vectors $u \in U$ and $w \in W$ such that $v = u + w$. The point at issue therefore is: are u, w uniquely determined by v? Suppose that $u + w = u' + w'$, where $u, u' \in U$ and $w, w' \in W$. Then $u - u' = w' - w$. This vector lies both in U and in W. By assumption, $U \cap W = \{0\}$ and so $u - u' = w' - w = 0$. Thus $u = u'$ and $w = w'$, so the decomposition of a vector v as $u + w$ with $u \in U$ and $w \in W$ is unique, as required.

Note: What we are discussing is sometimes (but rarely) called the "internal" direct sum to distinguish it from the natural construction which starts from two vector spaces V_1, V_2 over the same field F and constructs a new vector space whose set is the product set $V_1 \times V_2$ and in which the vector space structure is defined componentwise—compare the direct product of groups or of rings. This is (equally rarely) called the "external" direct sum. These are two sides of the same coin: the external direct sum of V_1 and V_2 is the internal direct sum of its subspaces $V_1 \times \{0\}$ and $\{0\} \times V_2$; while if $V = U \oplus W$ then V is naturally isomorphic with the external direct sum of U and W.

Exercise 10. Let V be a finite-dimensional vector space over $\mathbb{R}$ and let U be a non-trivial proper subspace. Prove that there are infinitely many different subspaces W of V such that $V = U \oplus W$. [*Hint:* think first what happens when V is 2-dimensional; then generalise.] How far can this be generalised to vector spaces over other fields F?

We come now to *projection operators.* Suppose that $V = U \oplus W$. Define $P : V \to V$ as follows. For $v \in V$ write $v = u + w$ where $u \in U$ and $w \in W$ and then define $P v := u$. Strictly P depends on the ordered pair (U, W) of summands of V, but to keep things simple we will not build this dependence into the notation.

Observations:

(1) *P is well-defined;*

(2) *P is linear;*

(3) $\operatorname{Im} P = U$, $\operatorname{Ker} P = W$;

(4) $P^2 = P$.

Proofs. That P is well-defined is an immediate consequence of the existence and uniqueness of the decomposition $v = u + w$ with $u \in U$, $v \in V$.

To see that P is linear, let $v_1, v_2 \in V$ and $\alpha_1, \alpha_2 \in F$ (where, as always, F is the field of scalars). Let $v_1 = u_1 + w_1$ and $v_2 = u_2 + w_2$ be the decompositions of v_1 and v_2. Then $P v_1 = u_1$ and $P v_2 = u_2$. What about $P(\alpha_1 v_1 + \alpha_2 v_2)$? Well,

$$\alpha_1 v_1 + \alpha_2 v_2 = \alpha_1(u_1 + w_1) + \alpha_2(u_2 + w_2) = (\alpha_1 u_1 + \alpha_2 u_2) + (\alpha_1 w_1 + \alpha_2 w_2).$$

Since $\alpha_1 u_1 + \alpha_2 u_2 \in U$ and $\alpha_1 w_1 + \alpha_2 w_2 \in W$ it follows that $P(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 u_1 + \alpha_2 u_2$. Therefore

$$P(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 P(v_1) + \alpha_2 P(v_2),$$

that is, P is linear.

For (3) it is clear from the definition that $\operatorname{Im} P \subseteq U$; but if $u \in U$ then $u = P u$, and therefore $\operatorname{Im} P = U$. Similarly, it is clear that $W \subseteq \operatorname{Ker} P$; but if $v \in \operatorname{Ker} P$ then $v = 0 + w$ for some $w \in W$, and therefore $\operatorname{Ker} P = W$.

Finally, if $v \in V$ and we write $v = u + w$ with $u \in U$ and $w \in W$ then

$$P^2 v = P(P v) = P u = u = P v\,,$$

and this shows that $P^2 = P$, as required.

Terminology: the operator (linear transformation) P is called the *projection* of V onto U along W.

Note 1. Suppose that V is finite-dimensional. Choose a basis $u_1, \ldots, u_r$ for U and a basis $w_1, \ldots, w_m$ for W. Then the matrix of P with respect to the basis $u_1, \ldots, u_r, w_1, \ldots, w_m$ of V is

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

Note 2. If P is the projection onto U along W then $I - P$, where $I : V \to V$ is the identity transformation, is the projection onto W along U.

Note 3. If P is the projection onto U along W then $u \in U$ if and only if $P u = u$. The fact that if $u \in U$ then $P u = u$ is immediate from the definition of P, while if $P u = u$ then obviously $u \in \operatorname{Im} P = U$.

Our next aim is to characterise projection operators algebraically. It turns out that Observation (4) above is the key:

Terminology: An operator T such that $T^2 = T$ is said to be *idempotent.*

Theorem. *A linear transformation is a projection operator if and only if it is idempotent.*

Proof. We have seen already that every projection is idempotent, so the problem is to prove that an idempotent linear transformation is a projection operator. Suppose that $P : V \to V$ is linear and idempotent. Define

$$U := \operatorname{Im} P, \qquad W := \operatorname{Ker} P.$$

Our first task is to prove that $V = U \oplus W$. Let $v \in V$. Then $v = Pv + (v - Pv)$. Now $Pv \in U$ (obviously), and $P(v - Pv) = Pv - P^2 v = 0$, so $v - Pv \in W$. Thus $V = U + W$. Now let $x \in U \cap W$. Since $x \in U$ there exists $y \in V$ such that $x = Py$, and since $x \in W$ also $Px = 0$. Then $x = Py = P^2 y = Px = 0$. Thus $U \cap W = \{0\}$, and so $V = U \oplus W$.

To finish the proof we need to convince ourselves (and others—such as examiners, tutors and other friends) that P is the projection onto U along W. For $v \in V$ write $v = u + w$ where $u \in U$ and $w \in W$. Since $u \in U$ there must exist $x \in V$ such that $u = Px$. Then

$$Pv = P(u + w) = Pu + Pw = P^2 x + 0 = Px = u\,,$$

and therefore P is indeed the projection onto U along W, as we predicted.

Exercise 11. Let V be a vector space (over some field F), and let E_1 and E_2 be projections on V.

(i) Show that $E_1 + E_2$ is a projection if and only if $E_1 E_2 + E_2 E_1 = 0$.

(ii) Prove that if $\operatorname{char} F \neq 2$ (that is if $1 + 1 \neq 0$ in F) then this happens if and only if $E_1 E_2 = E_2 E_1 = 0$. [*Hint:* Calculate $E_1 E_2 E_1$ in two different ways.]

(iii) Now suppose that $E_1 + E_2$ is a projection. Assuming that $\operatorname{char} F \neq 2$, find its kernel and image in terms of those of E_1 and E_2.

(iv) What can be said in (ii) and (iii) if $\operatorname{char} F = 2$?

The next result is a theorem which turns out to be very useful in many situations. It is particularly important in applications of linear algebra in Quantum Mechanics.

Theorem. *Let $P : V \to V$ be the projection onto U along W and let $T : V \to V$ be any linear transformation. Then $PT = TP$ if and only if U and W are T-invariant (that is $TU \leqslant U$ and $TW \leqslant W$).*

Proof. Suppose first that $PT = TP$. If $u \in U$, so that $Pu = u$, then $Tu = TPu = PTu \in U$, so $TU \leqslant U$. And if $w \in W$ then $P(Tw) = TPw = T0 = 0$, and therefore $TW \leqslant W$.

Now suppose conversely that U and W are T-invariant. Let $v \in V$ and write $v = u + w$ with $u \in U$ and $w \in W$. Then

$$PTv = PT(u + w) = P(Tu + Tw) = Tu\,,$$

since $Tu \in U$ and $Tw \in W$. Also,

$$TPv = TP(u + w) = Tu\,.$$

Thus $PTv = TPv$ for all $v \in V$ and therefore $PT = TP$, as asserted.

We turn now to direct sums of more than two subspaces. The vector space V is said to be the *direct sum* of subspaces $U_1, \ldots, U_k$ if for every $v \in V$ there exist unique vectors $u_i \in U_i$ for $1 \leqslant i \leqslant k$ such that $v = u_1 + \cdots + u_k$. We write $V = U_1 \oplus \cdots \oplus U_k$.

NOTE 1: If $k = 2$ then this reduces to exactly the same concept as we have just been studying. Moreover, if $k > 2$ then $U_1 \oplus U_2 \oplus \cdots \oplus U_k = (\cdots((U_1 \oplus U_2) \oplus U_3) \oplus \cdots \oplus U_k)$.

NOTE 2: If $U_i \leqslant V$ for $1 \leqslant i \leqslant k$ then $V = U_1 \oplus \cdots \oplus U_k$ if and only if $V = U_1 + U_2 + \cdots + U_k$ and $U_r \cap (\sum_{i \neq r} U_i) = \{0\}$ for $1 \leqslant r \leqslant k$. It is **NOT** sufficient that $U_i \cap U_j = \{0\}$ whenever $i \neq j$. Consider, for example, the 2-dimensional space F^2 of pairs (x_1, x_2) with $x_1, x_2 \in F$. Its three subspaces

$$U_1 := \{(x,0) \mid x \in F\}, \quad U_2 := \{(0,x) \mid x \in F\}, \quad U_3 := \{(x,x) \mid x \in F\}$$

satisfy

$$U_1 \cap U_2 = U_1 \cap U_3 = U_2 \cap U_3 = \{0\}$$

and yet it is clearly not true that F^2 is their direct sum.

NOTE 3: If $V = U_1 \oplus U_2 \oplus \cdots \oplus U_k$ and B_i is a basis of U_i then $B_1 \cup B_2 \cup \cdots \cup B_k$ is a basis of V. In particular, $\dim V = \sum_{i=1}^{k} \dim U_i$. The proof, which is not deep, is offered as an exercise:

EXERCISE 12. Prove that if $V = U_1 \oplus U_2 \oplus \cdots \oplus U_k$ and B_i is a basis of U_i then $B_1 \cup B_2 \cup \cdots \cup B_k$ is a basis of V.

Let $P_1, \ldots, P_k$ be linear mappings $V \to V$ such that $P_i^2 = P_i$ for all i and $P_i P_j = 0$ whenever $i \neq j$. If $P_1 + \cdots + P_k = I$ then $\{P_1, \ldots, P_k\}$ is known as a *partition of the identity* on V.

EXAMPLE: If P is a projection then $\{P, I - P\}$ is a partition of the identity.

THEOREM. *If $V = U_1 \oplus \cdots \oplus U_k$ and P_i is the projection of V onto U_i along $\bigoplus_{j \neq i} U_j$ then $\{P_1, \ldots, P_k\}$ is a partition of the identity on V.*

Conversely, if $\{P_1, \ldots, P_k\}$ is a partition of the identity on V and $U_i := \operatorname{Im} P_i$ then $V = U_1 \oplus \cdots \oplus U_k$.

Proof. Suppose first that $V = U_1 \oplus \cdots \oplus U_k$. Let P_i be the projection of V onto U_i along $\bigoplus_{j \neq i} U_j$. Certainly $P_i^2 = P_i$, and if $i \neq j$ then $\operatorname{Im} P_j \leqslant \operatorname{Ker} P_i$ so $P_i P_j = 0$. If $v \in V$ then there are uniquely determined vectors $u_i \in U_i$ for $1 \leqslant i \leqslant k$ such that $v = u_1 + \cdots + u_k$. Then $P_i v = u_i$ by definition of what it means for P_i to be projection of V onto U_i along $\bigoplus_{j \neq i} U_j$. Therefore

$$I v = v = P_1 v + \cdots + P_k v = (P_1 + \cdots + P_k) v\,.$$

Since this equation holds for all $v \in V$ we have $I = P_1 + \cdots + P_k$. Thus $\{P_1, \ldots, P_k\}$ is a partition of the identity.

To understand the converse, let $\{P_1, \ldots, P_k\}$ be a partition of the identity on V and let $U_i := \operatorname{Im} P_i$. For $v \in V$, defining $u_i := P_i v$ we have

$$v = I v = (P_1 + \cdots + P_k) v = P_1 v + \cdots + P_k v = u_1 + \cdots + u_k\,.$$

Suppose that $v = w_1 + \cdots + w_k$ where $w_i \in U_i$ for $1 \leqslant i \leqslant k$. Then $P_i\, w_i = w_i$ since P_i is a projection onto U_i. And if $j \neq i$ then $P_j\, w_i = P_j\,(P_i w_i) = (P_j\, P_i)\, w_i$, so $P_j\, w_i = 0$ since $P_j\, P_i = 0$. Therefore

$$P_i\, v = P_i\,(w_1 + \cdots + w_k) = P_i\, w_1 + \cdots + P_i\, w_k = w_i\,,$$

that is, $w_i = u_i$. This shows the uniqueness of the decomposition $v = u_1 + \cdots + u_k$ with $u_i \in U_i$ and so $V = U_1 \oplus \cdots \oplus U_k$, and the proof is complete.

Linear functionals and dual spaces

A *linear functional* on the vector space V over the field F is a function $f : V \to F$ such that

$$f(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 f(v_1) + \alpha_2 f(v_2)$$

for all $\alpha_1, \alpha_2 \in F$ and all $v_1, v_2 \in V$.

NOTE: A linear functional, then, is a linear transformation $V \to F$, where F is construed as a 1-dimensional vector space over itself.

EXAMPLE. If $V = F^n$ (column vectors) and y is a $1 \times n$ row vector then the map $v \mapsto y\,v$ is a linear functional on V.

The dual space V' of V is defined as follows:

$$\begin{aligned} \text{Set} &:= \text{set of linear functionals on } V \\ 0 &:= \text{zero function} && [v \mapsto 0 \text{ for all } v \in V] \\ (-f)(v) &:= -(f(v)) \\ (f_1 + f_2)(v) &:= f_1(v) + f_2(v) && [\text{pointwise addition}] \\ (\lambda f)(v) &:= \lambda f(v) && [\text{pointwise multiplication by scalars}] \end{aligned}$$

NOTE: It is a matter of important routine to check that the vector space axioms are satisfied (see the exercise below). It is also important that, when invited (for example by an examiner) to define the dual space of a vector space, you specify not only the set, but also the operations which make that set into a vector space.

EXERCISE 13 (worth doing carefully once in one's life, but not more than once—unless an examiner offers marks for it). Check that the vector space axioms are satisfied, so that V' defined as above really is a vector space over F.

NOTE: Some authors use V^* or $\mathrm{Hom}(V, F)$ or $\mathrm{Hom}_F\,(V, F)$ for the dual space V'.

THEOREM. *Let V be a finite-dimensional vector space over a field F. For every basis $v_1, v_2, \ldots, v_n$ of V there is a basis $f_1, f_2, \ldots, f_n$ of V' such that*

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

In particular, $\dim V' = \dim V$.

Proof. Define f_i as follows. For $v \in V$ we set $f_i(v) := \alpha_i$ where $\alpha_1, \dots, \alpha_n \in F$ are such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. This definition is acceptable because $v_1, \dots, v_n$ span V and so such scalars $\alpha_1, \dots, \alpha_n$ certainly exist; moreover, since $v_1, \dots, v_n$ are linearly independent the coefficients $\alpha_1, \dots, \alpha_n$ are uniquely determined by v. If $w \in V$, say $w = \beta_1 v_1 + \cdots + \beta_n v_n$, and $\lambda, \mu \in F$ then

$$\begin{aligned} f_i(\lambda v + \mu w) &= f_i\Big(\lambda \sum_j \alpha_j v_j + \mu \sum_j \beta_j v_j\Big) \\ &= f_i\Big(\sum_j (\lambda \alpha_j + \mu \beta_j) v_j\Big) \\ &= \lambda \alpha_i + \mu \beta_i \\ &= \lambda f_i(v) + \mu f_i(w), \end{aligned}$$

and so $f_i \in V'$. We have thus found elements $f_1, \dots, f_n$ of V' such that

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

To finish the proof we must show that they form a basis of V'.

To see that they are independent suppose that $\sum \mu_j f_j = 0$, where $\mu_1, \dots, \mu_n \in F$. Evaluate at v_i:

$$0 = \Big(\sum_j \mu_j f_j\Big)(v_i) = \sum_j \mu_j f_j(v_i) = \mu_i .$$

Thus $\mu_1 = \cdots = \mu_n = 0$ and so $f_1, \dots, f_n$ are linearly independent.

To see that they span V' let $f \in V'$ and define $g := \sum_j f(v_j) f_j$. Then also $g \in V'$ and for $1 \leqslant i \leqslant n$ we have

$$g(v_i) = \Big(\sum_j f(v_j) f_j\Big)(v_i) = \sum_j f(v_j) f_j(v_i) = f(v_i) .$$

Since f and g are linear and agree on a basis of V we have $f = g$, that is, $f = \sum_j f(v_j) f_j$. Thus $f_1, \dots, f_n$ is indeed a basis of V', as the theorem states.

Note. The basis $f_1, f_2, \dots, f_n$ is known as the *dual basis* of $v_1, v_2, \dots, v_n$. Clearly, it is uniquely determined by this basis of V.

Example. If $V = F^n$ ($n \times 1$ column vectors) then we may identify V' with the space of $1 \times n$ row vectors. The canonical basis $e_1, e_2, \dots, e_n$ then has dual basis $e'_1, e'_2, \dots, e'_n$, the canonical basis of the space of row vectors.

Exercise 14. Let F be a field with at least 4 elements and let V be the vector space of polynomials $c_0 + c_1 x + c_2 x^2 + c_3 x^3$ of degree $\leqslant 3$ with coefficients from F.

(i) Show that for $a \in F$ the map $f_a : V \to F$ given by evaluation of polynomial p at a (that is, $f_a(p) = p(a)$) is a linear functional.

(ii) Show that if a_1, a_2, a_3, a_4 are distinct members of F then $\{f_{a_1}, f_{a_2}, f_{a_3}, f_{a_4}\}$ is a basis of V', and find the basis $\{p_1, p_2, p_3, p_4\}$ of V of which this is the dual basis.

(ii) Generalise to the vector space of polynomials of degree $\leqslant n$ over F.

Let V be a vector space over the field F. For a subset X of V the *annihilator* is defined by

$$X^\circ := \{f \in V' \mid f(x) = 0 \text{ for all } x \in X\}.$$

NOTE: For any subset X the annihilator X° is a subspace. For, if $f_1, f_2 \in X^\circ$ and $\alpha_1, \alpha_2 \in F$ then for any $x \in X$

$$(\alpha_1 f_1 + \alpha_2 f_2)(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x) = 0 + 0 = 0,$$

and so $\alpha_1 f_1 + \alpha_2 f_2 \in X^\circ$.

NOTE: $X^\circ = \{f \in V' \mid X \subseteq \operatorname{Ker} f\}$.

THEOREM. *Let V be a finite-dimensional vector space over a field F and let U be a subspace. Then*

$$\dim U + \dim U^\circ = \dim V.$$

Proof. Let $u_1, \ldots, u_m$ be a basis for U and extend it to a basis $u_1, \ldots, u_m$, $u_{m+1}, \ldots, u_n$ for V. Thus $\dim U = m$ and $\dim V = n$. Let $f_1, \ldots, f_n$ be the dual basis of V'. We'll prove that $f_{m+1}, \ldots, f_n$ is a basis of U°. Certainly, if $m+1 \leqslant j \leqslant n$ then $f_j(u_i) = 0$ for $i \leqslant m$ and so $f_j \in U^\circ$ since $u_1, \ldots, u_m$ span U. Now let $f \in U^\circ$. There exist $\alpha_1, \ldots, \alpha_n \in F$ such that $f = \sum_j \alpha_j f_j$. Then

$$f(u_i) = \sum_j \alpha_j f_j(u_i) = \alpha_i,$$

and so $\alpha_i = 0$ for $1 \leqslant i \leqslant m$, that is, f is a linear combination of $f_{m+1}, \ldots, f_n$. Thus $f_{m+1}, \ldots, f_n$ span U° and so they form a basis of it. Therefore $\dim U^\circ = n - m$, that is, $\dim U + \dim U^\circ = \dim V$.

A WORKED EXAMPLE—part of an old FHS question (and useful to know). Let V be a finite-dimensional vector space over a field F. Show that if U_1, U_2 are subspaces then $(U_1 + U_2)^\circ = U_1^\circ \cap U_2^\circ$ and $(U_1 \cap U_2)^\circ = U_1^\circ + U_2^\circ$.

Response. If $f \in U_1^\circ \cap U_2^\circ$ then $f(u_1 + u_2) = f(u_1) + f(u_2) = 0 + 0 = 0$ for any $u_1 \in U_1$ and any $u_2 \in U_2$. Therefore $U_1^\circ \cap U_2^\circ \subseteq (U_1 + U_2)^\circ$. On the other hand, $U_1 \subseteq U_1 + U_2$ and $U_2 \subseteq U_1 + U_2$ and so if $f \in (U_1 + U_2)^\circ$ then $f \in U_1^\circ \cap U_2^\circ$, that is $(U_1 \cap U_2)^\circ \subseteq U_1^\circ + U_2^\circ$. Therefore in fact $(U_1 \cap U_2)^\circ = U_1^\circ + U_2^\circ$. Note that the assumption that V is finite-dimensional is not needed here.

For the second part, clearly $U_1^\circ \subseteq (U_1 \cap U_2)^\circ$ and $U_2^\circ \subseteq (U_1 \cap U_2)^\circ$ and so $U_1^\circ + U_2^\circ \subseteq (U_1 \cap U_2)^\circ$. Now we compare dimensions:

$$\begin{aligned}
\dim(U_1^\circ + U_2^\circ) &= \dim U_1^\circ + \dim U_2^\circ - \dim(U_1^\circ \cap U_2^\circ) \\
&= \dim U_1^\circ + \dim U_2^\circ - \dim(U_1 + U_2)^\circ \qquad \text{[by the above]} \\
&= (\dim V - \dim U_1) + (\dim V - \dim U_2) - (\dim V - \dim(U_1 + U_2)) \\
&= \dim V - (\dim U_1 + \dim U_2 - \dim(U_1 + U_2)) \\
&= \dim V - \dim(U_1 \cap U_2) \\
&= \dim(U_1 \cap U_2)^\circ.
\end{aligned}$$

Therefore $U_1^\circ + U_2^\circ = (U_1 \cap U_2)^\circ$, as required.

To finish this study of dual spaces we examine the second dual, that is, the dual of the dual. It turns out that if V is a finite-dimensional vector space then the second dual V'' can be naturally identified with V itself.

Theorem. *Let V be a vector space over a field F. Define $\Phi : V \to V''$ by $(\Phi\, v)(f) := f(v)$ for all $v \in V$ and all $f \in V'$. Then Φ is linear and one-one [injective]. If V is finite-dimensional then Φ is an isomorphism.*

Proof. We check linearity as follows. For $v_1, v_2 \in V$ and $\alpha_1, \alpha_2 \in F$, and for any $f \in V'$,

$$\begin{aligned}
\Phi(\alpha_1 v_1 + \alpha_2 v_2)(f) &= f(\alpha_1 v_1 + \alpha_2 v_2) \\
&= \alpha_1 f(v_1) + \alpha_2 f(v_2) \\
&= \alpha_1 (\Phi\, v_1)(f) + \alpha_2 (\Phi\, v_2)(f) \\
&= (\alpha_1 (\Phi\, v_1) + \alpha_2 (\Phi\, v_2))(f),
\end{aligned}$$

and so $\Phi(\alpha_1 v_1 + \alpha_2 v_2)\alpha_1 (\Phi\, v_1) + \alpha_2 (\Phi\, v_2)$.

Now

$$\begin{aligned}
\operatorname{Ker} \Phi &= \{v \in V \mid \Phi\, v = 0\} \\
&= \{v \in V \mid f(v) = 0 \text{ for all } f \in V'\} \\
&= \{0\}
\end{aligned}$$

and therefore Φ is injective. If V is finite-dimensional then

$$\dim \operatorname{Im}(\Phi) = \dim V = \dim V' = \dim V''$$

and so Φ is also surjective, that is, it is an isomorphism.

Exercise 15. Let V be a finite-dimensional vector space over a field F. If $Y \subseteq V'$ we'll use Y° to denote $\{v \in V \mid f(v) = 0 \text{ for all } f \in Y\}$. Prove that if $U \leqslant V$ (that is, U is a sub*space* of V) then $(U^\circ)^\circ = U$. Prove also that if $X \subseteq V$ (that is, X is a sub*set* of V) then $(X^\circ)^\circ$ is the subspace $\langle X \rangle_F$ spanned by X.

Dual transformations

Let V and W be vector spaces over a field F, and let $T : V \to W$ be a linear transformation. The *dual* transformation $T' : W' \to V'$ is defined by

$$T'(f) := f \circ T \textit{ for all } f \in W'.$$

Note that $T'f$ is often written for $T'(f)$. Thus

$$(T'f)(v) = f(Tv) \textit{ for all } v \in V\ .$$

Fact. *This specification does define a map $T' : W' \to V'$, which, moreover, is linear.*

Proof. We need first to show that if $f \in W'$ then $T'f \in V'$. But if $f \in W'$ then $f : W \to F$ and f is linear, so, since $T : V \to W$ is linear and composition of linear maps produces a linear map, also $f \circ T : V \to F$ is linear, that is, $T'f \in V'$. Now let $f_1, f_2 \in W'$ and $\alpha_1, \alpha_2 \in F$. Then for any $v \in V$ we have

$$\begin{aligned} T'(\alpha_1 f_1 + \alpha_2 f_2)(v) &= (\alpha_1 f_1 + \alpha_2 f_2)(Tv) \\ &= \alpha_1 f_1(Tv) + \alpha_2 f_2(Tv) \\ &= \alpha_1 T' f_1(v) + \alpha_2 T' f_2(v) \\ &= (\alpha_1 T' f_1 + \alpha_2 T' f_2)(v) \end{aligned}$$

and so $T'(\alpha_1 f_1 + \alpha_2 f_2) = (\alpha_1 T' f_1 + \alpha_2 T' f_2)$, that is, $T' : W' \to V'$ is linear.

Observations. *Let V and W be vector spaces over a field F, and let $T : V \to W$ be a linear transformation.*

(1) *If $V = W$ and $T = I$ then also $T' = I$.*

(2) *If $S : U \to V$ and $T : V \to W$ then $(T \circ S)' = S' \circ T'$.*

(3) *Hence if T is invertible then $(T^{-1})' = (T')^{-1}$.*

(4) *If $T_1, T_2 : V \to W$ and $\alpha_1, \alpha_2 \in F$ then $(\alpha_1 T_1 + \alpha_2 T_2)' = \alpha_1 T_1' + \alpha_2 T_2'$.*

(5) *Hence if $T : V \to V$ and $f(x) \in F[x]$ then $f(T)' = f(T')$*

Proofs. Clause (1) is immediate from the definition since $I'f = f \circ I = f$ for all $f \in W'$. Clauses (2), (3), (4) are straightforward routine and should be done as an exercise (see below). For (5) we need to know what is meant by $f(T)$ when f is a polynomial with coefficients from F and T is a linear transformation $V \to V$. It means precisely what you would expect it to mean: non-negative integral powers of T are defined by $T^0 := I$, $T^{n+1} := T \circ T^n$ for $n \geqslant 0$, and then if $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ then $f(T)$ is the corresponding linear combination $a_0 I + a_1 T + \cdots + a_n T^n$ of the powers of T. The fact that $f(T)' = f(T')$ therefore follows from (1), (2) and (4) by induction on the degree of f.

Exercise 16. Let U, V and W be vector spaces over a field F, and let $S : U \to V$, $T : V \to W$ be linear transformations. Show that $(T S)' = S' T'$. Deduce that if $T : V \to V$ is an invertible linear transformation then $(T^{-1})' = (T')^{-1}$.

Exercise 17. Let V and W be vector spaces over a field F and let $T_1, T_2 : V \to W$ be linear transformations. Show that if $\alpha_1, \alpha_2 \in F$ then $(\alpha_1 T_1 + \alpha_2 T_2)' = \alpha_1 T_1' + \alpha_2 T_2'$

We ask now what can be said about the matrix of a dual transformation with respect to suitable bases in W' and V'.

Theorem. *Suppose that V, W are finite-dimensional vector spaces over F and $T : V \to W$ is linear. Let $v_1, v_2, \ldots, v_m$ be a basis of V, and $w_1, w_2, \ldots, w_n$ a basis of W. Let A be the matrix of T with respect to these bases; let B be the matrix of the dual transformation T' with respect to the dual bases of W', V'. Then $B = A^{\mathrm{tr}}$.*

Proof. Let $f_1, \dots, f_m$ and $h_1, \dots, h_n$ be the relevant dual bases in V' and W' respectively, so that

$$f_p(v_i) = \begin{cases} 1 & \text{if } p = i\,, \\ 0 & \text{if } p \neq i\,, \end{cases} \qquad\qquad h_q(w_j) = \begin{cases} 1 & \text{if } q = j\,, \\ 0 & \text{if } q \neq j\,. \end{cases}$$

Let $a_{i,j}$, $b_{p,q}$ be the (i,j)- and (p,q)-coefficients of A and B respectively. By definition then

$$T\,v_i = \sum_{j=1}^{n} a_{j,i}\,w_j \qquad \text{and} \qquad T'\,h_q = \sum_{p=1}^{m} b_{p,q}\,f_p\,.$$

Now

$$(T'\,h_q)(v_i) = h_q(T\,v_i) = h_q\Big(\sum_{j=1}^{n} a_{j,i}\,w_j\Big) = a_{q,i}\,.$$

It follows that $T'\,h_q = \sum_p a_{q,p}\,f'_p$ (compare the proof on p. 11 that the dual basis spans). Therefore $b_{p,q} = a_{q,p}$ for $1 \leqslant p \leqslant m$ and $1 \leqslant q \leqslant n$, that is, $B = A^{\mathrm{tr}}$ as the theorem states.

Corollary. *For a linear transformation $T : V \to W$ of finite-dimensional vector spaces over any field F we have* $\mathrm{rank} T' = \mathrm{rank} T$ *.*

For we know that a matrix and its transpose have the same rank. By studying the kernel and image of a dual transformation we get a more "geometric" understanding of the corollary, however:

Theorem. *Let $T : V \to W$ be a linear transformation of finite-dimensional vector spaces over a field F. Then*

$$\operatorname{Ker} T' = (\operatorname{Im} T)^\circ \qquad \textit{and} \qquad \operatorname{Im} T' = (\operatorname{Ker} T)^\circ\,.$$

Proof. The first equation does not depend on finite-dimensionality—it is true in complete generality:

$$\begin{aligned} \operatorname{Ker} T' &= \{f \in W' \mid T'\,f = 0\} = \{f \in W' \mid f \circ T = 0\} \\ &= \{f \in W' \mid f(\operatorname{Im} T) = \{0\}\} \\ &= (\operatorname{Im} T)^\circ. \end{aligned}$$

For the second, suppose that $f \in \operatorname{Im} T'$, say $f = T'\,g$, where $g \in W'$, and let $u \in \operatorname{Ker} T$. Then

$$f(u) = (T'\,g)(u) = g(T\,u) = g(0) = 0\,.$$

Thus $\operatorname{Im} T' \leqslant (\operatorname{Ker} T)^\circ$. Again, this is true quite generally. But to get equality we use finite-dimensionality:

$$\begin{aligned} \dim(\operatorname{Im} T') &= \dim W' - \dim(\operatorname{Ker} T') && \text{[Rank-Nullity Theorem]} \\ &= \dim W - \dim((\operatorname{Im} T)^\circ) && [\dim W' = \dim W \text{ and first part}] \\ &= \dim(\operatorname{Im} T) && \text{[theorem on dimension of annihilator]} \\ &= \dim V - \dim(\operatorname{Ker} T) && \text{[Rank-Nullity Theorem]} \\ &= \dim((\operatorname{Ker} T)^\circ) && \text{[theorem on dimension of annihilator]} \end{aligned}$$

and therefore $\operatorname{Im} T' = (\operatorname{Ker} T)^\circ$, as required.

A WORKED EXAMPLE—FHS 2000, Paper a1, Qn 1:

> Let V be a finite-dimensional vector space over $\mathbb{R}$ and let $P : V \to V$ be a linear transformation of V. Let $V_1 = \operatorname{Ker}(P)$ and $V_2 = \operatorname{Ker}(I_V - P)$, where $I_V : V \to V$ is the identity map on V, and suppose that $V = V_1 \oplus V_2$. Prove that $P^2 = P$.
>
> Define the *dual space* V' of V and the *dual transformation* P' of P. Show that $(P')^2 = P'$. Hence or otherwise show that $V' = U_1 \oplus U_2$ where $U_1 = \operatorname{Ker}(P')$ and $U_2 = \operatorname{Ker}(I_{V'} - P')$.
>
> Let $\mathcal{E}$ be a basis for V. Define the *dual basis* $\mathcal{E}'$ for V' and show that it is indeed a basis. Suppose that $\mathcal{E} \subseteq V_1 \cup V_2$. Show that $\mathcal{E}' \subseteq U_1 \cup U_2$, and describe the matrices of P and P' with respect to the bases $\mathcal{E}$ and $\mathcal{E}'$ respectively.

Response. Note that in fact it is irrelevant that the field of coefficients is $\mathbb{R}$, so we'll do this for vector spaces over an arbitrary field F.

For the first part let $v \in V$ and write $v = v_1 + v_2$ where $v_1 \in V_1$ and $v_2 \in V_2$. Note that $P v_2 = v_2$ since $(I_V - P) v_2 = 0$. Then $P v = P v_1 + P v_2 = 0 + v_2 = v_2$, and $P^2 v = P v_2 = v_2$. Thus $P^2 v = P v$ for all $v \in V$ and so $P^2 = P$.

Defining dual spaces and dual transformations is bookwork treated earlier in these notes. Then for $f \in V'$ we have that

$$(P')^2 f = P'(P'(f)) = P'(f \circ P) = (f \circ P) \circ P = f \circ P^2 = f \circ P = P' f$$

and so $(P')^2 = P'$.

The proof of the theorem on p. 7 above that idempotent operators are projections includes a proof that $V' = U_1 \oplus U_2$ where $U_1 = \operatorname{Ker}(P')$ and $U_2 = \operatorname{Ker}(I_{V'} - P')$ [and this bookwork is what the examiner would have expected candidates to expound here].

Definition of dual basis and proof that it really is a basis is bookwork treated earlier in these notes (see p. 10 above). Now suppose that $\mathcal{E} \subseteq V_1 \cup V_2$. Thus $\mathcal{E} = \{v_1, v_2, \ldots, v_n\}$, where we may suppose that $v_1, \ldots, v_k \in V_1$ and $v_{k+1}, \ldots, v_n \in V_2$. Let $f_1, \ldots, f_n$ be the dual basis $\mathcal{E}'$ of V', so that $f_i(v_j) = 0$ if $i \neq j$ and $f_i(v_i) = 1$. Consider an index i such that $1 \leqslant i \leqslant k$. Now $(P' f_i)(v_j) = f_i(P v_j)$, and so if $1 \leqslant j \leqslant k$ then $(P' f_i)(v_j) = 0$ since $P v_j = 0$, while if $k+1 \leqslant j \leqslant n$ then $(P' f_i)(v_j) = 0$ since $P v_j = v_j$ and $f_i(v_j) = 0$. Thus $(P' f_i)(v_j) = 0$ for all relevant j and therefore $P' f_i = 0$, that is, $f_i \in U_1$. Similarly, if $k + 1 \leqslant i \leqslant n$ then $f_i \in U_2$. That is, $\mathcal{E}' \subseteq U_1 \cup U_2$ as required.

And now it is clear that the matrix of P' with respect to $\mathcal{E}'$ is the same as the matrix of P with respect to $\mathcal{E}$, namely $\begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix}$, where $r := n - k$, I_r is the $r \times r$ identity matrix and the three entries 0 represent $k \times k$, $k \times r$ and $r \times k$ zero matrices respectively.

Exercise 18. Let V be an n-dimensional vector space over a field F.

(i) Let U be an m-dimensional subspace of V and let

$$B(U) := \{T : V \to V \mid T \text{ is linear and } T(U) \leqslant U\}.$$

Show that $B(U)$ is a subspace of the space $\mathrm{Hom}(V, V)$ of all linear transformations $V \to V$ and that $\dim B(U) = n^2 - m\,n + m^2$.

(ii) A *flag* in V is an increasing sequence $\{0\} = U_0 < U_1 < U_2 < \cdots < U_k = V$ of subspaces beginning with $\{0\}$ and ending with V itself. For such a flag $\mathcal{F}$ we define

$$B(\mathcal{F}) := \{T : V \to V \mid T \text{ is linear and } T(U_i) \leqslant U_i \text{ for } 1 \leqslant i \leqslant k\}.$$

Show that $B(\mathcal{F})$ is a subspace of $\mathrm{Hom}(V, V)$ and calculate its dimension in terms of n and the dimensions m_i of the subspaces U_i.

Exercise 19. Let V be a vector space over a field F such that $\mathrm{char}\, F \neq 2$, and let E_1, E_2, E_3 be idempotent linear transformations $V \to V$ such that $E_1 + E_2 + E_3 = I$, where $I : V \to V$ is the identity transformation. Show that $E_i\, E_j = 0$ when $i \neq j$ (that is, $\{E_1, E_2, E_3\}$ is a partition of the identity on V). [*Hint*: recall Exercise 11 on p. 8.]

Give an example of four idempotent operators E_1, E_2, E_3, E_4 operators on V such that $E_1 + E_2 + E_3 + E_4 = I$ but $\{E_1, E_2, E_3, E_4\}$ is not a partition of the identity on V.

Exercise 20. Let $F := \mathbb{Z}_2$, the field with just two elements 0 and 1, let

$$V := \{f \in F^{\mathbb{N}} \mid \mathrm{supp}(f) \text{ is finite}\},$$

the subspace of $F^{\mathbb{N}}$ defined in Exercise 4 (see p. 3). Thus V may be thought of as the vector space of sequences $(x_0, x_1, x_2, \ldots)$ where each coordinate x_i is 0 or 1 and all except finitely many of the coordinates are 0. For each subset $S \subseteq \mathbb{N}$ define $\varphi_S : V \to F$ by $\varphi_S(f) = \sum_{n \in S} f(n)$.

(i) Show that $\varphi_S \in V'$ for all $S \subseteq \mathbb{N}$.

(ii) Show that in fact $V' = \{\varphi_S \mid S \subseteq \mathbb{N}\}$. [*Hint*: for each $n \in \mathbb{N}$ let $e_n \in V$ be the function such that $e_n(n) = 1$ and $e_n(m) = 0$ when $m \neq n$; then, given $\varphi \in V'$ define S by $S := \{n \in \mathbb{N} \mid \varphi(e_n) = 1\}$ and seek to show that $\varphi = \varphi_S$.]

(iii) Show that V is countable but V' is uncountable.

Part II: Some theory of a single linear transformation on a finite-dimensional vector space

We turn now to studying linear transformations on a finite-dimensional vector space to itself. The beginnings of this theory treat eigenvalues and eigenvectors, the characteristic polynomial, the minimal polynomial, and the Cayley–Hamilton Theorem.

Throughout this part V is a finite-dimensional vector space over a field F and $T : V \to V$ is linear. If $v_1, \ldots, v_n$ is a basis of V then T corresponds to an $n \times n$ matrix A over F. For, if $1 \leqslant i \leqslant n$ then $T v_i$ can be expressed uniquely as a linear combination of the members of the basis, and if

$$T v_i = \sum_{j=1}^{n} a_{ji} v_j$$

then $A = (a_{ij})$. Note that A is the transpose of the array of coefficients that would be written out on the page if we did not use summation notation. That might seem a little artificial, but this convention ensures that if $S : V \to V$ and S has matrix B then $S \circ T$ has matrix AB. The point to remember is that with respect to a given basis of V there is a one-to-one correspondence between linear transformations $T : V \to V$ and $n \times n$ square matrices over F, where $n := \dim V$.

For most of this section we will use linear transformations or $n \times n$ matrices over F interchangeably, whichever is the more convenient for the job in hand. Therefore you are advised to recall your facility for calculating with matrices. Here is an exercise to help.

Exercise 21. Suppose that $n = p + q$. Partition $n \times n$ matrices over a field F into the form $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$, where $A \in \mathrm{M}_{p\times p}(F)$, $B \in \mathrm{M}_{p\times q}(F)$, $C \in \mathrm{M}_{q\times p}(F)$, $D \in \mathrm{M}_{q\times q}(F)$. Show that if X is partitioned as $\begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}$ and Y is partitioned as $\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix}$ then

$$XY = \begin{pmatrix} X_{11}Y_{11} + X_{12}Y_{21} & X_{11}Y_{12} + X_{12}Y_{22} \\ X_{21}Y_{11} + X_{22}Y_{21} & X_{21}Y_{12} + X_{22}Y_{22} \end{pmatrix}.$$

Determinants and traces

Recall from Mods the definitions of $\det A$ and $\operatorname{trace} A$ where A is a square matrix. If A is the $n \times n$ matrix (a_{ij}) then

$$\det A = \sum_{\rho \in \mathrm{Sym}\,(n)} (-1)^{\mathrm{parity}(\rho)}\, a_{1\,1^\rho}\, a_{2\,2^\rho} \cdots a_{n\,n^\rho} \qquad \text{and} \qquad \operatorname{trace} A = \sum_{i=1}^{n} a_{ii}\,,$$

where $\mathrm{parity}(\rho)$ is 0 or 1 according to whether ρ is an even or odd permutation, and i^ρ denotes the image of i under ρ. Thus $\det A$ is described as a sum of $n!$ terms, each of which is plus or minus the product of n coefficients of A, chosen in such a way that there is just one from each row and just one from each column. In Mods the coefficients a_{ij} were real (or perhaps complex) numbers. But for us they come from an arbitrary field F (they could even come from a commutative ring). The basic properties are the same, for example that $\det(AB) = (\det A)(\det B)$, that $\mathrm{trace}(AB) = \mathrm{trace}(BA)$, and that A is invertible (in the sense that there exists B such that $AB = BA = I$) if and only if $\det A \neq 0$.

Now for our linear transformation $T : V \to V$ we define the determinant of T by $\det T := \det A$, and we define the trace by $\mathrm{trace} T := \mathrm{trace} A$ where A is the matrix of T with respect to some basis of V. On the face of it these definitions might depend on the particular basis that is used—in which case they would be of doubtful value. But in fact they do not:

Observation. *Determinant and trace of T depend only on T and do not depend on the basis of V used to compute them.*

Proof. We know from Mods that if B is the matrix representing T with respect to another basis then $B = U^{-1}AU$ where U is the matrix whose entries are the coefficients needed to express the elements of one basis as linear combinations of the elements of the other. Therefore

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

and

$$\mathrm{trace} B = \mathrm{trace}((U^{-1}A)\,U) = \mathrm{trace}(U\,(U^{-1}A)) = \mathrm{trace} A,$$

as required.

The characteristic polynomial and the minimal polynomial

The *characteristic polynomial* of an $n \times n$ matrix A is defined by

$$c_A(x) := \det(xI - A)\,.$$

For our linear transformation $T : V \to V$, the characteristic polynomial of T is defined by $c_T(x) := c_A(x)$ where A represents T with respect to some basis of V. By what has just been shown, $c_T(x)$ is well defined—that is, it is independent of the basis used to calculate it.

Note: If $n := \dim V$ then $c_T(x)$ is a monic polynomial (*monic* means leading coefficient $= 1$), and it is of degree n. In fact

$$c_T(x) = x^n - c_1 x^{n-1} + c_2 x^{n-2} - \cdots + (-1)^n c_n\,,$$

where

$$c_1 = \mathrm{trace} T\,, \qquad c_n = \det T, \qquad \textit{etc.}$$

Here '*etc.*' hides a great deal of detailed information. The other coefficients c_r are important but more complicated functions of A—in fact c_r is the sum of the determinants of all the so-called $r \times r$ principal submatrices of A, that is, square submatrices of A of which the diagonals coincide with part of the diagonal of A. For example, c_2 is the sum of the determinants of all the $\frac{1}{2}n(n-1)$ submatrices $\begin{pmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{pmatrix}$ for $1 \leqslant i < j \leqslant n$.

Just as in the case of linear algebra over $\mathbb{R}$, we define eigenvalues and eigenvectors of T as follows: a scalar $\lambda \in F$ is said to be an *eigenvalue* of T if there exists a non-zero vector $v \in V$ such that $Tv = \lambda v$; and a vector $v \in V$ is said to be an *eigenvector* of T if $v \neq 0$ and there exists a scalar $\lambda \in F$ such that $Tv = \lambda v$.

Theorem. *The characteristic polynomial of a linear transformation $T : V \to V$ has the following properties:*

(1) *if $S = U^{-1}TU$, where $U : V \to V$ is linear and invertible, then $c_S(x) = c_T(x)$;*

(2) *a scalar $\lambda \in F$ is an eigenvalue of T if and only if $c_T(\lambda) = 0$;*

Proof. If $S = U^{-1}TU$ then $c_S(x) = \det(xI - U^{-1}TU) = \det(U^{-1}(xI - T)U) = \det(xI - T) = c_T(x)$, and this proves (1).

For (2), λ is an eigenvalue if and only if there exists $v \in V \setminus \{0\}$ such that $(\lambda I - T)v = 0$, that is, if and only if $\operatorname{Ker}(\lambda I - T) \neq \{0\}$. But we know that this holds if and only if $\lambda I - T$ is not invertible, that is, if and only if $\det(\lambda I - T) = 0$, so $c_T(\lambda) = 0$, as required.

We turn now to the so-called *minimal polynomial* of T. We saw on p. 14 how $f(T)$ is defined for a polynomial $f \in F[x]$: if $f(x) = a_0 + a_1 x + \cdots + a_k x^k \in F[x]$ then $f(T) := a_0 I + a_1 T + \cdots + a_k T^k$, where $I : V \to V$ is the identity map. If the matrix representing T with respect to a given basis of V is A, then the matrix representing $f(T)$ with respect to this basis will be $f(A)$.

Exercise 22. Show that if $U : V \to V$ is invertible and $S := U^{-1}TU$ then $f(S) = U^{-1}f(T)U$ for any polynomial $f \in F[x]$.

Now we come to an important definition. A monic polynomial $f \in F[x] \setminus \{0\}$ of least degree such that $f(T) = 0$ is known as the *minimal polynomial* of T. Similarly, for an $n \times n$ matrix A, a monic polynomial $f \in F[x] \setminus \{0\}$ of least degree such that $f(A) = 0$ is known as the minimal polynomial of A.

For these definitions to make sense it must be the case that there exist non-zero polynomials $f \in F[x]$ such that $f(T) = 0$, or $f(A) = 0$ respectively. As preparation for the proof of this we need a lemma:

Lemma. *Let $n := \dim V$. The set of all linear transformations $S : V \to V$ forms a vector space of dimension n^2 over F.*

Proof. That the set of all linear transformations $V \to V$ forms a vector space should be clear since we can add linear transformations and multiply them by scalars, and the vector space axioms can easily be checked. The correspondence of linear transformations with matrices is obviously a vector-space isomorphism, and the space of $n \times n$ matrices has dimension n^2 since the matrices E_{pq}, where E_{pq} has 1 as its (p, q) entry and 0 elsewhere, form a basis.

Corollary. *There is a polynomial $f \in F[x] \setminus \{0\}$ such that $f(T) = 0$. Similarly for $n \times n$ matrices over F.*

Proof. Since the set of all linear transformations $V \to V$ forms a vector space of dimension n^2, the transformations I, T, T^2, $\ldots$, T^{n^2}, of which there are n^2+1, must be linearly dependent. Therefore for $0 \leqslant i \leqslant n^2$ there exist $c_i \in F$, not all 0, such that $\sum_{i=0}^{n^2} c_i T^i = 0$. So if $f(x) := \sum_{i=0}^{n^2} c_i x^i$ then $f(x) \in F[x] \setminus \{0\}$ and $f(T) = 0$.

It follows immediately that the minimal polynomial of our linear transformation $T : V \to V$ or of an $n \times n$ matrix A over F is well-defined.

Observation. *The minimal polynomial of T or of an $n \times n$ matrix A over F is unique.*

For, if $f_1, f_2 \in F[x]$ are minimal polynomials of T (or of A), then f_1, f_2 are monic and of the same degree, say m. Therefore if $g := f_1 - f_2$ then either $g = 0$ or $\deg g < m$. But $g(T) = f_1(T) - f_2(T) = 0$, and so since m is the least degree of non-zero polynomials which annihilate T, it must be that $g = 0$: that is, $f_1 = f_2$.

Notation. We'll write $m_T(x)$ (or $m_A(x)$ when A is an $n \times n$ matrix over F) for the minimal polynomial of T (or of A). Note that if $A \in \mathrm{M}_{n \times n}(F)$ and A represents T with respect to some basis of V then $m_T(x) = m_A(x)$.

Observation. *If $S = U^{-1}TU$, where $U : V \to V$ is linear and invertible, then $m_S(x) = m_T(x)$.*

For, if $f \in F[x]$, then $f(S) = U^{-1} f(T) U$ (see Exercise 21), and so $f(S) = 0$ if and only if $f(T) = 0$.

Examples:

- $m_I(x) = x - 1$;
- $m_0(x) = x$;
- if $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ then $m_A(x) = (x^2 - 3x + 2)$.

Exercise 23. Let $T : V \to V$ be a linear transformation of a finite-dimensional vector space over a field F. Show that T is invertible if and only if the constant term of $m_T(x)$ is non-zero.

Theorem. *For $f \in F[x]$, $f(T) = 0$ if and only if $m_T(x)$ divides $f(x)$ in $F[x]$. Similarly for $A \in \mathrm{M}_{n \times n}(F)$.*

Proof. Let $f \in F[x]$. Since $F[x]$ has a Division Algorithm we can find $q, r \in F[x]$ such that $f(x) = q(x)\, m_T(x) + r(x)$ and either $r = 0$ or $\deg r < \deg m_T$. Now $m_T(T) = 0$ and so $f(T) = 0$ if and only if $r(T) = 0$. It follows from the minimality of $\deg m_T$ that $f(T) = 0$ if and only if $r = 0$. That is $f(T) = 0$ if and only if $m_T(x)$ divides $f(x)$ in $F[x]$, as required.

Exercise 24. Let $\mathrm{Ann}(T) := \{f \in F[x] \mid f(T) = 0\}$, the so-called *annihilator* of T in $F[x]$. Show that $\mathrm{Ann}(T)$ is an ideal in $F[x]$, and that m_T is a generator—that is, $\mathrm{Ann}(T)$ is the principal ideal (m_T) in $F[x]$.

Next we examine the roots of the minimal polynomial: it turns out that they are precisely the eigenvalues of T in F:

Theorem. *For $\lambda \in F$, $m_T(\lambda) = 0$ if and only if $c_T(\lambda) = 0$. Similarly for $A \in \mathrm{M}_{n\times n}(F)$.*

Proof. Suppose first that $c_T(\lambda) = 0$, so that λ is an eigenvalue. Let v be a non-zero vector such that $T v = \lambda v$. A simple inductive argument shows that $T^n v = \lambda^n v$ for all $n \geqslant 0$, and then forming linear combinations we see that $f(T) v = f(\lambda) v$ for any polynomial $f \in F[x]$. In particular, $m_T(\lambda) v = m_T(T) v = 0$. But $v \neq 0$ and so $m_T(\lambda) = 0$.

For the converse, let $\lambda \in F$ and suppose that $m_T(\lambda) = 0$. We know then that there exists $g \in F[x]$ such that $m_T(x) = (x - \lambda) g(x)$. Now $\deg g < \deg m$ and so $g(T) \neq 0$. Therefore there exists $w \in V$ such that $g(T) w \neq 0$. Define $v := g(T) w$. Then $v \neq 0$ and $(T - \lambda I)v = (T - \lambda I) g(T) w = m_T(T) w = 0$, so λ is an eigenvalue of T, that is, $c_T(\lambda) = 0$.

Example. Let $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. Then

$$\begin{aligned} c_A(x) &= (x-1)(x-2)^2 \\ m_A(x) &= (x-1)(x-2)\,. \end{aligned}$$

Note. In fact $m_T(x)$ and $c_T(x)$ always have the same irreducible factors in $F[x]$.

Exercise 25. For each of the matrices $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 1 & 0 & 1 \\ -2 & -1 & -1 & 0 \\ 0 & 0 & 2 & -5 \\ 0 & 0 & 1 & -2 \end{pmatrix}$ find the characteristic polynomial and the minimal polynomial.

Exercise 26. [Part of a former FHS question.] (i) Let A be a 3×3 matrix whose characteristic polynomial is x^3. Show that there are exactly three possibilities for the minimal polynomial of A and give an example of matrices of each type.

(ii) Let V be a finite-dimensional vector space over some field F, and let $T : V \to V$ be a linear transformation whose minimal polynomial is x^k. Prove that

$$\{0\} < \operatorname{Ker} T < \operatorname{Ker} T^2 < \cdots < \operatorname{Ker} T^k = V$$

and deduce that $\dim V \geqslant k$.

The Primary Decomposition Theorem

The Primary Decomposition Theorem is a very useful result for understanding linear transformations and square matrices. Like many of the best results in mathematics it is more a theory than a single cut and dried theorem and I propose to present three forms of it. The first is the basic, all-purpose model which contains the main idea; the second gives some detail about the minimal polynomial; and the third is designed to extract a considerable amount of detail from the prime factorisation of the minimal polynomial.

Throughout this section notation is as before: F is a field, V is a finite-dimensional vector space over F, and $T : V \to V$ is linear. Recall that a subspace U is said to be T*-invariant* if $T(U) \leqslant U$, that is, $T u \in U$ for all $u \in U$. When this is the case we write $T|_U$ for the restriction of T to U. Thus $T|_U : U \to U$ and $T_U\, u = T\, u$ for all $u \in U$.

Reminder: although two linear operators do not usually commute, if they are polynomials in one and the same operator T then they certainly do commute. For, powers of T obviously commute and therefore so do linear combinations of powers of T.

Primary Decomposition Theorem (Mark 1). *Suppose that* $f(T) = 0$*, where* $f \in F[x]$*. Suppose also that* $f(x) = g(x)\, h(x)$*, where* $g, h \in F[x]$ *and* g*,* h *are coprime. Then there are* T*-invariant subspaces* U, W *of* V *such that* $V = U \oplus W$ *and* $g(T|_U) = 0$*,* $h(T|_W) = 0$*.*

Proof. Our problem is to find subspaces U and W of V that have the specified properties. Those properties include that $g(T)\, u = 0$ for all $u \in U$ and $h(T)\, w = 0$ for all $w \in W$. Therefore we know that we must seek U inside $\operatorname{Ker} g(T)$ and W inside $\operatorname{Ker} h(T)$. In fact we define

$$U := \operatorname{Ker} g(T) \quad \text{and} \quad W := \operatorname{Ker} h(T)$$

and prove that these subspaces do what is wanted. Certainly, if $u \in U$ then $g(T)(T\, u) = T\, g(T)\, u = T\, 0 = 0$. Thus if $u \in U$ then $T\, u \in U$, so U is T-invariant. Similarly, W is T-invariant. And the facts that $g(T|_U) = 0$ and $h(T|_W) = 0$ are immediate from the definitions of U and of W. It remains to prove therefore that $V = U \oplus W$.

From the theory of polynomials rings over a field we know that, since g, h are coprime, there exist $a, b \in F[x]$ such that

$$a(x)\, g(x) + b(x)\, h(x) = 1.$$

Then

$$a(T)\, g(T) + b(T)\, h(T) = I,$$

where $I : V \to V$ is the identity as usual. For $v \in V$ define

$$u := b(T)\, h(T)\, v \quad \text{and} \quad w := a(T)\, g(T)\, v\,.$$

Then $v = u + w$. Moreover, $g(T)\, u = g(T)\, b(T)\, h(T)\, u = b(T)\, f(T)\, u = 0$, and so $u \in U$. Similarly $w \in W$. Thus $V = U + W$. Now let $v \in U \cap W$. Then

$$v = a(T)\, g(T)\, v + b(T)\, h(T)\, v = a(T)\, 0 + b(T)\, 0 = 0\,.$$

Thus $U \cap W = \{0\}$ and $V = U \oplus W$ as required.

Exercise 27. With the notation and assumptions of the Primary Decomposition Theorem, let P be the projection of V onto U along W. Find $p(x) \in F[x]$ such that $P = p(T)$.

Primary Decomposition Theorem (Mark 2). *Suppose that* $m_T(x) = g(x)h(x)$ *where* $g, h \in F[x]$ *are monic and co-prime. Let* U*,* W *be as in the previous theorem. Then* $m_{T|_U} = g$ *and* $m_{T|_W} = h$ *.*

Proof. Define $f(x) := m_{T|_U}(x) \times m_{T|_W}(x)$. For $v \in V$ write $v = u + w$, where $u \in U$ and $w \in W$. Then

$$\begin{aligned} f(T)\,v &= m_{T|_U}(T)\,m_{T|_W}(T)\,(u+w) \\ &= m_{T|_W}(T)\,m_{T|_U}(T)\,u + m_{T|_U}(T)\,m_{T|_W}(T)\,w \\ &= 0 + 0 = 0\,. \end{aligned}$$

Thus $f(T)\,v = 0$ for all $v \in V$, that is $f(T) = 0$. Therefore $m_T(x)$ divides $f(x)$ in $F[x]$.

Since $g(T|_U) = 0$ we know that $m_{T|_U}(x)$ divides $g(x)$ in $F[x]$. Similarly $m_{T|_W}(x)$ divides $h(x)$ in $F[x]$. Therefore $f(x)$ divides $g(x)\,h(x)$ in $F[x]$, that is $f(x)$ divides $m_T(x)$ in $F[x]$. It follows that $m_T(x) = c\,f(x)$ for some non-zero $c \in F$. Since $m_T(x)$ and $f(x)$ are both monic, in fact $c = 1$, that is $m_T(x) = f(x)$. And now, if we write $g(x) = c_1(x)\,m_{T|_U}(x)$ and $h(x) = c_2(x)\,m_{T|_W}(x)$, where $c_1, c_2 \in F[x]$, then we see that $c_1(x)\,c_2(x) = 1$, so both c_1 and c_2 are constant polynomials. Since all of g, h, $m_{T|_U}$, $m_{T|_W}$ are monic, in fact $c_1 = c_2 = 1$. Thus $m_{T|_U} = g$ and $m_{T|_W} = h$, as required.

Example. If $m_T(x) = x^2 - x$ then (as we already know) there exist $U, W \leqslant V$ such that $V = U \oplus W$, $T|_U = I_U$ and $T|_W = 0_W$.

Primary Decomposition Theorem (Mark 3). *If*

$$m_T(x) = f_1(x)^{a_1}\,f_2(x)^{a_2}\;\cdots\;f_k(x)^{a_k}\,,$$

where f_1, f_2, …, f_k *are distinct monic irreducible polynomials over* F*, then*

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k\,,$$

where V_1, V_2, …, V_k *are* T*-invariant subspaces and the minimal polynomial of* $T|_{V_i}$ *is* $f_i^{a_i}$ *for* $1 \leqslant i \leqslant k$*.*

Proof. We use induction on k. The result is trivially true if $k = 1$, that is, if $m_T(x)$ is simply a power of some irreducible polynomial. Our inductive hypothesis is that if U is a finite-dimensional vector space over F and if the minimal polynomial of $S : U \to U$ factorises as a product of $k - 1$ powers of irreducible polynomials, then U decomposes as a direct sum as described in the statement of the theorem (with S replacing T).

So now suppose that $m_T(x) = f_1(x)^{a_1}\,f_2(x)^{a_2}\;\cdots\;f_k(x)^{a_k}$, where f_1, f_2, …, f_k are distinct monic irreducible polynomials over F. Let

$$g(x) := \prod_{i=1}^{k-1} f_i(x)^{a_i} \quad \text{and} \quad h(x) := f_k(x)^{a_k}.$$

By the Primary Decomposition Theorem, Mark 2, $V = U \oplus W$, where U, W are T-invariant and $m_{T|_U} = g$, $m_{T|_W} = h$. Applying the induction hypothesis to U and $T|_U$ we see that $U = U_1 \oplus \cdots \oplus U_{k-1}$, where the subspaces U_i are $T|_U$-invariant and the minimal polynomial of the restriction of $T|_U$ to U_i is $f_i^{a_i}$ for $1 \leqslant i \leqslant k-1$. But of course this simply means that U_i is T-invariant and the minimal polynomial of the restriction of T to U_i is $f_i^{a_i}$ for $1 \leqslant i \leqslant k-1$. Define $V_i := U_i$ for $1 \leqslant i \leqslant k-1$ and $V_k := W$ to complete the proof.

An important application of the Primary Decomposition Theorem is a criterion for diagonalisability of a linear transformation or a square matrix. Our linear transformation T is said to be *diagonalisable* if there is a basis of V consisting of eigenvectors of T. Thus T is diagonalisable if and only if there is a basis of V with respect to which its matrix is diagonal. Correspondingly therefore, a matrix $A \in \mathrm{M}_{n\times n}(F)$ is said to be *diagonalisable* if there exists an invertible $n \times n$ matrix P over F such that $P^{-1}AP$ is diagonal.

Theorem. *Our linear transformation $T : V \to V$ is diagonalisable if and only if $m_T(x)$ may be factorised as a product of distinct linear factors in $F[x]$.*

Proof. Suppose first that $m_T(x)$ may be factorised as a product of distinct linear factors in $F[x]$. This means that there exist distinct scalars $\lambda_1, \ldots, \lambda_k$ such that $m_T(x) = (x-\lambda_1)\cdots(x-\lambda_k)$. By the Primary Decomposition Theorem (Mark 3), $V = V_1 \oplus \cdots \oplus V_k$ where V_i is T-invariant for $1 \leqslant i \leqslant k$ and $T|_{V_i} - \lambda_i I_{V_i} = 0$. Thus all vectors v in V_i satisfy the equation $Tv = \lambda_i v$. If B_i is a basis of V_i then B_i consists of eigenvectors of T with eigenvalue λ_i, and so if $B := B_1 \cup \cdots \cup B_k$ then B is a basis of V consisting of eigenvectors. Thus T is diagonalisable.

Now suppose conversely that T is diagonalisable and let B be a basis of V consisting of eigenvectors of T. Let $\lambda_1, \ldots, \lambda_k$ be the distinct members of F that occur as eigenvalues for the vectors in B. Define $f(x) := (x-\lambda_1)\cdots(x-\lambda_k)$. We propose to show that $f(T) = 0$. Let $v \in B$. Then there exists $i \in \{1, \ldots, k\}$ such that $Tv = \lambda_i v$. Therefore $(T - \lambda_i I)v = 0$ and so $f(T)v = 0$. Since $f(T)$ annihilates all members of a basis of V its null-space (kernel) is V, that is, $f(T) = 0$. It follows that m_T divides f in $F[x]$. That would be enough to show that $m_T(x)$ is a product of *some of* the factors $(x-\lambda_i)$ and therefore factorises as a product of distinct linear factors in $F[x]$. But in fact we can go a little further—we know that every eigenvalue is a root of m_T and therefore in fact $m_T = f$, that is, $m_T(x) = (x-\lambda_1)\cdots(x-\lambda_k)$.

A worked example—Part of FHS 2001, Paper a1, Question 1:

> State a criterion for the diagonalizability of a linear transformation in terms of its minimum polynomial, and show that if two linear transformations S and T of V are diagonalizable and $ST = TS$, then there is a basis of V with respect to which *both* S and T have diagonal matrices.

Response. The first clause is the 'bookwork' we have just treated. For the second part, let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T and for $\lambda \in F$ define

$$U_i := \{v \in V \mid Tv = \lambda_i v\}$$

(the so-called λ_i-*eigenspace* of T). Since T is diagonalisable there is a basis of V

consisting of eigenvectors and this means that $V = U_1 \oplus \cdots \oplus U_k$. This is part of what we have just proved; note that the fact that eigenvectors for distinct eigenvalues are linearly independent was what you proved in Exercise 4 of the 'Mods Revision' Sheet.

Now we prove (see that same Exercise 4) that each subspace U_i is S-invariant. For, if $v \in U_i$ then $T(S\,v) = S(T\,v) = S(\lambda_i v) = \lambda_i S\,v$, and so $S\,v \in U_i$. Now the minimal polynomial $m_{S|_{U_i}}$ divides m_S, and m_S may be factorised as a product of distinct linear factors in $F[x]$, so the same is true of $m_{S|_{U_i}}$. It follows that there is a basis $u_{i\,1}, \ldots, u_{i\,d_i}$ of U_i consisting of eigenvalues of S. Note, however, that since these are non-zero vectors in U_i they are automatically eigenvectors also for T. And now if

$$B := \{u_{i\,j} \mid 1 \leqslant i \leqslant k,\ 1 \leqslant j \leqslant d_i\}$$

then B, being the union of bases of U_i for $1 \leqslant i \leqslant k$, is a basis of V, and it consists of vectors that are simultaneously eigenvectors of S and of T. Thus the matrices of both S and T with respect to the basis B of V are diagonal.

Exercise 28. For each of the matrices in Exercise 23, thought of as matrices with coefficients in $\mathbb{C}$, find an invertible matrix Y over $\mathbb{C}$ such that $Y^{-1}XY$ is diagonal, or prove that no such Y exists. Can such matrices be found, whose coefficients are real?

Exercise 29. Let $A := \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$. Is A diagonalisable when considered as a matrix over the following fields: (i) $\mathbb{C}$; (ii) $\mathbb{Q}$; (iii) $\mathbb{Z}_2$; (iv) $\mathbb{Z}_5$?

Exercise 30. Let $T : V \to V$ be a linear transformation of a finite-dimensional vector space over a field F, and $T' : V' \to V'$ the dual transformation. Show that T and T' have the same characteristic polynomial and the same minimal polynomial. Deduce that T is diagonalisable if and only if T' is diagonalisable.

Exercise 31. Let $T : V \to V$ be a linear transformation of a finite-dimensional vector space over a field F.

(i) Let $m_T(x)$ be its minimal polynomial and let $g(x)$ be a polynomial which is coprime with $m_T(x)$. Show that $g(T)$ is invertible. [*Hint*: use the existence of $u(x), v(x) \in F[x]$ such that $u(x)m_T(x) + v(x)g(x) = 1$.]

(ii) Using the Primary Decomposition Theorem, or otherwise, deduce that $V = V_1 \oplus V_2$ where V_1, V_2 are T-invariant subspaces (that is $T(V_i) \subseteq V_i$) such that the restriction T_1 of T to V_1 is invertible and the restriction T_2 of T to V_2 is nilpotent (that is, $T_2^m = 0$ for some $m \in \mathbb{N}$).

Triangular form

There are some linear transformations T and some square matrices A that cannot be diagonalised—for example the transformation $T : F^2 \to F^2$ given by $T : (x, y) \mapsto (x + y, y)$ or the matrix $A := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Nevertheless, one can choose bases with respect to which the matrices are particularly simple or particularly well-adapted to show the behaviour of T or of A. The most sophisticated of these give the so-called Rational Canonical Form and the Jordan Normal Form of matrices, but these are quite a long way beyond the scope of this course. Triangular form is a step in the direction of the Jordan Normal Form, and although it is quite a small step, it is extremely useful.

As previously, throughout this section F is a field, V is a finite-dimensional vector space over F, $n := \dim V$, and $T : V \to V$ is a linear transformation.

An $n \times n$ matrix A with entries $a_{ij} \in F$ is said to be *upper triangular* if $a_{ij} = 0$ when $i > j$. The following is almost trivial:

Observation. *If A is upper triangular then $c_A(x) = \displaystyle\prod_{i=1}^{n} (x - a_{ii})$, and therefore the eigenvalues of A are its diagonal entries $a_{11}, \ldots, a_{nn}$.*

Our transformation T is said to be *triangularisable* if there is a basis of V with respect to which its matrix is upper triangular.

Observation. *The matrix of T with respect to the basis $v_1, \ldots, v_n$ of V is upper triangular if and only if each subspace $\langle v_1, \ldots, v_r \rangle$ (for $1 \leqslant r \leqslant n$) is T-invariant.*

Theorem. *Our transformation T is triangularisable if and only if $c_T(x)$ may be factorised as a product of (not necessarily distinct) linear factors in $F[x]$.*

Note the comparison with the diagonalisability theorem on p. 25. There it was the minimal polynomial that mattered, and it had to factorise completely with distinct roots in F. Here it is the characteristic polynomial that matters, and it must factorise completely over F but can have multiple roots.

Proof. This is clear one way round: we have already seen that if the matrix of T is the upper triangular matrix a_{ij} with respect to a suitable basis then $c_T(x) = \det(xI - A) = \prod(x - a_{ii})$.

For the converse we use induction on the dimension n. There is nothing to prove if $n = 0$ or if $n = 1$. So suppose that $n \geqslant 2$ and that the theorem is known to hold for linear transformations of smaller-dimensional vector spaces. Suppose that $c_T(x) = \prod_{i=1}^{n}(x - \lambda_i)$, where $\lambda_1, \ldots, \lambda_n \in F$. Let λ be any one of the λ_i. For the sake of definiteness let's define $\lambda := \lambda_n$. Let $W := \operatorname{Im}(T - \lambda I)$ and let $m := \dim W$. Since λ is an eigenvalue, $\dim(\operatorname{Ker}(T - \lambda I)) \geqslant 1$ and so (by the Rank-Nullity Theorem) $m \leqslant n - 1$. Let $w_1, \ldots, w_m$ be a basis of W and extend this to a basis $w_1, \ldots, w_m, v_{m+1}, \ldots, v_n$ of V. Now W is T-invariant because if $w \in W$ then $w = Tv - \lambda v$ for some $v \in V$, and so $Tw = T^2 v - \lambda T v = (T - \lambda I) T v \in W$. Therefore there is an $m \times m$ matrix (a_{ij}) such that

$$T w_j = \sum_{i=1}^{m} a_{ij} w_i \quad \text{for } 1 \leqslant j \leqslant m.$$

Also, since for $m + 1 \leqslant j \leqslant n$ we know that $T v_j - \lambda v_j \in W$ there is an $m \times (n - m)$ matrix (b_{ij}) such that

$$T v_j = \lambda v_j + \sum_{i=1}^{m} b_{ij} w_i \quad \text{for } m + 1 \leqslant j \leqslant n.$$

The matrix A of T with respect to this basis has the partitioned form $\begin{pmatrix} A_1 & B_1 \\ 0 & \lambda I_{n-m} \end{pmatrix}$, where $A_1 = (a_{ij})$ (the matrix of $T|_W$ with respect to the basis $w_1, \ldots, w_m$ of W), $B_1 = (b_{ij})$, and the matrix in the south-west corner is an $(n-m) \times m$ zero matrix. Now

$$c_T(x) = \det(xI - A) = \det(xI - A_1)\,(x - \lambda)^{n-m},$$

and so $c_{T|_W}(x)$ divides $c_A(x)$ in $F[x]$. Therefore $c_{T|_W}(x)$ may be written as a product of linear factors in $F[x]$ and so by inductive hypothesis there is a basis $v_1, \ldots, v_m$ of W with respect to which the matrix A' of $T|_W$ is upper triangular. Then the matrix of T with respect to the basis $v_1, \ldots, v_m, v_{m+1}, \ldots, v_n$ of V is $\begin{pmatrix} A' & B' \\ 0 & \lambda I_{n-m} \end{pmatrix}$, for some $m \times (n-m)$ matrix B', and this is upper triangular, as required.

Note. In particular, if $F = \mathbb{C}$ then *every* linear transformation $V \to V$ is triangularisable. For, by the so-called Fundamental Theorem of Algebra (which is much more a theorem in Analysis than in Algebra) every polynomial with complex coefficients can be factorised as a product of linear factors in $\mathbb{C}[x]$.

The proof of the theorem gives a practical method for finding a basis with respect to which the matrix of T is triangular. We find an eigenvalue λ of T; then $(T - \lambda I)V$ is a proper T-invariant subspace of V. Find a triangularising basis there, and extend to V.

Worked example. Find a triangular form of A where

$$A := \begin{pmatrix} 6 & 2 & 3 \\ -3 & -1 & -1 \\ -5 & -2 & -2 \end{pmatrix}.$$

Response. We find that $c_A(x) = x^3 - 3x^2 + 3x - 1 = (x-1)^3$. Now if $V = F^3$ (column vectors) and $W := (A - I)V$ then W is spanned by the columns of $A - I$, that is of

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

Restricted to this subspace W the only eigenvalue of A (that is, of the linear transformation produced by multiplication of column vectors by A) is 1, and so we consider the space $(A - I)W$. This is spanned by the columns of $(A - I)^2$, which is

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

Choose the vector $(1, -1, -1)^{\text{tr}}$ to span this one-dimensional space; then by inspection we find that W is spanned by this vector together with $(0, 1, 0)^{\text{tr}}$, and so as triangularising basis for V we can take

$$\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

And in fact, with respect to this basis the matrix becomes

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

Exercise 32. Let $A := \begin{pmatrix} 4 & 9 \\ -1 & -2 \end{pmatrix}$, construed as a matrix over an arbitrary field F. Find an invertible 2×2 matrix P over F for which $P^{-1}AP$ is triangular. Are there any fields F over which P can be found such that $P^{-1}AP$ is diagonal?

Exercise 33. For each of the matrices in Exercise 24, thought of as matrices with coefficients in $\mathbb{C}$, find an invertible matrix Y over $\mathbb{C}$ such that $Y^{-1}XY$ is upper triangular. Can such matrices be found, whose coefficients are real?

The Cayley–Hamilton Theorem

We come now to one of the classic and lovely theorems of linear algebra:

Cayley–Hamilton Theorem. *Let V be a finite-dimensional vector space over a field F, and let $T : V \to V$ be a linear transformation. Then $c_T(T) = 0$.*

Equivalently: *If $A \in \mathrm{M}_{n\times n}(F)$ then $c_A(A) = 0$.*

Equivalently: *The minimal polynomial $m_T(x)$ divides the characteristic polynomial $c_T(x)$ in $F[x]$.*

Equivalently: *If $A \in \mathrm{M}_{n\times n}(F)$ then the minimal polynomial $m_A(x)$ divides the characteristic polynomial $c_A(x)$ in $F[x]$.*

Note. Historically this was a theorem about square matrices rather than linear transformations. Indeed, it was the second of the four assertions above. For reasons both historical and practical, in these notes I propose to work with $n \times n$ matrices over F. The translation to linear transformations $T : V \to V$ (for a finite-dimensional vector space V over F) is, however, absolutely routine.

There are many proofs of the theorem. It lies deeper than other parts of this course, and some of those proofs give little insight into just *why* the theorem holds. They merely prove it. In order to give you some insight into why it is true I propose to begin with three simple but suggestive observations which lead directly to an easy proof of the theorem over any subfield of $\mathbb{C}$.

Observation 1. *Let $A, B \in \mathrm{M}_{n\times n}(F)$. If $B = P^{-1}AP$, where P is invertible in $\mathrm{M}_{n\times n}(F)$, then $c_A(A) = 0$ if and only if $c_B(B) = 0$.*

Proof. Let P be an invertible $n \times n$ matrix over F and let $B := P^{-1}AP$. We know then that if $f(x) := c_A(x) = \det(xI - A)$ then also $f(x) = c_B(x) = \det(xI - B)$. It is an easy calculation (see Exercise 22 on p. 20) that

$$f(B) = f(P^{-1}AP) = P^{-1}f(A)\,P$$

and so $f(B) = 0$ if and only if $f(A) = 0$. That is, $c_A(A) = 0$ if and only if $c_B(B) = 0$, as required.

This can be thought of in another way. Choose a basis for an n-dimensional vector space over V and let $T : V \to V$ be the linear transformation whose matrix is A with respect to this basis. The correspondence between matrices and linear transformations says that $c_A(A) = 0$ if and only if $c_T(T) = 0$. But now if $B = P^{-1}AP$ then B is simply the matrix of the same linear transformation T with respect to a different basis, so $c_B(B) = 0$ if and only if $c_T(T) = 0$. Therefore $c_A(A) = 0$ if and only if $c_B(B) = 0$.

Observation 2. *If $A \in \mathrm{M}_{n\times n}(F)$ and A is diagonalisable then $c_A(A) = 0$.*

For, by the previous observation this is true if and only if it is true for diagonal matrices. Suppose then that, in an obvious notation, $A = \mathrm{Diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. Then $c_A(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$, and so $c_A(A) = (A-\lambda_1 I)(A-\lambda_2 I)\cdots(A-\lambda_n I)$. Each factor $(A - \lambda_i I)$ is a diagonal matrix, and its i^{th} diagonal entry is 0. Therefore $c_A(A)$ is a diagonal matrix and its i^{th} diagonal entry is 0 for every i; that is, $c_A(A) = 0$.

As it happens, there is a strong sense in which ‘almost all’ matrices are diagonalisable, and therefore these simple arguments have already proved the Cayley–Hamilton Theorem for ‘most’ matrices (and therefore for ‘most’ linear transformations). We can however take one further step and prove the theorem for triangularisable matrices.

Observation 3. *If $A \in \mathrm{M}_{n\times n}(F)$ and A is triangularisable then $c_A(A) = 0$.*

Proof. Again, by the first observation above, we may assume that in fact A is upper triangular. Then this observation was Qn 6 on the preliminary exercise sheet (Mods Revision). We think of A as partitioned in the form $\begin{pmatrix} A_1 & B \\ 0 & \lambda_n \end{pmatrix}$ where A_1 is an $(n-1)\times(n-1)$ upper triangular matrix, B is a $(n-1)\times 1$ column vector, and 0 denotes the $1\times(n-1)$ zero row vector. Then

$$\det(xI - A) = \det(xI - A_1) \times \det(x - \lambda_n),$$

that is, $c_A(x) = c_{A_1}(x)\,(x - \lambda_n)$. For any polynomial $f \in F[x]$ we find that

$$f(A) = \begin{pmatrix} f(A_1) & C \\ 0 & f(\lambda_n) \end{pmatrix},$$

where C is some $(n-1)\times 1$ column vector, and 0 again denotes the $1\times(n-1)$ zero row vector. As inductive assumption we may assume that $c_{A_1}(A_1) = 0$. Then, for suitable column vectors C, C',

$$\begin{aligned} c_A(A) &= c_{A_1}(A)\,(A - \lambda_n I) \\ &= \begin{pmatrix} c_{A_1}(A_1) & C \\ 0 & c_{A_1}(\lambda_n) \end{pmatrix} \begin{pmatrix} A_1 - \lambda_n I & C' \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & C \\ 0 & c_{A_1}(\lambda_n) \end{pmatrix} \begin{pmatrix} A_1 - \lambda_n I & C' \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Corollary. *The Cayley–Hamilton Theorem holds for matrices over $\mathbb{C}$ and for matrices over any subfield F of $\mathbb{C}$, such as $\mathbb{Q}$ or $\mathbb{R}$.*

Proof. We know (see the note on p. 28) that every matrix over $\mathbb{C}$ is trangularisable, and so the result follows immediately from Observation 3.

If $A \in \mathrm{M}_{n\times n}(F)$, where F is a subfield of $\mathbb{C}$ then we can think of A as a matrix over $\mathbb{C}$. As such it is known to be annihilated by its characteristic polynomial, which is what we wanted to show.

It is worth a digression to take a brief look at the original source of the theorem. This is A. Cayley, 'A memoir on the theory of matrices', *Phil. Trans Roy. Soc. London*, 148 (1858), 17–37, which is reprinted in *The Collected Mathematical Papers of Arthur Cayley*, Vol. II, pp. 475–495. Hamilton's connection with the theorem seems to have been more tenuous and to have come from one of his theorems about his quaternions. Here is a quotation from Cayley's introduction to his paper:

> I obtain the remarkable theorem that any matrix whatever satisfies an algebraical equation of its own order, [...] viz. the determinant, formed out of the matrix diminished by the matrix considered as a single quantity involving the matrix unity, will be equal to zero.

And here is an extract from §§21–23 (I have tried to reproduce Cayley's notation for matrices faithfully, with the first row enclosed in round brackets, the rest enclosed in vertical bars):

> **21.** The general theorem before referred to will be best understood by a complete development of a particular case. Imagine a matrix
>
> $$\begin{array}{l} (\ a,\quad b\ ), \\ \left|\ c,\quad d\ \right| \end{array}$$
>
> and form the determinant
>
> $$\left|\begin{array}{ll} a-M, & b \\ c \qquad , & d-M \end{array}\right|,$$
>
> the developed expression of this determinant is
>
> $$M^2-(a+d)\,M^1+(ad-bc)\,M^0\,;$$
>
> the values of M^2, M^1, M^0 are
>
> $$\begin{array}{l} (\ a^2+bc\ ,\quad b(a+d)\ ), \\ \left|\ c(a+d),\quad d^2+bc\ \right| \end{array} \qquad \begin{array}{l} (\ a,\quad b\ ), \\ \left|\ c,\quad d\ \right| \end{array} \qquad \begin{array}{l} (\ 1,\quad 0\ ), \\ \left|\ 0,\quad 1\ \right| \end{array}$$
>
> and substituting these values the determinant becomes equal to the matrix zero, [...].
>
> [...]
>
> **23.** I have verified the theorem, in the next simplest case of a matrix of the order 3, viz. if M be such a matrix, suppose
>
> $$\begin{array}{l} (\ a,\quad b,\quad c\ ), \\ \left|\ d,\quad e,\quad f\ \right| \\ \left|\ g,\quad h,\quad i\ \right| \end{array}$$

> then the derived determinant vanishes, or we have
>
> $$\begin{vmatrix} a-M, & b \quad , & c \\ d \quad , & e-M, & f \\ g \quad , & h \quad , & i-M \end{vmatrix} = 0\,,$$
>
> or expanding
>
> $$M^3-(a+e+i)\,M^2+(ei+ia+ae-fh-cg-bd)\,M-(aei+bfg+cdh-afh-bdi-ceg)=0\,;$$
>
> but I have not thought it necessary to undertake the labour of a formal proof of the theorem in the general case of a matrix of any degree.

What a charming piece of 19th Century chutzpah: "I have not thought it necessary to undertake the labour of a formal proof of the theorem in the general case of a matrix of any degree." It is very unlikely that the method which Cayley uses for 2×2 matrices, and sketches for the 3×3 case could be practical for the $n\times n$ case: even if one could write down explicitly the characteristic polynomial of an $n\times n$ matrix A, it seems unrealistic to expect to write down the (i,j) coefficient of a general power A^k for all k up to n, and then evaluate $c_A(A)$ in the way that it is possible in the 2×2 case. So the fact is, he can't really have had a proof. But there's another fact: he may not have appreciated rigorous thinking in the same way as Oxford students now do (the poor chap went to Cambridge and missed out on the Oxford experience) but he did have a wonderful insight. He *knew* that the theorem was right, and for him proof, though it would have been nice to have, was less important than having an insight which he could use in all sorts of ways to solve other mathematical problems.

That is as far as I am going to go with the Cayley–Hamilton Theorem. We have a proof of it over any field F contained as a subfield of $\mathbb{C}$. That proof can, once one knows more about fields in general, be easily adapted to other fields. For a general proof using adjoint matrices (a proof which, though pleasantly short, to me seems rather unnatural and gives no insight into why the theorem holds) see, for example, T. S. Blyth & E. F. Robertson, *Basic Linear Algebra*, p. 169 or Richard Kaye & Robert Wilson *Linear Algebra*, p. 170. For a general proof using 'rational canonical form' see, for example, Peter J. Cameron, *Introduction to Algebra*, p. 154 or Charles W. Curtis, *Linear Algebra*, p. 226.

Further exercises II

Exercise 35. Let V be a finite-dimensional vector space over a field F and let $T: V\to V$ be a linear transformation.

(i) Suppose that $F=\mathbb{C}$ and that $T^4=T$. Show that T is diagonalisable.

(ii) Now suppose that $F=\mathbb{Z}_3$ and that $m_T(x)=x^4-x$. Is T diagonalisable?

Exercise 36 Let V be a finite-dimensional vector space and let $S,T: V\to V$ be linear transformations. Let m_1 and m_2 denote the minimal polynomials of ST and TS respectively. By considering relations such as $T(ST)^rS=(TS)^{r+1}$ show that $m_2(x)=x^i m_1(x)$, where $i\in\{1,\,0,\,-1\}$. Show that λ is an eigenvalue of ST if and only if λ is an eigenvalue of TS.

Exercise 37 [FHS 1995, A1, 3]. Let V be a finite-dimensional vector space over a field F, and let $T : V \to V$ be a linear transformation. Let $v \in V$, $v \neq 0$, and suppose that V is spanned by $\{v, Tv, T^2v, \ldots, T^jv, \ldots\}$.

(i) Show that there is an integer $k \geqslant 1$ such that $v, Tv, T^2v, \ldots, T^{k-1}v$ are linearly independent but $T^kv = \alpha_0 v + \alpha_1 Tv + \cdots + \alpha_{k-1}T^{k-1}v$ for some $\alpha_0, \alpha_1, \ldots, \alpha_{k-1} \in F$.

(ii) Prove that $\{v, Tv, T^2v, \ldots, T^{k-1}v\}$ is a basis for V.

Let T have minimum polynomial $m(x)$ and characteristic polynomial $c(x)$.

(iii) Prove that $m(x) = x^k - \alpha_{k-1}x^{k-1} - \cdots - \alpha_1 x - \alpha_0$.

(iv) By considering the matrix of T with respect to the basis $\{v, Tv, T^2v, \ldots, T^{k-1}v\}$, prove (without using the Cayley–Hamilton Theorem) that $m(x) = c(x)$.

Now let A be the matrix $\begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. Show that there does *not* exist a column vector $v \in \mathbb{R}^4$ such that $\mathbb{R}^4$ is spanned by $\{v, Av, A^2v, \ldots, A^jv, \ldots\}$.

Part III: Inner Product Spaces

In the first two parts of this course the focus has been on linear algebra—vector spaces, linear transformations and matrices—over an arbitrary field F. We come now to that part of linear algebra which is rather more geometric, the theory of inner products and inner product spaces. Inner products are generalisations of the familiar dot product $u \cdot v$ of n-vectors with real coordinates, defined by $u \cdot v = \sum x_i y_i$ (where u has coordinates $x_1, \ldots, x_n$ and v has coordinates $y_1, \ldots, y_n$). Although this theory can be extended to arbitrary fields, or at least parts of it can, its most important and most natural manifestations are over $\mathbb{R}$ and $\mathbb{C}$. Therefore from now on we restrict to these fields.

Real inner product spaces and their geometry

Let V be a vector space over $\mathbb{R}$. An *inner product* on V is a function $B : V \times V \to F$ such that for all $u, v, w \in V$ and all $\alpha, \beta \in \mathbb{R}$,

(1) $B(\alpha u + \beta v, w) = \alpha B(u, w) + \beta B(v, w)$

(2) $B(u, v) = B(v, u)$ [B is *symmetric*]

(3) if $u \neq 0$ then $B(u, u) > 0$ [B is *positive definite*]

Note: From (1) and (2) follows

(1′) $B(u, \alpha v + \beta w) = \alpha B(u, v) + \beta B(u, w)$

A function $V \times V \to \mathbb{R}$ satisfying (1) and (1′) is said to be *bilinear* and called a *bilinear form.* Thus an inner product on a real vector space is a positive definite symmetric bilinear form. A *real inner product space* is a vector space over $\mathbb{R}$ equipped with an inner product.

Notation. Often we find $\langle u, v\rangle$ or $\langle u \,|\, v\rangle$ used to denote what here is $B(u, v)$. I propose to use the former. In an inner product space we define

$$||u|| := \langle u, u\rangle^{\frac{1}{2}}.$$

Thus $||u|| \geqslant 0$ and $||u|| = 0$ if and only if $u = 0$; this is known as the *length* or as the *norm* of the vector u.

Example 1: $V = \mathbb{R}^n$ and $\langle u, v\rangle = u \cdot v = u^{\mathrm{tr}}\, v$. This example is our standard real inner product space and we call it euclidean space.

Example 2: $V = C[a, b]$ (continuous functions $f : [a, b] \to \mathbb{R}$) and

$$\langle f, g\rangle = \int_a^b f(t)\, g(t)\, \mathrm{d}t\,.$$

Note. From condition (1′) for an inner product it follows that $\langle u, 0\rangle = 0$ for any $u \in V$. Conversely, Suppose that $u \in V$ and $\langle u, v\rangle = 0$ for all $v \in V$. Then in particular $\langle u, u\rangle = 0$ and so $u = 0$. This property of an inner product is expressed by saying that it is a *non-degenerate* or *non-singular* bilinear form.

Let V be a real inner product space. If $u, v \in V$ and $\langle u, v\rangle = 0$ then we say that u, v are *orthogonal*. For $X \subseteq V$ we define

$$X^{\perp} := \{v \in V \mid \langle u, v\rangle = 0 \text{ for all } u \in X\}.$$

For $u \in V$ we write $u^{\perp}$ for $\{u\}^{\perp}$.

Lemma. *Let V be a real inner product space and let $X \subseteq V$. Then $X^{\perp}$ is a subspace of V.*

Proof. We have already seen in the note above that $0 \in X^{\perp}$. Also, if $v, w \in X^{\perp}$ and $\alpha, \beta \in \mathbb{R}$ then $\langle u, \alpha v + \beta w\rangle = \alpha\langle u, v\rangle + \beta\langle u, w\rangle = 0$ for any $u \in X$, and so $\alpha v + \beta w \in X^{\perp}$. Thus $X^{\perp}$ is a subspace.

Lemma. *Let V be a real inner product space and let $u \in V$. Then*

$$V = \operatorname{Span}(u) \oplus u^{\perp}.$$

Proof. If $u = 0$ there is nothing to prove since then $u^{\perp} = V$. So suppose that $u \neq 0$. If $u_1 := ||u||^{-1} u$ then

$$||u_1|| = \langle u_1, u_1\rangle = ||u||^{-2}\langle u, u\rangle = 1.$$

Moreover, $\operatorname{Span}(u) = \operatorname{Span}(u_1)$ and $u^{\perp} = u_1^{\perp}$. Thus we may assume without loss of generality that $||u|| = 1$. Now for $v \in V$ define $\alpha := \langle u, v\rangle$ and $w := v - \alpha u$. Then

$$\langle u, w\rangle = \langle u, v - \alpha u\rangle = \langle u, v\rangle - \alpha\langle u, u\rangle = \langle u, v\rangle - \alpha = 0,$$

and so $w \in u^{\perp}$. This shows that $V = \operatorname{Span}(u) + u^{\perp}$. If $x \in \operatorname{Span}(u) \cap u^{\perp}$ then $x = \lambda u$ for some $\lambda \in \mathbb{R}$ but also $\langle u, x\rangle = 0$, so $0 = \langle u, \lambda u\rangle = \lambda\langle u, u\rangle = \lambda$, that is, $x = 0$. Therefore also $\operatorname{Span}(u) \cap u^{\perp} = \{0\}$ and so $V = \operatorname{Span}(u) \oplus u^{\perp}$ as we claimed.

Let V be a real inner product space. Vectors u_1, u_2, ..., u_k in V are said to form an *orthogonal* set if $\langle u_i, u_j\rangle = 0$ whenever $i \neq j$.

Lemma. *An orthogonal set of non-zero vectors in a real inner product space is linearly independent.*

Proof. Let V be a real inner product space and let u_1, u_2, ..., u_k be an orthogonal set in $V \setminus \{0\}$. Suppose that $\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_k u_k = 0$ where $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}$. Then for $1 \leqslant i \leqslant k$ we have

$$0 = \Big\langle \sum \alpha_j u_j, u_i \Big\rangle = \sum \alpha_j \langle u_j, u_i\rangle = \alpha_i \langle u_i, u_i\rangle$$

by orthogonality, and so $\alpha_i = 0$ since $\langle u_i, u_i\rangle \neq 0$. Thus u_1, u_2, ..., u_k are linearly independent.

We are particularly interested in orthogonal sets of vectors of length 1. Vectors u_1, u_2, ..., u_k in a real inner product space V are said to form an *orthonormal* set if

$$\langle u_i, u_j\rangle = \begin{cases} 1 & \text{if } i = j\,, \\ 0 & \text{if } i \neq j\,. \end{cases}$$

THEOREM. *Let V be a finite-dimensional real inner product space, let $n := \dim V$, and let $u \in V\backslash\{0\}$. There is an orthonormal basis $v_1, v_2, \ldots, v_n$ in which $v_1 = ||u||^{-1} u$.*

Proof. This is trivially true if $n = 0$ or if $n = 1$. Now suppose that $n > 1$ and as inductive hypothesis suppose it is true for real inner product spaces of dimension $n - 1$. Define $v_1 := ||u||^{-1} u$ and $V_1 := v_1^\perp$. We know then that $V = \mathrm{Span}(v_1) \oplus V_1$ and, in particular, $\dim V_1 = n - 1$. Obviously, the restriction of our inner product to V_1 is an inner product on V_1. Starting from any non-zero vector in V_1 the inductive hypothesis yields that there is an orthonormal basis $v_2, \ldots, v_n$ of V_1. Then, since $\langle v_1, v_i\rangle = 0$ for $2 \leqslant i \leqslant n$, the vectors $v_1, v_2, \ldots, v_n$ form an orthonormal basis of V_1.

THEOREM. *Let V be a real inner product space. If U is a finite-dimensional subspace then $V = U \oplus U^\perp$.*

Proof. Certainly $U \cap U^\perp = \{0\}$ because if $u \in U \cap U^\perp$ then $\langle x, x\rangle = 0$. What needs to be proved therefore is that $V = U + U^\perp$.

The restriction to U of our inner product on V is obviously an inner product on U. Since U is finite-dimensional we know it has an orthonormal basis $u_1, \ldots, u_n$, say. For $v \in V$ define $x_i := \langle u_i, v\rangle$ for $1 \leqslant i \leqslant n$, and then define $u := \sum_i x_i u_i$ and $w := v - u$. Now

$$\langle u_j, w\rangle = \langle u_j, v - \sum_i x_i u_i\rangle = \langle u_j, v\rangle - \sum_i x_i \langle u_j, u_i\rangle = \langle u_j, v\rangle - x_j \langle u_j, u_j\rangle = 0\,,$$

and so $w \in U^\perp$ since w is orthogonal to all elements of a basis of U. This shows that $v = u + w \in U + U^\perp$, and completes the proof.

EXERCISE 38. Let V be a real inner product space and let U be a subspace of V. Show that $(U^\perp)^\perp = U$.

THEOREM. *Let V be a finite-dimensional real inner product space and let $v_1, \ldots, v_n$ be an orthonormal basis of V. Let $u, w \in V$ and suppose that*

$$u = x_1 v_1 + \cdots + x_n v_n\,, \quad w = y_1 v_1 + \cdots + y_n v_n\,,$$

where $x_i, y_j \in \mathbb{R}$ for all i, j. Then

(1) $x_i = \langle u, v_i\rangle$;

(2) $||u|| = \left(\sum |x_i|^2\right)^{\frac{1}{2}}$;

(3) $\langle u, w\rangle = \sum x_i y_i$.

Proof. First, $\langle u, v_i\rangle = \langle \sum_j x_j v_j, v_i\rangle = \sum_j x_j \langle v_j, v_i\rangle = x_i$ since the terms in which $j \neq i$ are 0 and $\langle v_i, v_i\rangle = 1$.

Next, $||u||^2 = \langle \sum_i x_i v_i, \sum_j x_j v_j\rangle = \sum_{i,j} x_i x_j \langle v_i, v_j\rangle = \sum_i x_i^2$ and therefore $||u|| = (\sum_i x_i^2)^{1/2}$, as stated in (2).

Finally, $\langle u, w\rangle = \langle \sum_i x_i v_i, \sum_j y_j v_j\rangle = \sum_{i,j} x_i y_j \langle v_i, v_j\rangle = \sum_i x_i y_j$ as stated in (3).

The significance of this theorem is twofold. First, it shows the value of an orthonormal basis for computing norms and inner products. Secondly, it shows that if we use an orthonormal basis to identify our inner product space with $\mathbb{R}^n$ then our abstract inner product is identified with the familiar dot product:

COROLLARY. *Let V be a real inner product space of dimension n. Then there is an isomorphism (bijective linear transformation) $\varphi : \mathbb{R}^n \to V$ such that*

$$\textit{if } \varphi((x_1, \ldots, x_n)^{\mathrm{tr}}) = u \textit{ and } \varphi((y_1, \ldots, y_n)^{\mathrm{tr}}) = v \textit{ then } \langle u, v\rangle = \sum x_i\, y_i\,.$$

If V, W are real inner product spaces, with inner products $\langle\ ,\ \rangle_V$ and $\langle\ ,\ \rangle_W$ respectively, we say that V is *isometric* with W if there exists a linear isomorphism $\varphi : V \to W$ such that $\langle \varphi u, \varphi v\rangle_W = \langle u, v\rangle_V$ for all $u, v \in V$. Thus the above corollary says that any n-dimensional real inner product space is isometric with $\mathbb{R}^n$ equipped with its usual scalar (dot) product.

EXERCISE 39. Let V be a vector space over $\mathbb{R}$, and let $\langle u, v\rangle_1$ and $\langle u, v\rangle_2$ be inner products defined on V. Prove that if $\langle x, x\rangle_1 = \langle x, x\rangle_2$ for all $x \in V$, then $\langle u, v\rangle_1 = \langle u, v\rangle_2$ for all $u, v \in V$.

Complex inner product spaces

Let V be a vector space over $\mathbb{C}$. An *inner product* on V is a function $B : V \times V \to \mathbb{C}$ such that for all $u, v, w \in V$ and all $\alpha, \beta \in \mathbb{C}$

C(1) $\quad B(\alpha\, u + \beta\, v, w) = \alpha\, B(u, w) + \beta\, B(v, w)$

C(2) $\quad B(u, v) = \overline{B(v, u)}$ $\qquad$ [B is *conjugate symmetric*]

C(3) $\quad$ if $u \neq 0$ then $B(u, u) > 0$ $\qquad$ [B is *positive definite*]

NOTE. From C(1) and C(2) follows

C(2′) $\quad B(u, \alpha\, v + \beta\, w) = \bar{\alpha}\, B(u, v) + \bar{\beta}\, B(u, w)$.

We express this by saying that B is *semilinear* or *conjugate linear* in its second variable, or that B is *sesquilinear* (one-and-a-half linear). A complex inner product is often called a *Hermitian form* in honour of CHARLES HERMITE (1822–1901). A *complex inner product space* is a complex vector space equipped with an inner product in this sense.

NOTATION. As in the real case $\langle u, v\rangle$ or $\langle u \,|\, v\rangle$ are often used for inner products. And as in the real case we define $||u|| := \langle u, u\rangle^{\frac{1}{2}}$.

EXAMPLE 1: $V = \mathbb{C}^n$ and $\langle u, v\rangle = u^{\mathrm{tr}}\, \bar{v}$. This example is our standard complex inner product space.

EXAMPLE 2: $V = \{f : [a, b] \to \mathbb{C} \mid f \text{ is continuous}\}$ and

$$\langle f, g\rangle = \int_a^b f(t)\, \overline{g(t)}\, \mathrm{d}t\,.$$

Let V be a complex inner product space. *Orthogonality* is defined just as in real inner product spaces. A finite-dimensional complex inner product space has an orthonormal basis. If $U \leqslant V$ and U is finite-dimensional then $V = U \oplus U^{\perp}$. The proofs of these simple facts are almost exactly the same as in the real case and are offered as exercises for the reader:

Exercise 40. Prove

(i) that if V is a finite-dimensional complex inner product space then V has an orthonormal basis;

(ii) that if V is a complex inner product space and U is a finite-dimensional subspace then $V = U \oplus U^{\perp}$.

Theorem. *Let V be a finite-dimensional complex inner product space and let $v_1, \ldots, v_n$ be an orthonormal basis of V. Let $u, w \in V$ and suppose that*

$$u = x_1 v_1 + \cdots + x_n v_n\,, \quad w = y_1 v_1 + \cdots + y_n v_n\,,$$

where $x_i, y_j \in \mathbb{C}$ for all i, j. Then

(1) $\quad x_i = \langle u, v_i \rangle\,;$

(2) $\quad ||u|| = \left(\sum |x_i|^2 \right)^{\frac{1}{2}}\,;$

(3) $\quad \langle u, w \rangle = \sum x_i\, \overline{y_i}\,.$

Exercise 41. Prove this theorem.

Isometry of complex inner product spaces is defined exactly as in the real case and we have the following important consequence of the theorem:

Corollary. *Let V be a complex inner product space of dimension n. Then V is isometric with $\mathbb{C}^n$ with its standard inner product: $\langle x, y \rangle = x^{\mathrm{tr}}\, \bar{y} = \sum x_i\, \overline{y_i}$ where x is the column vector $(x_1, \ldots, x_n)^{\mathrm{tr}}$ and y is the column vector $(y_1, \ldots, y_n)^{\mathrm{tr}}$.*

Exercise 42. Let V be a vector space over $\mathbb{C}$, and let $\langle u,\, v \rangle_1$ and $\langle u,\, v \rangle_2$ be inner products defined on V. Is it true that if $\langle x,\, x \rangle_1 = \langle x,\, x \rangle_2$ for all $x \in V$, then $\langle u,\, v \rangle_1 = \langle u,\, v \rangle_2$ for all $u, v \in V$? [Compare Exercise 39.]

The Gram–Schmidt process

Although I have introduced real and complex inner product spaces separately, we have already seen that the two theories are very similar. Indeed, they have much in common. Since complex conjugation leaves real numbers unchanged, one could in fact, had one so wished, have used C(1), C(2) and C(3) to define inner products in the real case as well as the complex case. I prefer not to do that (indeed, I'm inclined to think that that would be ill-advised) because the real world and the complex world are quite distinct entities, and although they are of course intimately related to each other it seems best to keep them separate. Nevertheless, from now on I propose to treat their theories together.

THEOREM. *Let V be a real or complex inner product space and let $u_1, \ldots, u_n$ be linearly independent vectors in V. Then there exists an orthonormal set $v_1, \ldots, v_n$ in V such that*

$$\mathrm{Span}\,(v_1, \ldots, v_k) = \mathrm{Span}\,(u_1, \ldots, u_k) \quad \textit{for} \quad 0 \leqslant k \leqslant n\,.$$

Proof. Since $u_1 \neq 0$ we can define $v_1 := ||u_1||^{-1}\, u_1$. Then $||v_1|| = 1$ and $\mathrm{Span}\,(v_1) = \mathrm{Span}\,(u_1)$. Suppose as inductive hypothesis that $1 \leqslant k < n$ and we have found an orthonormal set $v_1, \ldots, v_k$ in V such that $\mathrm{Span}\,(v_1, \ldots, v_k) = \mathrm{Span}\,(u_1, \ldots, u_k)$. For $1 \leqslant i \leqslant k$ define $\alpha_i := \langle u_{k+1}, v_i\rangle$ and $w := \sum_i \alpha_i\, v_i$. (We should think geometrically of w as the orthogonal projection of u_{k+1} into the subspace spanned by $v_1, \ldots, v_k$, which is the same as the subspace spanned by $u_1, \ldots, u_k$.) Now $w \in \mathrm{Span}\,(v_1, \ldots, v_k) = \mathrm{Span}\,(u_1, \ldots, u_k)$ whereas $u_{k+1} \notin \mathrm{Span}\,(u_1, \ldots, u_k)$. Therefore if $v := u_{k+1} - w$ then $v \neq 0$ and we can define $v_{k+1} := ||v||^{-1}\, v$. Then $||v_{k+1}|| = 1$. Also,

$$\begin{aligned}\langle v, v_i\rangle = \langle u_{k+1} - w, v_i\rangle &= \langle u_{k+1}, v_i\rangle - \langle w, v_i\rangle \\ &= \langle u_{k+1}, v_i\rangle - \Big\langle \sum_{j=1}^{k} \alpha_j\, v_j, v_i \Big\rangle \\ &= \langle u_{k+1}, v_i\rangle - \alpha_i = 0\end{aligned}$$

and it follows that also $\langle v_{k+1}, v_i\rangle = 0$ for $1 \leqslant i \leqslant k$. Thus $v_1, \ldots, v_k, v_{k+1}$ is an orthonormal set. Finally,

$$\begin{aligned}\mathrm{Span}\,(v_1, \ldots, v_k, v_{k+1}) &= \mathrm{Span}\,(v_1, \ldots, v_k, v) \\ &= \mathrm{Span}\,(u_1, \ldots, u_k, u_{k+1} - w) \\ &= \mathrm{Span}\,(u_1, \ldots, u_k, u_{k+1}),\end{aligned}$$

where the last equation comes from the fact that $w \in \mathrm{Span}\,(u_1, \ldots, u_k)$.

NOTE 1. The construction in the proof is known as the *Gram–Schmidt orthogonalisation process.*

NOTE 2 If the Gram–Schmidt process gives the orthonormal basis $v_1, \ldots, v_n$ from a basis $u_1, \ldots, u_n$ and T is the transition matrix from the latter to the former then T is *positive* upper triangular—that is, upper triangular with positive real diagonal entries. For, in the course of the proof we found that $v_j = \sum_{i=1}^{j} \beta_{ij}\, u_i$ where $\beta_{jj} = ||v||^{-1}$ for a certain non-zero vector v, and therefore $\beta_{jj} > 0$.

A WORKED EXAMPLE—FHS 1999, Paper a1, Qn 3:

> Let V be a finite-dimensional vector space over $\mathbb{R}$. Define what is meant by saying that V *has an inner product* $\langle\ ,\ \rangle$ *over* $\mathbb{R}$.
>
> Let $\{b_1, b_2, \ldots, b_n\}$ be a basis for V. Prove that there exists an orthonormal basis $\{e_1, e_2, \ldots, e_n\}$ of V such that for each $k = 1, 2, \ldots, n$ the set

$\{e_1, e_2, \ldots, e_k\}$ is a basis for the subspace spanned by $\{b_1, b_2, \ldots, b_k\}$. Deduce that there are $t_{ij} \in \mathbb{R}$ with $t_{ii} \neq 0$ such that

$$b_1 = t_{11}e_1\,, \quad b_2 = t_{12}e_1 + t_{22}e_2\,, \quad \ldots\,, \quad b_k = t_{1k}e_1 + \cdots + t_{kk}e_k\,, \quad \ldots$$

(for $1 \leqslant k \leqslant n$). Show that $\langle b_i, b_j\rangle = \sum_{k=1}^n \langle b_i, e_k\rangle\langle b_j, e_k\rangle$ (for $1 \leqslant i, j \leqslant n$).

Now let G be the $n \times n$ matrix with (i,j)th entry equal to $\langle b_i, b_j\rangle$. Show that $\det G = \prod_{i=1}^{n} {t_{ii}}^2$ and deduce that G is non-singular. Show also that $\det G \leqslant \prod_{i=1}^{n} \langle b_i, b_i\rangle$.

Response. The first three instructions ask for bookwork that has just been treated above. For the fourth note that since $\{e_1, e_2, \ldots, e_n\}$ is an orthonormal basis of V and $b_j = \sum_{r=1}^{j} t_{rj}\, e_r$ we have $\langle b_j, e_i\rangle = t_{ij}$. So we calculate as follows:

$$\langle b_i, b_j\rangle = \Big\langle \sum_r t_{ri}\, e_r, \sum_s t_{sj}\, e_s \Big\rangle = \sum_{r,s} t_{ri}\, t_{sj}\, \langle e_r, e_s\rangle = \sum_{k=1}^{n} t_{ki}\, t_{kj}\,,$$

since $\langle e_r, e_s\rangle$ is 0 if $r \neq s$ and is 1 if $r = s = k$. That is, $\langle b_i, b_j\rangle = \sum_{k=1}^n \langle b_i, e_k\rangle\langle b_j, e_k\rangle$, as required.

Now let G be the matrix $(\langle b_i, b_j\rangle)$ [which is known as the *Gram* matrix of the inner product with respect to the basis $\{b_1, b_2, \ldots, b_n\}$], and let T be the matrix (t_{ij}). The formula that has just been proved tells us that $G = T^{\text{tr}}\, T$. Therefore $\det G = (\det T)^2$. But T is upper triangular and therefore $\det T = \prod t_{ii}$. Hence $\det G = \prod_{i=1}^n t_{ii}^2$ as required. Since $t_{ii} \neq 0$ for all relevant i we see that $\det G \neq 0$ and so G is non-singular.

Finally, $\langle b_i, b_i\rangle = \langle \sum_r t_{ri}\, e_r, \sum_s t_{si}\, e_s\rangle = \sum_r t_{ri}^2$ and so, since the coefficients are real, $t_{ii}^2 \leqslant \langle b_i, b_i\rangle$. Therefore $\det G = \prod t_{ii}^2 \leqslant \prod \langle b_i, b_i\rangle$, as required.

Exercise 43. Let V be the vector space of polynomials of degree $\leqslant 3$ with real coefficients. Define $\langle f, g\rangle := \int_{-1}^{1} f(t)\, g(t)\, \mathrm{d}t$. Show that this is an inner product on V. Use the Gram–Schmidt process to find an orthonormal basis for V.

Exercise 44. How does the answer to Exercise 43 change if $\langle f, g\rangle := \int_0^1 f(t)\, g(t)\, \mathrm{d}t$?

Bessel's Inequality

Bessel's Inequality. *Let V be a real or complex inner product space and let $v_1, \ldots, v_m$ be an orthonormal set in V. If $u \in V$ then*

$$\sum_{1}^{m} |\langle u, v_i\rangle|^2 \leqslant ||u||^2\,.$$

Equality holds if and only if $u \in \mathrm{Span}(v_1, \ldots, v_m)$.

Proof. In the following calculation we shall use notation appropriate to the complex case. For $u \in V$ define $w := u - \sum_{i=1}^m \langle u, v_i\rangle\, v_i$. Then $\langle w, w\rangle \geqslant 0$, but also

$$\begin{aligned}\langle w, w\rangle &= \Big\langle u - \sum_{i=1}^m \langle u,\ v_i\rangle\, v_i, u - \sum_{i=1}^m \langle u, v_i\rangle\, v_i\Big\rangle \\ &= \langle u, u\rangle - \Big\langle u, \sum_{i=1}^m \langle u, v_i\rangle\, v_i\Big\rangle - \Big\langle \sum_{i=1}^m \langle u, v_i\rangle\, v_i, u\Big\rangle + \Big\langle \sum_{i=1}^m \langle u, v_i\rangle\, v_i, \sum_{j=1}^m \langle u, v_j\rangle\, v_j\Big\rangle \\ &= \langle u, u\rangle - \sum_{i=1}^m \overline{\langle u, v_i\rangle}\, \langle u, v_i\rangle - \sum_{i=1}^m \langle u, v_i\rangle\, \langle v_i, u\rangle + \sum_{i,j=1}^m \langle u, v_i\rangle\, \overline{\langle u, v_j\rangle} \langle v_i, v_j\rangle.\end{aligned}$$

Since $\langle v_i, u\rangle = \overline{\langle u, v_i\rangle}$ and $\langle v_i, v_j\rangle$ is 1 if $i = j$ and 0 otherwise, we find that

$$\langle w, w\rangle = \langle u, u\rangle - \sum_{i=1}^m |\langle u, v_i\rangle|^2.$$

Therefore $\langle u, u\rangle \geqslant \sum_{i=1}^m |\langle u, v_i\rangle|^2$ as the theorem states.

Suppose that equality holds. The argument shows that then $w = 0$ and so $u \in \mathrm{Span}\,(v_1, \ldots, v_m)$. Conversely, if $u \in \mathrm{Span}\,(v_1, \ldots, v_m)$, so that $u = \sum x_i\, v_i$ for suitable scalars $x_1, \ldots, x_m$ then, as we have seen before, $x_i = \langle u, v_i\rangle$ and $\langle u, u\rangle = \sum |x_i|^2$, so $\langle u, u\rangle = \sum_{i=1}^m |\langle u, v_i\rangle|^2$.

Note that in the real case absolute values are not needed on the left side of the inequality. It states that $\sum_1^m \langle u, v_i\rangle^2 \leqslant ||u||^2$. Also, complex conjugation is not needed in the proof. Nor is it harmful, though if a friend (such as a tutor or an examiner) specifically asks for a proof of Bessel's Inequality for real inner product spaces then one should expound the proof without it.

Example: $V = \mathbb{R}^n$ with the usual inner product; $v_k = e_k$ for $k = 1, 2, \ldots, m$; $u = (x_1, x_2, \ldots, x_n)^{\mathrm{tr}}$. In this case $\langle u, v_i\rangle = x_i$ and Bessel's Inequality tells us that

$$\sum_{i=1}^m x_i^2 \leqslant \sum_1^n x_i^2,$$

which is of course obvious, but gives us some insight into what the theorem is saying in its general and abstract setting.

Example: V is the space of continuous functions $f : [0, 1] \to \mathbb{C}$;

$$\langle f, g\rangle = \int_0^1 f(t)\, \overline{g(t)}\, \mathrm{d}t\,; \qquad v_k(t) = \mathrm{e}^{2\pi ikt} \text{ for } k \in \mathbb{Z}.$$

If $r, s \in \mathbb{Z}$ then

$$\langle v_r, v_s\rangle = \int_0^1 \mathrm{e}^{2\pi\mathrm{i}(r-s)t}\, \mathrm{d}t = \begin{cases} 1 & \text{if } r = s, \\ 0 & \text{if } r \neq s, \end{cases}$$

so $\{v_k(t) \mid k \in \mathbb{Z}\}$ is an orthonormal set in V. For $f \in V$ define

$$c_k := \langle f, e_k\rangle = \int_0^1 f(t)\, \mathrm{e}^{-2\pi\mathrm{i}kt}\, \mathrm{d}t\,.$$

Then Bessel's Inequality tells us that for $m, n \in \mathbb{N}$

$$\sum_{-m}^{n} |c_k|^2 \leqslant \int_0^1 |f(t)|^2 \, \mathrm{d}t,$$

a fact which is of immense importance in Analysis, in the theory of Fourier series in particular.

The Cauchy–Schwarz Inequality

Cauchy–Schwarz Inequality. *Let V be a real or complex inner product space and let $u, v \in V$. Then*

$$|\langle u, v\rangle| \leqslant ||u||.||v||\,.$$

Moreover, equality holds if and only if u, v are linearly dependent.

Proof. If $v = 0$ there is nothing to prove, so we may suppose that $v \neq 0$. Define $v_1 := ||v||^{-1}\, v$. Then, trivially, $\{v_1\}$ is an orthonormal set in V. By Bessel's Inequality $|\langle u, v_1\rangle| \leqslant ||u||$. But $|\langle u, v_1\rangle| = ||v||^{-1}\langle u, v\rangle$. Since $||v|| > 0$ the inequality can be multiplied by $||v||$ and we get that $|\langle u, v\rangle| \leqslant ||u||.||v||$, as required.

If $v = 0$ then equality holds and u, v are linearly dependent. Otherwise equality holds if and only if $u \in \operatorname{Span}(v_1)$, that is, if and only if u, v are linearly dependent.

Classic alternative proof for real inner product spaces. Suppose that V is a real inner product space and let $u, v \in V$. If $v = 0$ the result holds for trivial reasons, so we suppose that $v \neq 0$. For $x \in \mathbb{R}$ define

$$f(x) := ||u - x\, v||^2 = \langle u - x\, v, u - x\, v\rangle = ||u||^2 - 2\, x\, \langle u, v\rangle + x^2\, ||v||^2.$$

Thus $f(x)$ is a quadratic function of x and it is positive semi-definite in the sense that $f(x) \geqslant 0$ for all $x \in \mathbb{R}$. Therefore its discriminant is $\leqslant 0$, and so $\langle u, v\rangle^2 \leqslant ||u||^2\, ||v||^2$, that is, $|\langle u, v\rangle| \leqslant ||u||.||v||$, as required.

Worked example [FHS 1998, Paper a1, Qn 3].

> Let V be a vector space over the complex numbers. Define what is meant by the statement that V has an inner product over $\mathbb{C}$, and explain how this can be used to define the norm $||v||$ of a vector v in V.
>
> Prove *Bessel's Inequality*, that if $\{u_1, \ldots, u_n\}$ is an orthonormal set in V and $v \in V$ then
>
> $$\sum_{i=1}^{n} |\langle u_i, v\rangle|^2 \leqslant ||v||^2,$$
>
> where $\langle\, ,\, \rangle$ denotes the inner product. Deduce, or prove otherwise, the Cauchy–Schwarz Inequality,
>
> $$|\langle u, v\rangle| \leqslant ||u||\,||v|| \text{ for any } u, v \in V.$$
>
> (i) Show that if $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are any complex numbers then
>
> $$\Big|\sum a_i b_i\Big|^2 \leqslant \Big(\sum |a_i|^2\Big)\Big(\sum |b_i|^2\Big).$$

(ii) Show that if $a_1, \ldots, a_n$ are strictly positive real numbers then

$$\Big(\sum a_i\Big)\Big(\sum \frac{1}{a_i}\Big) \geqslant n^2.$$

(iii) Show that if f, g are continuous real-valued functions on an interval $[a,b] \subseteq \mathbb{R}$, then

$$\int_a^b |f(x)|^2\,\mathrm{d}x \int_a^b |g(x)|^2\,\mathrm{d}x \geqslant \Big(\int_a^b f(x)\,g(x)\,\mathrm{d}x\Big)^2.$$

Response. The first parts are 'bookwork' done above (but perhaps now is a good moment for you to close this book and try to reconstruct the proofs for yourself).

For (i) take $V := \mathbb{C}^n$, the space of $1 \times n$ column vectors with its usual hermitian inner product. Take $u := (a_1, \ldots, a_n)^{\mathrm{tr}}$ and $v := (\overline{b_1}, \ldots, \overline{b_n})^{\mathrm{tr}}$ in the Cauchy–Schwarz Inequality. Since $\langle u, v\rangle = u^{\mathrm{tr}}\,\bar{v} = \sum a_i\, b_i$, $||u||^2 = \sum |a_i|^2$ and $||v||^2 = \sum |b_i|^2$, this theorem tells us that $|\sum a_i\, b_i|^2 \leqslant (\sum |a_i|^2)\,(\sum |b_i|^2)$, as required.

Now for (ii) replace a_i and b_i in (i) with $a_i^{1/2}$ and $a_i^{-1/2}$ respectively to see that

$$\Big(\sum_{i=1}^{n} 1\big)^2 \leqslant \Big(\sum_{i=1}^{n} a_i\Big)\Big(\sum_{i=1}^{n} a_i^{-1}\Big),$$

that is, $(\sum a_i)\,\sum 1/a_i) \geqslant n^2$.

For (iii) take V to be the vector space of complex valued continuous functions on $[a,b]$ with inner product $\langle f, g\rangle := \int_a^b f(t)\,\overline{g(t)}\,\mathrm{d}t$. It needs to be checked that this does define an inner product on V [bookwork]. Then the Cauchy–Schwarz inequality says that for any $f, g \in V$, $(\int_a^b |f(x)|^2\,\mathrm{d}x)\,(\int_a^b |g(x)|^2\,\mathrm{d}x) \geqslant (\int_a^b f(x)\,\overline{g(x)}\,\mathrm{d}x)^2$. Specialising to the case where f, g are real-valued, we get the required inequality

$$\Big(\int_a^b f(x)^2\,\mathrm{d}x\Big)\Big(\int_a^b g(x)^2\,\mathrm{d}x\Big) \geqslant \Big(\int_a^b f(x)\,g(x)\,\mathrm{d}x\Big)^2.$$

[Comment: note that in the statement of part (iii) of the question the modulus bars are unnecessary. Also, it seems a bit odd to define V in this last part to consist of complex-valued functions on $[a,b]$ when the question asks about real-valued functions, but since the first part asks for a proof of Bessel's Inequality and the Cauchy–Schwarz Inequality for complex inner product spaces, that is what we have available.]

Isometries of inner product spaces

Let V be a real or complex inner product space of finite dimension n. An *isometry* of V is a linear transformation $P : V \to V$ such that $\langle P\,u, P\,v\rangle = \langle u, v\rangle$ for all $u, v \in V$.

Note 1. An isometry is invertible. For, if P is an isometry of V and $u \in \operatorname{Ker} P$ then $\langle u, u\rangle = \langle P\,u, P\,u\rangle = 0$ and so $u = 0$. Thus $\operatorname{Ker} P = \{0\}$ and so P is injective and since V is finite-dimenisonal it must also be surjective, hence invertible.

Note 2. It follows that the isometries of our inner product space V form a group: certainly I is an isometry; if P is an isometry then P^{-1} is an isometry; if P_1, P_2 are isometries then so is $P_2 \circ P_1$.

Note 3. In the real case an isometry is known as an *orthogonal* transformation; the group is the *orthogonal* group denoted $\mathrm{O}(V)$. The group $\mathrm{O}(\mathbb{R}^n)$ is often denoted $\mathrm{O}(n)$, sometimes $\mathrm{O}(n,\mathbb{R})$ or $\mathrm{O}_n(\mathbb{R})$.

Note 4. In fact, $\mathrm{O}(\mathbb{R}^n) = \{A \in \mathrm{M}_{n\times n}(\mathbb{R}) \mid A^{-1} = A^{\mathrm{tr}}\}$. For,

$$\langle A\,u, A\,v\rangle = (A\,u)^{\mathrm{tr}}\,(A\,v) = u^{\mathrm{tr}}\,A^{\mathrm{tr}}\,A\,v$$

and this is $u^{\mathrm{tr}}\,v$ for all $u, v \in \mathbb{R}^n$ if and only if $A^{\mathrm{tr}}\,A = I$.

Note 5. Thus an $n \times n$ matrix A with real entries is orthogonal if and only if the columns of A form an orthonormal basis for $\mathbb{R}^n$.

Note 6. In the complex case an isometry is known as a *unitary* transformation; the group is the *unitary* group $\mathrm{U}(V)$. The group $\mathrm{U}(\mathbb{C}^n)$ is often denoted $\mathrm{U}(n)$, sometimes $\mathrm{U}(n,\mathbb{C})$ or $\mathrm{U}_n(\mathbb{C})$.

Note 7. $\mathrm{U}(\mathbb{C}^n) = \{A \in \mathrm{M}_{n\times n}(\mathbb{C}) \mid A^{-1} = \bar{A}^{\mathrm{tr}}\}$. The proof is similar to the real case and is offered as an exercise:

Exercise 45. Prove that $\mathrm{U}(\mathbb{C}^n) = \{A \in \mathrm{M}_{n\times n}(\mathbb{C}) \mid A^{-1} = \bar{A}^{\mathrm{tr}}\}$.

Note 8. Thus an $n \times n$ matrix A with complex entries is unitary if and only if the columns of A form an orthonormal basis for $\mathbb{C}^n$.

Exercise 46. Let X be a non-singular $n \times n$ matrix over $\mathbb{R}$.

(i) Use the Gram–Schmidt orthogonalisation process to show that there exist $n \times n$ matrices P, U over $\mathbb{R}$ such that U is upper triangular with positive entries on its main diagonal, P is orthogonal, and $X = PU$.

(ii) Suppose that $X = PU = QV$, where U, V are upper triangular with positive entries on their main diagonals, and P, Q are orthogonal. How must Q be related to P, and V to U?

(iii) What are the corresponding results for non-singular matrices over $\mathbb{C}$?

Observation. *the group* $\mathrm{O}(n)$ *is a closed bounded subset of* $\mathrm{M}_{n\times n}(\mathbb{R})$. *Similarly,* $\mathrm{U}(n)$ *is a closed bounded subset of* $\mathrm{M}_{n\times n}(\mathbb{C})$.

The group $\mathrm{O}(n)$ is closed in $\mathrm{M}_{n\times n}(\mathbb{R})$ because if A_m are orthogonal matrices and $A = \lim_{m\to\infty} A_m$ then $A^{\mathrm{tr}}\,A = \lim A_m^{\mathrm{tr}} \lim A_m = \lim(A_m^{\mathrm{tr}} A_m) = \lim I = I$, which shows that A is orthogonal. It is bounded because if $A = (a_{ij}) \in \mathrm{O}(n)$ then for each i we have $\sum_j a_{ij}^2 = 1$ and so $|a_{ij}| \leqslant 1$ for all relevant i, j. The proof for $\mathrm{U}(n)$ is almost the same.

Representation of linear functionals on an inner product space

We come now to a very useful result often known as the Riesz Representation Lemma (although that name more properly belongs to a theorem in functional analysis about infinite dimensional spaces—so-called Hilbert spaces). This will be the foundation for our treatment of adjoint transformations in the fourth and final part of these notes.

THEOREM. *Let V be a finite-dimensional real or complex inner product space. For every $f \in V'$ there exists $v_f \in V$ such that*

$$f(u) = \langle u, v_f \rangle \quad \textit{for all } u \in V\,.$$

Moreover, v_f is unique.

Proof. Let $f \in V'$. If $f = 0$ then we take $v_f := 0$ as we clearly must. So suppose now that $v \neq 0$. Where should we seek v_f? Well, let $U := \operatorname{Ker} f$. Since $f : V \to F$ (where F is $\mathbb{R}$ or $\mathbb{C}$) is linear and non-trivial it is surjective and so $\operatorname{Ker} U = \dim V - 1$ by the Rank-Nullity Theorem. We certainly want that $\langle u, v_f \rangle = 0$ for all $u \in U$, that is, we should seek v_f in $U^\perp$. Now $U^\perp$ is 1-dimensional. We can choose $v \in U^\perp$ such that $||v|| = 1$, and then we'll want $v_f = \lambda v$ for a suitable scalar λ. And λ can be determined from the requirement that $f(v) = \langle v, v_f \rangle = \langle v, \lambda v \rangle = \bar{\lambda} \langle v, v \rangle = \bar{\lambda}$. Thus we define $v_f := \overline{f(v)}\, v$ (but note that in the real case complex conjugation, though harmless, is not needed). We check: if $u \in V$ then, since $V = \operatorname{Span}(v) \oplus U$, there is a unique vector $w \in U$ and there is a unique scalar α such that $u = \alpha v + w$; then $f(u) = \alpha f(v) + f(w) = \alpha \lambda$, while

$$\langle u, f_v \rangle = \langle \alpha v + w, \bar{\lambda} v \rangle = \alpha \langle v, \bar{\lambda} v \rangle + \langle w, \bar{\lambda} v \rangle = \alpha \lambda \langle v, v \rangle = \alpha \lambda\,.$$

Thus $f(u) = \langle u, f_v \rangle$, as required.

For uniqueness note that if $v_1, v_2 \in V$ are such that $f(u) = \langle u, v_1 \rangle = \langle u, v_1 \rangle$ for all $u \in V$ then $\langle u, v_0 \rangle = 0$ for all $u \in V$ where $v_0 := v_1 - v_2$. In particular, $\langle v_0, v_0 \rangle = 0$ and so $v_0 = 0$, that is $v_1 = v_2$. Therefore the vector v_f corresponding to the linear functional f is unique.

NOTE. The map $f \mapsto v_f$ from V' to V is linear in the real case, semilinear in the complex case:

EXERCISE 47. Prove this. That is, show that if $F = \mathbb{R}$ then in the theorem above $v_{\alpha f + \beta g} = \alpha v_f + \beta v_g$ for all $f, g \in V'$ and all $\alpha, \beta \in \mathbb{R}$, and that if $F = \mathbb{C}$ then $v_{\alpha f + \beta g} = \bar{\alpha} v_f + \bar{\beta} v_g$ for all $f, g \in V'$ and all $\alpha, \beta \in \mathbb{C}$.

Further exercises III

EXERCISE 48 [A former FHS question (corrected)]. Let V be a finite-dimensional inner-product space over $\mathbb{R}$. Prove that if $w_1, \ldots, w_m$ is an orthonormal set of vectors in V then

$$||v||^2 \geqslant \sum_1^m (v, w_i)^2 \qquad \text{for all } v \in V\,. \qquad \text{[Bessel's inequality]}$$

Now suppose that $u_1, \ldots, u_l$ are unit vectors in V. Prove that

$$||v||^2 = \sum_1^l (v, u_i)^2 \quad \text{for all } v \in V$$

if and only if $u_1, \ldots, u_l$ is an orthonormal basis for V.

EXERCISE 49. Prove that if $z_1, \ldots, z_n \in \mathbb{C}$ then $\left|\sum_{i=1}^{n} z_i\right|^2 \leqslant n \sum_{i=1}^{n} |z_i|^2$. When does equality hold?

EXERCISE 50 [Part of FHS 2005, AC2, 2]. Prove that for any positive real number a

$$\sum_{i=1}^{k} a^i \sum_{i=1}^{k} a^{-i} \geqslant k^2.$$

When does equality hold?

Part IV: Adjoints of linear transformations on finite-dimensional inner product spaces

In this fourth and final part of these notes we study linear transformations of finite-dimensional real or complex inner product spaces. What we are aiming for is the so-called Spectral Theorem for self-adjoint transformations.

Adjoints of linear transformations

We begin with the definition of adjoints.

THEOREM. *Let V be a finite-dimensional inner product space. For each linear transformation $T : V \to V$ there is a unique linear transformation $T^* : V \to V$ such that*

$$\langle T\,u, v\rangle = \langle u, T^* v\rangle \quad \textit{for all } u, v \in V\,.$$

Proof. For each fixed $v \in V$ the map $u \mapsto \langle T\,u, v\rangle$ is a linear functional on V because

$$\langle T(\alpha_1 u_1 + \alpha_2 u_2), v\rangle = \langle \alpha_1 T\,u_1 + \alpha_2 T\,u_2, v\rangle = \alpha_1 \langle T\,u_1, v\rangle + \alpha_2 \langle T\,u_2, v\rangle.$$

By the representation theorem (see p. 45) there exists $w \in V$ such that

$$\langle T\,u, v\rangle = \langle u, w\rangle \text{ for all } u \in U\,.$$

This vector w depends on and is uniquely determined by v so we write it as $T^* v$, where $T^* : V \to V$. Thus to prove the theorem we need to prove that T^* is linear. Well, let $v_1, v_2 \in V$ and let $\alpha_1, \alpha_2 \in F$, where F is the field of scalars ($\mathbb{R}$ or $\mathbb{C}$). Then for any $u \in V$, using the definition of T^* and the bilinearity or sesquilinearity of the inner product, we find that

$$\begin{aligned}\langle u, T^*(\alpha_1 v_1 + \alpha_2 v_2)\rangle &= \langle T\,u, \alpha_1 v_1 + \alpha_2 v_2\rangle \\ &= \overline{\alpha}_1\, \langle T\,u, v_1\rangle + \overline{\alpha}_2\, \langle T\,u, v_2\rangle \\ &= \overline{\alpha}_1\, \langle u, T^* v_1\rangle + \overline{\alpha}_2\, \langle u, T^* v_2\rangle \\ &= \langle\, u, \alpha_1 T^* v_1 + \alpha_2 T^* v_2\rangle.\end{aligned}$$

We have seen before that if $\langle u, w_1\rangle = \langle u, w_2\rangle$ for all $u \in V$ then $w_1 = w_2$. It follows that

$$T^*(\alpha_1 v_1 + \alpha_2 v_2) = \alpha_1 T^* v_1 + \alpha_2 T^* v_2\,,$$

and therefore T^* is linear, as required.

EXAMPLE. Let $V := \mathbb{R}^n$ with its usual inner product $\langle u, v\rangle = u^{\mathrm{tr}} v$. Suppose $T : V \to V$ is given by $T : v \mapsto Av$ where $A \in \mathrm{M}_{n\times n}(\mathbb{R})$. Then $T^* : v \mapsto A^{\mathrm{tr}} v$. For,

$$\langle T\,u, v\rangle = \langle A\,u, v\rangle = (A\,u)^{\mathrm{tr}} v = (u^{\mathrm{tr}} A^{\mathrm{tr}}) v = u^{\mathrm{tr}}\,(A^{\mathrm{tr}} v) = \langle u, A^{\mathrm{tr}} v\rangle.$$

Example. Let $V := \mathbb{C}^n$ with its usual inner product $\langle u, v\rangle = u^{\rm tr}\bar{v}$. Suppose $T : V \to V$, $T : v \mapsto Av$ where $A \in \mathrm{M}_{n\times n}(\mathbb{C})$. Then $T^* : v \mapsto \bar{A}^{\,\rm tr} v$. The proof is very similar to that worked through in the real case.

Theorem. *Let $S : V \to V$, $T : V \to V$ be linear transformations of the real or complex inner product space V. Then:*

$$\text{(i) } (S+T)^* = S^* + T^*; \quad \text{(ii) } (\alpha T)^* = \bar{\alpha} T^*; \quad \text{(iii) } (ST)^* = T^* S^*; \quad \text{(iv) } S^{**} = S.$$

Note: Complex conjugation of α in (ii) is unnecessary, but harmless, in the real case.

Proof. We deal with (ii) and (iii), leaving (i) and (iv) as exercises. For (ii), for all $u, v \in V$,

$$\langle (\alpha T)\, u, v\rangle = \alpha \langle T u, v\rangle = \alpha \langle u, T^* v\rangle = \langle u, \bar{\alpha} T^* v\rangle = \langle u, (\bar{\alpha} T^*)\, v\rangle,$$

and therefore (by uniqueness) $(\alpha T)^* = \bar{\alpha} T^*$.

For (iii), for all $u, v \in V$,

$$\langle (S T)\, u, v\rangle = \langle T u, S^* v\rangle = \langle u, T^*(S^* v)\rangle = \langle u, (T^* S^*)\, v\rangle,$$

and therefore $(ST)^* = T^* S^*$, as required.

Exercise 51. Show that if $S : V \to V$, $T : V \to V$ are linear transformations of the real or complex inner product space V then: $(S+T)^* = S^* + T^*$ and $S^{**} = S$.

Exercise 52. Show that if $T : V \to V$ is a linear transformation of a real or complex inner product space V and $f \in F[x]$ (where F is $\mathbb{R}$ or $\mathbb{C}$ appropriately) then $f(T)^* = f(T^*)$ in the real case and $f(T)^* = \overline{f}(T^*)$ in the complex case (where if $f(x) = \sum a_i x^i$ then $\overline{f}(x) = \sum \overline{a_i}\, x^i$).

Theorem. *Let V be a finite-dimensional real or complex inner product space, let $T : V \to V$ be a linear transformation, let $\{e_1, e_2, \ldots, e_n\}$ be an orthonormal basis of V, and let A, A^* be the matrices of T and T^* respectively with respect to this basis. Then $A^* = \bar{A}^{\rm tr}$.*

Note. Conjugation is of course not needed in the real case.

Proof. The matrix A is (a_{ij}), say, where the coefficients are defined by the equations $T e_j = \sum_i a_{ij}\, e_i$. Similarly, if $A^* = (b_{ij})$ then the coefficients b_{ij} are defined by the equations $T^* e_j = \sum_i b_{ij}\, e_i$. Now $\langle T e_p, e_q\rangle = \langle e_p, T^* e_q\rangle$ for all relevant p, q by definition of the adjoint. But

$$\langle T e_p, e_q\rangle = \langle\, \textstyle\sum_i a_{ip}\, e_i\,,\, e_q\rangle = a_{qp}\,,$$

while

$$\langle e_p, T^* e_q\rangle = \langle e_p, \textstyle\sum_i b_{iq}\, e_i\rangle = \overline{b_{pq}}\,.$$

Therefore $b_{pq} = \overline{a_{qp}}$, that is, $A^* = \bar{A}^{\rm tr}$, as the theorem states.

Next we investigate the kernel and the image of an adjoint transformation.

Theorem. *Let V be a finite-dimensional real or complex inner product space and let $T : V \to V$ be a linear transformation. Then* $\operatorname{Ker} T^* = (\operatorname{Im} T)^\perp$ *and* $\operatorname{Im} T^* = (\operatorname{Ker} T)^\perp$.

Proof. Recall that if $v \in V$ then $\langle u, v\rangle = 0$ for all $u \in V$ if and only if $v = 0$. Therefore

$$\begin{aligned}\operatorname{Ker} T^* &= \{v \in V \mid T^* v = 0\}\\ &= \{v \in V \mid \langle u, T^* v\rangle = 0 \text{ for all } u \in V\}\\ &= \{v \in V \mid \langle T u, v\rangle = 0 \text{ for all } u \in V\}\\ &= \{v \in V \mid \langle w, v\rangle = 0 \text{ for all } w \in \operatorname{Im} T\}\\ &= (\operatorname{Im} T)^\perp.\end{aligned}$$

Now if $v \in \operatorname{Im} T^*$ then $v = T^* w$ for some $w \in V$ and so if $u \in \operatorname{Ker} T$ then

$$\langle u, v\rangle = \langle u, T^* w\rangle = \langle T u, w\rangle = \langle 0, w\rangle = 0.$$

This shows that $\operatorname{Im} T^* \leqslant (\operatorname{Ker} T)^\perp$. Compare dimensions: if $n := \dim V$ then

$$\begin{aligned}\dim \operatorname{Im} T^* &= n - \dim(\operatorname{Ker} T^*) && \text{[Rank-Nullity theorem]}\\ &= n - \dim(\operatorname{Im} T)^\perp && \text{[since } \operatorname{Ker} T^* = (\operatorname{Im} T)^\perp\text{]}\\ &= \dim(\operatorname{Im} T) && \text{[theorem on p. 36]}\\ &= n - \dim(\operatorname{Ker} T) && \text{[Rank-Nullity theorem]}\\ &= \dim(\operatorname{Ker} T)^\perp && \text{[theorem on p. 36]}\end{aligned}$$

and therefore $\operatorname{Im} T^* = (\operatorname{Ker} T)^\perp$, and the proof of the theorem is complete.

Exercise 53 [Part of an old FHS question]. Consider the vector space of real-valued polynomials of degree $\leqslant 1$ in a real variable t, equipped with the inner product $\langle f, g\rangle = \int_0^1 f(t)\, g(t)\, \mathrm{d}t$. Let D be the operation of differentiation with respect to t. Find the adjoint D^*.

Exercise 54 [Compare FHS 2005, AC1, Qn 2]. In the previous exercise, how does D^* change if the inner product is changed to $\int_{-1}^1 f(t)\, g(t)\, \mathrm{d}t$?

Self-adjoint linear transformations

The transformation $T : V \to V$, where V is a finite-dimensional real or complex inner product space, is said to be *self-adjoint* if $T^* = T$. So T is self-adjoint if and only if $\langle T u, v\rangle = \langle u, T v\rangle$ for all $u, v \in V$.

Important example. If $S : V \to V$ is any linear transformation then $S S^*$ and $S^* S$ are self-adjoint. For $(S S^*)^* = (S^*)^* S^* = S S^*$ and $(S^* S)^* = S^* (S^*)^* = S^* S$ by clauses (iii), (iv) of the theorem on p. 48.

Lemma. *Let V be a finite-dimensional real or complex inner product space. If $T : V \to V$ is self-adjoint and U is a T-invariant subspace (that is, recall, $TU \leqslant U$) then also $U^\perp$ is T-invariant.*

Proof. Let U be a T-invariant subspace of V and let $v \in U^\perp$. Then for any $u \in U$

$$\begin{aligned}\langle u, T v\rangle &= \langle T u, v\rangle \qquad &&[\text{since } T \text{ is self-adjoint}]\\ &= 0 &&[\text{since } Tu \in U \text{ and } v \in U^\perp].\end{aligned}$$

Thus $Tv \in U^\perp$ and so $U^\perp$ is T-invariant.

Exercise 55 [Part of FHS 2003, A1, Qn 4]. Let V be a 3-dimensional complex inner product space, let $\{e_1,\, e_2,\, e_3\}$ be an orthonormal basis for V, and let $S : V \to V$ be a linear transformation such that

$$S(e_1) = e_1\,, \quad S(e_1 + e_2) = 2e_1 + 2e_2\,, \quad S(e_1 + e_3) = 0\,.$$

Is S self-adjoint? Justify your answer.

Exercise 56 [Part of FHS 1990, A1, Qn 4]. Let V be the vector space of all $n \times n$ real matrices with the usual addition and scalar multiplication. For A, B in V, let $\langle A, B\rangle = \mathrm{Trace}(AB^{\mathrm{tr}})$, where B^{tr} denotes the transpose of B.

(i) Show that this defines an inner product on V.

(ii) Let P be an invertible $n\times n$ matrix and let $\theta : V \to V$ be the linear transformation given by $\theta(A) = P^{-1}AP$. Find the adjoint θ^* of θ.

(iii) Prove that θ is self-adjoint if and only if P is either symmetric or skew-symmetric (P is *skew-symmetric* if $P^{\mathrm{tr}} = -P$).

Eigenvalues and eigenvectors of self-adjoint linear transformations

Theorem. *Let V be a finite-dimensional real or complex inner product space and let $T : V \to V$ be a self-adjoint linear transformation. Then all eigenvalues of T are real—that is, $c_T(x)$ may be factorised as a product of linear factors in $\mathbb{R}[x]$.*

Proof. Let A be the matrix of T with respect to some orthonormal basis of V. We know that the matrix of T^* with respect to that same basis is $\bar{A}^{\mathrm{tr}}$ and since $T = T^*$ it follows that $A = \bar{A}^{\mathrm{tr}}$. Now let λ be an eigenvalue of T, that is, a root of $c_T(x)$ or, which is the same thing, of $c_A(x)$. Of course, by the so-called Fundamental Theorem of Algebra, even if V is a real inner product space we'll know only that $\lambda \in \mathbb{C}$. But then there exists $v \in \mathbb{C}^n \setminus \{0\}$ such that $Av = \lambda v$. Now we compute the standard inner product of v and Av:

$$\langle v, Av\rangle = v^{\mathrm{tr}}\,(\overline{Av}) = v^{\mathrm{tr}} A^{\mathrm{tr}}\,\bar{v} = (Av)^{\mathrm{tr}}\,\bar{v} = \lambda\, v^{\mathrm{tr}}\,\bar{v} = \lambda\,||v||^2.$$

On the other hand,

$$\langle v, Av\rangle = \langle v, \lambda v\rangle = v^{\mathrm{tr}}\,\overline{\lambda v} = \bar{\lambda}\, v^{\mathrm{tr}}\,\bar{v} = \bar{\lambda}\,||v||^2.$$

Therefore $\lambda\,||v||^2 = \bar{\lambda}\,||v||^2$ and since $||v||^2 > 0$ we must have $\lambda = \bar{\lambda}$, that is, λ is real.

Theorem. *Let V be a finite-dimensional real or complex inner product space and let $T : V \to V$ be a self-adjoint linear transformation. If u, v are eigenvectors of T for distinct eigenvalues λ, μ then $\langle u, v\rangle = 0$.*

Proof. We have $Tu = \lambda u$, $Tv = \mu v$ and $\lambda \neq \mu$. Therefore

$$\lambda\langle u, v\rangle = \langle Tu, v\rangle = \langle u, Tv\rangle = \langle u, \mu v\rangle = \bar{\mu}\,\langle u, v\rangle.$$

We have just seen that $\bar{\mu} = \mu$ and so we see that $\lambda\langle u, v\rangle = \mu\langle u, v\rangle$. Since $\lambda \neq \mu$ we must have $\langle u, v\rangle = 0$ as the theorem states.

Diagonalisability and the spectral theorem for self-adjoint linear transformations

One of the most important theorems in the area tells us that a self-adjoint transformation, or equivalently, a symmetric or conjugate-symmetric matrix, is diagonalisable, and moreover, the diagonalisation can be accomplished with an orthogonal or unitary transformation. We give this theorem in three different forms.

THEOREM. *Let V be a finite-dimensional inner product space over $\mathbb{R}$ or $\mathbb{C}$ and let $T : V \to V$ be a self-adjoint linear transformation. There is an orthonormal basis of V consisting of eigenvectors of T.*

Proof. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T, for $1 \leqslant i \leqslant k$ let

$$V_i := \{v \in V \mid Tv = \lambda_i v\},$$

and let $U := V_1 + \cdots + V_k$, so that U is the subspace of V spanned by the eigenvectors of T. We aim to show that $U = V$. To this end, let $W := U^{\perp}$. We know that W is T-invariant (see p. 49). If W were non-trivial then it would contain an eigenvector of T, but all of these lie in U. Therefore $W = \{0\}$, and so $U = V$.

We know from the previous theorems that if $v_i \in V_i$ then $v_1, \ldots, v_k$ is an orthogonal set of vectors in V and is therefore linearly independent (see p. 35). Therefore $U = V_1 \oplus \cdots \oplus V_k$. Now V_i is a finite-dimensional inner product space and so it has an orthonormal basis B_i, and of course B_i consists of eigenvectors of T for eigenvalue λ_i. Thus, since members of B_i are orthogonal to members of B_j when $i \neq j$, if $B := B_1 \cup \cdots \cup B_k$ then B is an orthonormal basis of V consisting of eigenvalues of T.

THEOREM.

(1) *If $A \in \mathrm{M}_{n\times n}(\mathbb{R})$ and $A^{\mathrm{tr}} = A$ then there exists $U \in \mathrm{O}(n)$ and there exists a diagonal matrix $D \in \mathrm{M}_{n\times n}(\mathbb{R})$ such that $U^{-1}AU = D$. (Recall: $U \in \mathrm{O}(n)$ means that $U^{-1} = U^{\mathrm{tr}}$.)*

(2) *If $A \in \mathrm{M}_{n\times n}(\mathbb{C})$ and $\bar{A}^{\mathrm{tr}} = A$ then there exists $U \in \mathrm{U}(n)$ and there exists a diagonal matrix $D \in \mathrm{M}_{n\times n}(\mathbb{R})$ such that $U^{-1}AU = D$. (Recall: $U \in \mathrm{U}(n)$ means that $U^{-1} = \bar{U}^{\mathrm{tr}}$.)*

The force of this theorem is that real symmetric matrices can be diagonalised by an orthogonal change of basis—that is, by a rotation or a reflection (though in fact one can always do it with a rotation). Similarly, a conjugate-symmetric complex matrix can be diagonalised by a unitary transformation. This is simply the special case of the previous theorem where V is $\mathbb{R}^n$ or $\mathbb{C}^n$ and $T : v \mapsto Av$.

As preparation for our third version of the diagonalisablity theorem, the so-called Spectral Theorem, we need to examine self-adjoint projection operators

Lemma. *Let $P : V \to V$ be a projection operator (idempotent) where V is a finite-dimensional real or complex inner product space. Then P is self-adjoint if and only if* $\operatorname{Im} P = (\operatorname{Ker} P)^{\perp}$.

Proof. Let $U := \operatorname{Im} P$ and $W := \operatorname{Ker} P$, so that $V = U \oplus W$ and P is the projection onto U along W. Suppose first that P is self-adjoint. If $u \in U$, $w \in W$, then

$$\langle u, w\rangle = \langle P u, w\rangle = \langle u, P w\rangle = \langle u, 0\rangle = 0,$$

and so $W \leqslant U^{\perp}$. Comparing dimensions we see that $W = U^{\perp}$.

Now suppose that $W = U^{\perp}$. For v_1, $v_2 \in V$ write $v_1 = u_1 + w_1$, $v_2 = u_2 + w_2$, where $u_1, u_2 \in U$, $w_1, w_2 \in W$. Then

$$\langle P v_1, v_2\rangle = \langle u_1, u_2 + w_2\rangle = \langle u_1, u_2\rangle = \langle u_1 + w_1, u_2\rangle = \langle v_1, P v_2\rangle,$$

and this shows that P is self-adjoint.

The Spectral Theorem. *Let V be a finite-dimensional real or complex inner product space and let $T : V \to V$ be a self-adjoint linear transformation. If the distinct eigenvalues of T are $\lambda_1, \ldots, \lambda_k$ then $\lambda_i \in \mathbb{R}$ for $1 \leqslant i \leqslant k$ and there are uniquely determined self-adjoint projection operators $P_i : V \to V$ such that $P_i P_j = 0$ whenever $i \neq j$ and*

$$\sum_i P_i = I \quad \textit{and} \quad T = \sum_i \lambda_i P_i\,.$$

Note that $P_1, \ldots, P_k$ is a partition of the identity (see p. 9).

Again, this is simply a restatement of previous theorems. The projection operator P_i is the projection of V onto V_i along $\bigoplus_{j \neq i} V_j$, where V_i is the eigenspace for eigenvalue λ_i as in the proof of the first of these three theorems.

Worked example [FHS 1999, Paper a1, Qn 4].

> Let V be a finite-dimensional complex inner product space. Let A be a linear transformation of V; define the *adjoint* A^* of A and show that it is unique. Show also that $\operatorname{Ker}(A) = \operatorname{Ker}(A^* A)$. If A, B are linear transformations of V show that $(A B)^* = B^* A^*$.
>
> What does it mean to say that S is *self-adjoint*? Suppose S is self-adjoint. Show that
>
> $$(tr(S))^2 \leqslant r(S) tr(S^2)\,,$$
>
> where r denotes the rank, and tr denotes the trace of a linear transformation of V. [You may use without proof that a self-adjoint linear transformation has only real eigenvalues and that there exists an orthonormal basis of eigenvectors of V.]
>
> Deduce that for an arbitrary linear transformation A
>
> $$(tr(A^* A))^2 \leqslant r(A) tr((A^* A)^2)\,.$$

Response. Definition of adjoint and proof that it is unique is bookwork done above. To see that $\mathrm{Ker}\,(A) = \mathrm{Ker}\,(A^*A)$ note first that if $u \in \mathrm{Ker}\,(A)$ then certainly $(A^*A)\,u = A^*(A\,u) = 0$, so $\mathrm{Ker}\,(A) \leqslant \mathrm{Ker}\,(A^*A)$. On the other hand, if $u \in \mathrm{Ker}\,(A^*A)$ then

$$\langle A\,u, A\,u\rangle = \langle u, A^*A\,u\rangle = \langle u, 0\rangle = 0$$

and so $A\,u = 0$. This shows that $\mathrm{Ker}\,(A^*A) \leqslant \mathrm{Ker}\,(A)$ and so $\mathrm{Ker}\,(A^*A) = \mathrm{Ker}\,(A)$.

The definition of *self-adjoint* is bookwork given above. So now suppose that S *is* self-adjoint. We know (and the examiner lets us quote) that there is a basis of V with respect to which the the matrix of S is partitioned as $\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$ where D is a diagonal matrix which in a self-explanatory notation may be written $\mathrm{Diag}\,(\lambda_1, \ldots, \lambda_r)$ where $\lambda_1, \ldots, \lambda_r$ are non-zero real numbers (not necessarily distinct). Thus in this notation $r(S) = r$. With respect to that same basis the matrix of S^2 is $\begin{pmatrix} D^2 & 0 \\ 0 & 0 \end{pmatrix}$. Since $D^2 = \mathrm{Diag}\,(\lambda_1^2, \ldots, \lambda_r^2)$ we see that $tr(S^2) = \sum \lambda_i^2$. [In fact, this is true for *any* linear transformation as one sees from consideration of its triangular form.] Taking vectors $u := (\lambda_1, \ldots, \lambda_r)^{\mathrm{tr}} \in \mathbb{R}^r$ and $v := (1, \ldots, 1)^{\mathrm{tr}} \in \mathbb{R}^r$ and applying the Cauchy–Schwarz Inequality in the form $\langle u, v\rangle^2 \leqslant ||u||^2\,||v||^2$ to them we see that $(\sum \lambda_i)^2 \leqslant r \sum \lambda_i^2$, that is, $(tr(S))^2 \leqslant r(S)\,tr(S^2)$, as required.

For the last part let $A : V \to V$ be any linear transformation and define $S := A^*A$. Then S is self-adjoint (since $S^* = A^*A^{**} = A^*A = S$) and from the first part of the question $\mathrm{Ker}\,S = \mathrm{Ker}\,A$. By the Rank-Nullity Theorem therefore $r(S) = r(A)$. Applying what has just been proved we see that $(tr(A^*A))^2 \leqslant r(A)\,tr((A^*A)^2)$, as required.

Exercise 57 [FHS 1988, A1, Qn 3]. Let V be a finite-dimensional real inner product space, $T : V \to V$ a linear transformation and T^* its adjoint. Prove the following:

(i) $(\mathrm{im}\,T)^\perp = \ker T^*$

(ii) $\ker T^*T = \ker T$

(iii) $\dim\ker\,(TT^*) = \dim\ker\,(T^*T)$

(iv) if v is an eigenvector of T^*T with eigenvalue $\lambda \neq 0$, then Tv is an eigenvector of TT^* with eigenvalue λ.

Deduce that there exists an orthogonal linear transformation P such that $P^{-1}TT^*P = T^*T$. [You may assume that any self-adjoint linear transformation has an orthonormal basis of eigenvectors.]

An application: quadratic forms

As an application of the theory of inner product spaces and self-adjoint linear transformations we treat the beginnings of the theory of quadratic forms. A *quadratic form* in n variables $x_1, x_2, \ldots, x_n$ over a field F is a homogeneous quadratic polynomial $\sum a_{ij}\,x_i\,x_j$ where $a_{ij} \in F$ for $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant n$.

Note. If $\mathrm{char}\,F \neq 2$ then we may replace each of a_{ij}, a_{ji} by $\frac{1}{2}(a_{ij} + a_{ji})$. Thus we may (and we always **do**) assume that $a_{ij} = a_{ji}$.

Note. Then $\sum a_{ij}\,x_i\,x_j = x^{\mathrm{tr}}\,A\,x$ where x is the column vector $(x_1, x_2, \ldots, x_n)^{\mathrm{tr}}$ and A is the matrix (a_{ij}), which, it is worth emphasising, now is *symmetric*.

There is a close connection between quadratic and bilinear forms (which, recall, were defined on p. 34).

Lemma. *Suppose that* $\mathrm{char}\,F \neq 2$. *To each quadratic form* $Q(x_1, \ldots, x_n)$ *over* F *there corresponds a unique symmetric bilinear form* $B(u, v)$ *on* F^n *such that*

$$Q(v) = B(v, v).$$

Proof Define $B(u, v) := \frac{1}{2}(Q(u+v) - Q(u) - Q(v))$. Thus if A is the symmetric matrix representing Q then

$$B(u, v) = \frac{1}{2}\Big((u+v)^{\mathrm{tr}}\,A\,(u+v) - u^{\mathrm{tr}}\,A\,u - v^{\mathrm{tr}}\,A\,v\Big) = u^{\mathrm{tr}}\,A\,v\,.$$

The function $(u, v) \mapsto u^{\mathrm{tr}}\,A\,v$ is obviously bilinear (by the distributive laws for matrix multiplication) so B is bilinear. And clearly $B(u, u) = Q(u)$ for all $u \in F^n$.

Uniqueness goes as follows: suppose that $B_1(u, u) = B_2(u, u) = Q(u)$ for all $u \in F^n$, where B_1 and B_2 are symmetric bilinear forms. Now

$$B_1(u+v, u+v) - B_1(u, u) - B_1(v, v) = B_1(u, v) + B_1(v, u) = 2B_1(u, v)$$

since B_1 is symmetric. Similarly, $2B_2(u, v) = B_2(u+v, u+v) - B_2(u, u) - B_2(v, v)$, and it follows (given that $\mathrm{char}\,F \neq 2$) that $B_1 = B_2$.

Real quadratic forms, that is to say, quadratic forms over $\mathbb{R}$ are of particular importance in most branches of mathematics, both pure and applied.

Theorem. *Let* $Q(x_1, \ldots, x_n)$ *be a real quadratic form. There exists an invertible matrix* P *such that if*

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = P \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

then

$$Q(x_1, \ldots, x_n) = X_1^2 + \cdots + X_p^2 - X_{p+1}^2 - \cdots - X_r^2$$

for all $(x_1, \ldots, x_n)^{\mathrm{tr}} \in \mathbb{R}^n$

Proof. Write $Q(x_1, \ldots, x_n) = u^{\mathrm{tr}}\,A\,u$, where $u = (x_1, \ldots, x_n)^{\mathrm{tr}} \in \mathbb{R}^n$ and A is the real symmetric matrix of the quadratic form. By the theorem about diagonalisability of real symmetric matrices (see p. 51) there exists $U \in \mathrm{O}(n)$ and there exists a diagonal matrix $D \in \mathrm{M}_{n\times n}(\mathbb{R})$ such that $U^{-1}AU = D$. Now U is orthogonal and so $U^{-1} = U^{\mathrm{tr}}$. Therefore $U^{\mathrm{tr}}AU = D$. We will see the significance of this in a moment.

We may write $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$, where

$$\lambda_i > 0 \text{ if } 1 \leqslant i \leqslant p, \quad \lambda_i < 0 \text{ if } p+1 \leqslant i \leqslant r, \quad \lambda_i = 0 \text{ if } r+1 \leqslant i \leqslant n.$$

Define

$$\mu_i := \begin{cases} \lambda_i^{-1/2} & \text{if } 1 \leqslant i \leqslant p, \\ (-\lambda_i)^{-1/2} & \text{if } p+1 \leqslant i \leqslant r, \\ 1 & \text{if } r+1 \leqslant i \leqslant n. \end{cases}$$

and $E := \mathrm{Diag}\,(\mu_1, \ldots, \mu_n)$. Clearly, E is invertible and

$$E^{\mathrm{tr}}\, D\, E = \begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $q := r - p$. It follows that

$$(U\,E)^{\mathrm{tr}}\, A\,(U\,E) = E^{\mathrm{tr}}\, D\, E = \begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and so if $P := (U\,E)^{-1}$ and $v := (X_1, \ldots, X_n)^{\mathrm{tr}}$ where $(X_1, \ldots, X_n)^{\mathrm{tr}} := P\,u$ then

$$u^{\mathrm{tr}}\, A\, u = (P^{-1}\, v)^{\mathrm{tr}}\, A\,(P^{-1}\, v) = v^{\mathrm{tr}}\Big((U\,E)^{\mathrm{tr}}\, A\,(U\,E)\Big)\, v$$

that is, $Q(x_1, \ldots, x_n) = X_1^2 + \cdots + X_p^2 - X_{p+1}^2 - \cdots - X_r^2$, as required.

Note 1. The parameter r in the theorem is called the *rank* of Q. It is the matrix rank of A.

Note 2. The number p is also an invariant of Q. Define $q := r - p$. Sometimes p, sometimes $p - q$, sometimes the pair (p, q) is known as the *signature* of Q. Although the invariance of signature is not part of the syllabus, it is worth understanding. Here is a simple proof.

Let R be an invertible $n \times n$ matrix over $\mathbb{R}$ such that if

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = R \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

then

$$Q(x_1, \ldots, x_n) = Y_1^2 + \cdots + Y_{p'}^2 - Y_{p'+1}^2 - \cdots - Y_{r'}^2$$

for all $(x_1, \ldots, x_n)^{\mathrm{tr}} \in \mathbb{R}^n$. Define $q := r - p$ (as above) and $q' := r' - p'$. Call a subspace U of $\mathbb{R}^n$ *positive* if $Q(u) > 0$ for all $u \in U \setminus \{0\}$, and call a subspace W *non-positive* if $Q(u) \leqslant 0$ for all $u \in W \setminus \{0\}$. Clearly, if U is a positive subspace and W is a non-positive subspace then $U \cap W = \{0\}$, so $\dim U + \dim W \leqslant n$. Define subspaces U_1, U_2, W_1, W_2 by

$$\begin{aligned} U_1 &:= \{u \in \mathbb{R}^n \mid X_{p+1} = \ldots = X_n = 0\}, \\ W_1 &:= \{u \in \mathbb{R}^n \mid X_1 = \ldots = X_p = 0\}, \\ U_2 &:= \{u \in \mathbb{R}^n \mid Y_{p'+1} = \ldots = Y_n = 0\}, \\ W_2 &:= \{u \in \mathbb{R}^n \mid Y_1 = \ldots = Y_{p'} = 0\}. \end{aligned}$$

(Note that here $X_i = 0$ and $Y_j = 0$ are to be construed as linear equations in the coordinates $x_1, \ldots, x_n$ of u.) Then U_1 and U_2 are positive subspaces of dimensions p, p' respectively and W_1, W_2 are non-positive subspaces of dimensions $n-p$, $n-p'$ respectively. It follows that $p+(n-p') \leqslant n$ so that $p \leqslant p'$, and $p'+(n-p) \leqslant n$ so that $p' \leqslant p$. Therefore $p = p'$.

Note that a similar argument with negative and non-negative subspaces will prove that $q = q'$, and therefore $r = p+q = p'+q' = r'$. But of course the fact that $r = r'$ also comes from the fact that this is the rank of the matrix A of the quadratic form.

Note 3. The invariance of rank and signature is known as *Sylvester's Law of Inertia.*

Note 4. If $Q(u) > 0$ whenever $u \in \mathbb{R}^n \setminus \{0\}$ then Q is said to be *positive definite.* This holds if and only if $r = p = n$. Then the associated bilinear form is an inner product on $\mathbb{R}^n$.

Exercise 58 [FHS 1992, A1, Qn 3]. For $u := (x,y,z,t) \in \mathbb{R}^4$ we define $q(u) := (x^2+y^2+z^2-t^2)$. A subspace U of $\mathbb{R}^4$ is said to be 'positive' if $q(u) > 0$ for all non-zero u in U, it is said to be 'negative' if $q(u) < 0$ for all non-zero u in U, and it is said to be 'null' if $q(u) = 0$ for all $u \in U$. Prove or disprove each of the following assertions:

(i) if U is positive then $\dim U \leqslant 3$;

(ii) there is a unique positive subspace of dimension 3;

(iii) if U is negative then $\dim U \leqslant 1$;

(iv) there exist non-zero subspaces U, V, W such that U is positive, V is null, W is negative and $\mathbb{R}^4 = U \oplus V \oplus W$.

In geometry and in mechanics we often need to study two real quadratic forms simultaneously. The following theorem has a number of applications, in particular to the study of small vibrations of a mechanical system about a state of stable equilibrium.

Theorem. *Let $Q(x_1,\ldots,x_n)$, $R(x_1,\ldots,x_n)$ be real quadratic forms. Suppose that Q is positive definite. Then there is an invertible $n\times n$ matrix P over $\mathbb{R}$ and there exist $a_1,\ldots,a_n \in \mathbb{R}$ such that if*

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = P^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

then the equations

$$\begin{aligned} Q(x_1,\ldots,x_n) &= X_1^2 + \cdots + X_n^2 \\ R(x_1,\ldots,x_n) &= a_1 X_1^2 + \cdots + a_n X_n^2 \end{aligned}$$

hold simultaneously for all $(x_1,\ldots,x_n)^{\mathrm{tr}} \in \mathbb{R}^n$.

The coefficients $a_1, a_2, \ldots, a_n$ are the roots of the equation

$$\det(xA - B) = 0$$

where A, B are the symmetric matrices associated with Q, R respectively.

Proof. Let A and B be the real symmetric matrices associated with Q and R respectively, so that

$$Q(x_1, \ldots, x_n) = u^{\rm tr} A\, u, \qquad R(x_1, \ldots, x_n) = u^{\rm tr} B\, u,$$

where $u = (x_1, \ldots, x_n)^{\rm tr} \in \mathbb{R}^n$. The bilinear form associated with Q is $\langle\, ,\, \rangle_Q$, say, where $\langle u, v \rangle_Q = u^{\rm tr} A\, v$. Since Q is positive definite this is an inner product on $\mathbb{R}^n$. Let $u_1, \ldots, u_n$ be an orthonormal basis for $\mathbb{R}^n$ with respect to this new inner product, and let U be the $n \times n$ matrix whose columns are $u_1, \ldots, u_n$. The equations $u_i A\, u_j = 1$ if $i = j$ and $u_i A\, u_j = 0$ if $i \neq j$ may be written $U^{\rm tr} A\, U = I$.

Now let $C := U^{\rm tr} B\, U$. Then, since B is symmetric,

$$C^{\rm tr} = U^{\rm tr} B^{\rm tr} (U^{\rm tr})^{\rm tr} = U^{\rm tr} B\, U = C,$$

that is C is symmetric. By the Diagonalisability Theorem, Version 2, there exists an orthogonal matrix T such that $T^{-1} C\, T = D$, where D is diagonal, say $D = \mathrm{Diag}(a_1, \ldots, a_n)$. Since T is orthogonal $T^{-1} = T^{\rm tr}$ and so $T^{\rm tr} C\, T = D$. Now if $P := U\, T$ then

$$P^{\rm tr} A\, P = T^{\rm tr}\, U^{\rm tr} A\, U\, T = T^{\rm tr} I\, T = I,$$

and

$$P^{\rm tr} B\, P = T^{\rm tr}\, U^{\rm tr} B\, U\, T = T^{\rm tr} C\, T = D.$$

This means that if $v = P^{-1} u$ and $v = (X_1, \ldots, X_n)$ then

$$Q(x_1, \ldots, x_n) = (P\, v)^{\rm tr} A\, (P\, v) = v^{\rm tr}\, (P^{\rm tr} A\, P)\, v = X_1^2 + \cdots + X_n^2$$

and

$$R(x_1, \ldots, x_n) = (P\, v)^{\rm tr} B\, (P\, v) = v^{\rm tr}\, (P^{\rm tr} B\, P)\, v = a_1\, X_1^2 + \cdots + a_n\, X_n^2,$$

as required.

We have seen that $U^{\rm tr} A U = I$ and $U^{\rm tr} B U = D = \mathrm{Diag}(a_1, \ldots, a_n)$. Clearly, $\det(xI - D) = \prod(x - a_i)$. Thus $\det(x\, U^{\rm tr} A U - U^{\rm tr} B U) = \prod(x - a_i)$. Therefore $\det(x\, A - B) = c \prod(x - a_i)$ where $c := (\det U)^{-2}$. So $a_1, \ldots, a_n$ are the roots of the equation $\det(x\, A - B) = 0$ and this competes the proof of the theorem.

Exercise 59 [An old FHS question]. (i) Let A, B be real symmetric $n \times n$ matrices. Suppose that all the eigenvalues of A are positive and let T be an invertible $n \times n$ matrix such that $T^{\rm tr} A T = I_n$. [It is a theorem of the course that such a matrix exists]. Show that $T^{\rm tr} B T$ is symmetric and deduce that there exists an invertible $n \times n$ matrix S such that both $S^{\rm tr} A S = I_n$ and $S^{\rm tr} B S$ is diagonal. Show also that the diagonal entries of $S^{\rm tr} B S$ are the roots of the equation $\det(xA - B) = 0$.

(ii) Show that if

$$\begin{aligned} Q(x, y, z) &= 3x^2 + 5y^2 + 3z^2 + 2xy - 2xz + 2yz, \\ R(x, y, z) &= x^2 + y^2 + z^2 + 10xy + 2xz - 6yz, \end{aligned}$$

then there exist linear forms l, m, n in x, y, z such that

$$Q(x, y, z) = l^2 + m^2 + n^2 \quad \text{and} \quad R(x, y, z) = l^2 + \sqrt{2}\, m^2 - \sqrt{2}\, n^2.$$

Further exercises IV

EXERCISE 61. Express the following quadratic forms as sums or differences of squares of linearly independent linear forms in x, y, z: $x^2 + 2xy + 2y^2 - 2yz - 3z^2$; $xy + yz + xz$. [NOTE: it is probably quicker to use the method of 'completing the square' than to use the method given by the proof of the theorem on diagonalisation of real quadratic forms.]

EXERCISE 62 [FHS 1983, I, Qn 3]. Let α be a self-adjoint linear transformation of a finite dimensional real inner-product space V. Show that V has an orthonormal basis consisting of eigenvectors of α.

Suppose the eigenvalues of α are strictly positive, and let

$$G = \{\pi \mid \pi \text{ is a linear transformation of } V \text{ and } \pi^*\alpha\pi = \alpha\}.$$

Show that G is a group isomorphic to $\mathrm{O}(V)$, the group of orthogonal transformations of V.

EXERCISE 63. Let X_1 and X_2 be subspaces of a real finite-dimensional vector space V, and suppose that $V = X_1 + X_2$. Suppose that X_1 and X_2 have inner products $\langle\,,\,\rangle_1$ and $\langle\,,\,\rangle_2$ which agree on $X_1 \cap X_2$. Show that:

(i) there exists a basis for V which contains an orthonormal basis for X_1 and an orthonormal basis for X_2;

(ii) there exists an inner product on V which agrees with $\langle\,,\,\rangle_1$ on X_1 and $\langle\,,\,\rangle_2$ on X_2.

Is this inner product unique? Justify your answer.

EXERCISE 64 [Cambridge Tripos Part IA, 1995]. Let V be a real inner product space, and let $||v|| := \langle v, v\rangle^{\frac{1}{2}}$. Prove that

$$||x - y|| \leqslant ||z - x|| + ||y - z||$$

for all x, y, $z \in V$. When does equality occur?

Prove that
$$\left|\left|\frac{x-y}{||x-y||^2} - \frac{x}{||x||^2}\right|\right| = \frac{||y||}{||x-y||\;||x||}.$$

Hence show that $||y||\;||z|| \leqslant ||z - x + y||\;||x|| + ||z - x||\;||x - y||$.

EXERCISE 65 Let V be the vector space of infinitely differentiable functions $f : [-\pi, \pi] \to \mathbb{R}$ equipped with the inner product $\langle f, g\rangle := \int_{-\pi}^{\pi} f(t)\,g(t)\,dt$, and let $\Delta : V \to V$ be the linear transformation $\Delta : f \mapsto f''$.

(i) Show that if $V_0 := \{f \in V \mid f(-\pi) = f(\pi) = 0\}$ then V_0 is a subspace of V and $\dim(V/V_0) = 2$.

(ii) Show that if $f, g \in V_0$ then $\langle \Delta f, g\rangle = \langle f, \Delta g\rangle$.

(iii) What is wrong with the assertion that Δ is a self-adjoint transformation of V_0?

Exercise 66. [FHS 1996, A1, Qn 3.] Let V be a finite-dimensional real inner product space.

(a) If $v \in V$ show that there is a unique element θ_v in the dual space V' of V such that
$$\theta_v(u) = \langle u, v\rangle \text{ for all } u \in V.$$

(b) Show that the map $\theta : V \to V'$ given by $\theta(v) = \theta_v$ (for $v \in V$) is an isomorphism. [You may assume that $\dim V = \dim V'$.]

(c) Let W be a subspace of V, and let $W^{\perp}$ be the orthogonal complement of W in V. Show that $\theta(W^{\perp}) = W^{\circ}$, where W° is the annihilator of W in V'.

Now let V be the space of polynomials in x of degree at most 2, with real coefficients. Define an inner product on V by setting
$$\langle f, g\rangle = \int_0^1 f(x)\, g(x)\, \mathrm{d}x$$
for $f, g \in V$. You may assume that $\{1,\ \sqrt{3}(1-2x),\ \sqrt{5}(6x^2-6x+1)\}$ is an orthonormal basis for V. Show that the map $\phi : V \to \mathbb{R}$ given by $\phi(f) = f''(0)$ defines a linear functional on V (*i.e.* show that $\phi \in V'$), and find $v \in V$ such that $\theta_v = \phi$.

Exercise 67 [FHS 1995, A1, Qn 3]. Let V be a finite-dimensional vector space over a field F, and let $T : V \to V$ be a linear transformation. Let $v \in V$, $v \neq 0$, and suppose that V is spanned by $\{v,\ T v,\ T^2 v,\ \ldots,\ T^j v,\ \ldots\}$.

(i) Show that there is an integer $k \geqslant 1$ such that $v,\ T v,\ T^2 v,\ \ldots,\ T^{k-1} v$ are linearly independent but
$$T^k v = \alpha_0\, v + \alpha_1\, T\, v + \cdots + \alpha_{k-1}\, T^{k-1} v$$
for some $\alpha_0,\ \alpha_1,\ \ldots,\ \alpha_{k-1} \in F$.

(ii) Prove that $\{v,\ Tv,\ T^2 v,\ \ldots,\ T^{k-1} v\}$ is a basis for V.

Let T have minimum polynomial $m(x)$ and characteristic polynomial $c(x)$.

(iii) Prove that $m(x) = x^k - \alpha_{k-1}\, x^{k-1} - \cdots - \alpha_1\, x - \alpha_0$.

(iv) By considering the matrix of T with respect to the basis $\{v,\ T v,\ T^2 v,\ \ldots,\ T^{k-1} v\}$, prove (without using the Cayley–Hamilton Theorem) that $m(x) = c(x)$.

Now let A be the matrix $\begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. Show that there does *not* exist a column vector $v \in \mathbb{R}^4$ such that $\mathbb{R}^4$ is spanned by $\{v,\ A v,\ A^2 v,\ \ldots,\ A^j v,\ \ldots\}$.

Exercise 68. Let V be a finite-dimensional vector space, and let $S, T : V \to V$ be linear transformations. Let m_1 and m_2 denote the minimal polynomials of $S T$ and $T S$ respectively. By considering relations such as $T(S T)^r S = (T S)^{r+1}$ show that $m_2(x) = x^i m_1(x)$, where $i = -1,\ 0$, or $+1$. Show that λ is an eigenvalue of $S T$ if and only if λ is an eigenvalue of $T S$.

Exercise 69. Let A denote the real matrix $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ -4 & -11 & -4 \end{pmatrix}$. Calculate the matrix $\sum_{r=1}^{99} A^r$.

ΠMN: Queen's: 13.xi.2007