

Linear Mathematics

Lent Term 2002.

My purpose here is to provide a *terse* set of notes for the course, including all the definitions, theorems, lemmata, etc.. I have focused on the most important aspects of the material, especially in the techniques and outlines of proofs (which require students to fill in the easier details — the real way to learn the material). I trust that these notes will permit me to be more intuitive and pictorial in lectures and give more examples; I hope that they will allow you to spend less time copying and more time understanding why the theorems are true, the patterns of proofs, and the need for the extra hypotheses in the statements of results.

We will generalise the ideas from 3-space that you met in the Algebra & Geometry course to cover arbitrary (but usually finite dimensional) real and complex vector spaces. This requires abstraction but the advantages will be considerable. Although matrices are very concrete, they can be very messy in computations. To overcome this, we will introduce an equivalent notion of linear transformations (the vector space analogue of group homomorphisms) and often use that instead. Just as many maps in 3-space are best understood by considering them with respect to another basis, the same is true in a more general setting. The proofs (other than those which are simply symbol manipulations to equivalent statements) are repeatedly by induction (on the dimension) and take the following form: Find a subspace on which the linear transformation acts “nicely”; find a complementary subspace on which the linear transformation acts, and apply the inductive hypothesis to this complementary subspace of smaller dimension; put the two pieces together to get the result for the entire space. One major tool which will often help us achieve this is the “minimal polynomial”. If you remember this outline throughout the course, you won’t lose sight of the wood for the trees as the number of results grow.

Any errata in these notes are entirely due to my incompetence as a proof reader. Please alert me to any that you find. Thank you.

A. M. W. Glass (amwg@dpmms.cam.ac.uk)

0.1 Contents

Chapter 1 Vector spaces and linear maps	3
1.1. Definitions, etc.....	3
1.2. Quotient spaces.....	7
1.3. Spanning and linear dependence.....	8
1.4. Direct sums.....	11
1.5. Change of basis.....	13
1.6. Column reduction.....	16
1.7. Exercises.....	18
Chapter 2 Endomorphisms	25
2.1. Eigenvalues and eigenvectors.....	25
2.2. The minimal polynomial.....	27
2.3. Summary.....	36
2.4. Exercises.....	37
Chapter 3 Jordan normal form	41
3.1. Nilpotent matrices.....	41
3.2. Jordan canonical form.....	45
3.3. Differential equations.....	46
3.4. Exercises.....	47
Chapter 4 Determinants	51
4.1. The desired properties.....	51
4.2. The characteristic polynomial.....	52
4.3. Volumes.....	54
4.4. Properties of the determinant.....	57
4.5. Cofactors and the adjugate matrix.....	59
4.6. Exercises.....	60
Chapter 5 The Dual space	63
5.1. The dual and double dual.....	63
5.2. Exercises.....	67

Chapter 1

Vector Spaces and Linear Maps

1.1 Definitions, etc.

Definition 1.1.1 Let G be a set with a binary operation $+$ defined on it. Then G is called an **Abelian group** if $+$ is associative ($a + (b + c) = (a + b) + c$ for all $a, b, c \in G$) and commutative ($a + b = b + a$ for all $a, b \in G$), there is a zero in G (denoted 0) such that $a + 0 = a$ for all $a \in G$) and for each $a \in G$, there is $a^* \in G$ such that $a + a^* = 0$. We will write $-a$ for a^* . If the operation on G is denoted by \cdot , we will write 1 for the “zero” and a^{-1} or $1/a$ for a^* .

Definition 1.1.2 Let F be a set with two operations defined on it denoted by $+$ and \cdot . Suppose that $(F, +)$ is an Abelian group with 0 as zero. Then F is a **field** if $(F \setminus \{0\}, \cdot)$ is an Abelian group and $a(b + c) = ab + ac$ for all $a, b, c \in F$.

Example 1.1.3 \mathbb{Z} is not a field since only ± 1 have multiplicative inverses.

Example 1.1.4 $\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Q}[\sqrt{2}]$ are all fields. $\mathbb{Z}/p\mathbb{Z}$ is a field if and only if p is prime: if p is not prime, then no proper divisor of p has an inverse.

[**Aside:**

Definition 1.1.5 A field F is said to be algebraically closed if every polynomial with coefficients in F has a solution in F . Equivalently, every polynomial with coefficients in F can be written as a product of linear polynomials (polynomials of degree at most 1, the coefficients being in F).

Example 1.1.6 \mathbb{C} is the smallest algebraically closed field containing \mathbb{R} . Gauss gave four proofs of this, the first of which was his Ph.D. in 1800.

FACT: Every field F is contained in an algebraically closed field K ; the intersection of all algebraically closed subfields of K containing F is an algebraically closed field called the algebraic closure of F .]

Definition 1.1.7 Let V be a set and F be a field. A map from $F \times V$ into V is called a scalar multiplication (with scalars in F).

Definition 1.1.8 A vector space V over a field F is an Abelian group under addition that has a scalar multiplication with scalars in F satisfying (for all $\lambda, \mu \in F$ and $\mathbf{u}, \mathbf{v} \in V$)

- (i) $\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$,
- (ii) $(\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$,
- (iii) $(\lambda \cdot \mu)\mathbf{v} = \lambda(\mu\mathbf{v})$ and
- (iv) $1\mathbf{v} = \mathbf{v}$.

Definition 1.1.9 The subsets of a vector space which are closed under addition and scalar multiplication are precisely the subsets that are vector spaces (under the induced operations). These are called vector subspaces. We will often write subspace for vector subspace.

Example 1.1.10 (i) \mathbb{R}^3 is a vector space (over \mathbb{R}) as is \mathbb{R}^n for any positive integer n .

(ii) The set of all $m \times n$ matrices with entries from a field F is a vector space $\mathcal{M}_{m \times n}(F)$ over F .

(iii) The set of all diagonal $n \times n$ matrices over \mathbb{C} is a subspace of the vector space of all $n \times n$ matrices with complex entries; so are the set of all $n \times n$ with trace 0 ($\sum_{i=1}^n A_{i,i} = 0$), the set of all $n \times n$ symmetric matrices with complex entries ($A_{i,j} = A_{j,i}$), the set of all antisymmetric

complex $n \times n$ matrices ($A_{i,j} = -A_{j,i}$) and the set of all Hermitian $n \times n$ matrices ($A_{i,j} = \overline{A_{j,i}}$).

(iv) The set $C(\mathbb{R}, \mathbb{R})$ of all continuous functions from \mathbb{R} into \mathbb{R} is a real vector space under functional addition and the usual scalar multiplication. The set of all real polynomials is a subspace thereof, as is the set of all twice differentiable real functions.

(v) The set of all sequences of real numbers is a vector space over \mathbb{R} under sequence addition and the usual scalar multiplication.

Definition 1.1.11 Cf., groups, we seek maps that preserve the vector space operations. A map τ from a vector space V to a vector space W over the same field is called a linear map if $\tau(\mathbf{u} + \mathbf{v}) = \tau(\mathbf{u}) + \tau(\mathbf{v})$ and $\tau(\lambda\mathbf{v}) = \lambda\tau(\mathbf{v})$ for all scalars λ and $\mathbf{u}, \mathbf{v} \in V$. Note that the operations on the right hand sides are in W . So linear maps are just vector space homomorphisms.

Example 1.1.12 Let $\mathbb{R}^{\mathbb{N}}$ be the real vector space of all real sequences under addition, and V be the subspace of all convergent real sequences. Define $\tau : V \rightarrow \mathbb{R}$ by $\tau((x_n)) = \lim_{n \rightarrow \infty} x_n$. Then τ is a linear map. (Verify this.)

Example 1.1.13 Let τ and σ be linear maps from vector space V into vector space W . If we define $(\lambda\tau)(\mathbf{v}) = \lambda(\tau(\mathbf{v}))$ and $(\sigma + \tau)(\mathbf{v}) = \sigma(\mathbf{v}) + \tau(\mathbf{v})$, then the set of all linear maps from V into W is itself a vector space which we denote by $\mathcal{L}(V, W)$.

If $\sigma \in \mathcal{L}(U, V)$, then the map $\Theta : \mathcal{L}(V, W) \rightarrow \mathcal{L}(U, W)$ given by $\Theta : \tau \mapsto \tau\sigma$ is linear (why?) where we write $\tau\sigma$ for $\tau \circ \sigma$.

The idea of a linear map is tied up very closely with matrices.

Proposition 1.1.14 Let F be a field and m, n be positive integers. There is a linear bijection between $\mathcal{L}(F^m, F^n)$ and $\mathcal{M}_{n \times m}(F)$.

Proof: Let $\mathbf{e}_1, \dots, \mathbf{e}_m$ be the standard coordinate basis for F^m and $\mathbf{f}_1, \dots, \mathbf{f}_n$ be the standard coordinate basis for F^n . Given $\tau \in \mathcal{L}(F^m, F^n)$ we have $\tau(\mathbf{e}_j) = \sum_{i=1}^n t_{i,j} \mathbf{f}_i$ for some $t_{i,j} \in F$ ($j = 1, \dots, m$). Let $\Phi(\tau)$ be the $n \times m$ matrix whose (i, j) entry is $t_{i,j}$. An easy computation shows that Φ is linear. (Fill in this gap.) Let S be any $n \times m$ matrix and let $\sigma(\mathbf{v}) = S\mathbf{v}$ ($\mathbf{v} \in F^m$); then σ is easily seen to be linear (do this) and so

belongs to $\mathcal{L}(F^m, F^n)$. Let $\Psi(S) = \sigma$. Then $\Psi\Phi(\tau) = \tau$ and $\Phi\Psi(S) = S$. Hence Φ is a bijection. //

Definition 1.1.15 A bijective map τ (for which both τ and τ^{-1} are linear) is called an isomorphism. If $\tau : U \rightarrow V$ is an isomorphism, then we write $U \cong V$.

If $U \cong V$, then we can view U and V as the same (as vector spaces); only the names are changed.

Proposition 1.1.16 Any bijective map in $\mathcal{L}(U, V)$ is an isomorphism.

Proof: We need only show that τ^{-1} is linear. But $\tau(\tau^{-1}(\mathbf{v}_1 + \mathbf{v}_2)) = \mathbf{v}_1 + \mathbf{v}_2 = \tau(\tau^{-1}(\mathbf{v}_1)) + \tau(\tau^{-1}(\mathbf{v}_2)) = \tau(\tau^{-1}(\mathbf{v}_1) + \tau^{-1}(\mathbf{v}_2))$. Since τ is injective, we get $\tau^{-1}(\mathbf{v}_1 + \mathbf{v}_2) = \tau^{-1}(\mathbf{v}_1) + \tau^{-1}(\mathbf{v}_2)$. Similarly, $\tau^{-1}(\lambda\mathbf{v}) = \lambda\tau^{-1}(\mathbf{v})$. //

Proposition 1.1.17 If $\tau \in \mathcal{L}(U, V)$, then $\text{Im}(\tau) = \{\tau(\mathbf{u}) : \mathbf{u} \in U\}$ is a subspace of V and $\ker(\tau) = \{\mathbf{u} \in U : \tau(\mathbf{u}) = \mathbf{0}\}$ is a subspace of U . Moreover, $\ker(\tau) = \{\mathbf{0}\}$ if and only if τ is injective.

We explore the connection between linear maps and matrices further.

Let $\sigma \in \mathcal{L}(F^k, F^m)$ and $\tau \in \mathcal{L}(F^m, F^n)$ correspond to an $m \times k$ matrix S and a $n \times m$ matrix T , respectively. If $\mathbf{g}_1, \dots, \mathbf{g}_k$ is the standard basis of F^k , then $\tau(\sigma(\mathbf{g}_\ell)) = \tau(\sum_{j=1}^m s_{j,\ell} \mathbf{e}_j) = \sum_{j=1}^m s_{j,\ell} \tau(\mathbf{e}_j) = \sum_{j=1}^m \sum_{i=1}^n s_{j,\ell} t_{i,j} \mathbf{f}_i = (TS)(\mathbf{g}_\ell)$ for $\ell = 1, \dots, k$.

MORAL: Composition of linear maps corresponds to the product of their corresponding matrices. So we can pass between linear maps on various F^n 's and matrices without worrying.

Since composition of functions is associative, we get that matrix multiplication is also associative; i.e., $A(BC) = (AB)C$ for matrices (This is the real reason why matrix multiplication is associative, without any subscript nonsense.)

1.2 Quotient Spaces

Let $\tau \in \mathcal{L}(V, W)$. For any $\mathbf{w} \in W$, let $\tau^{-1}(\mathbf{w}) = \{\mathbf{v} \in V : \tau(\mathbf{v}) = \mathbf{w}\}$. [Caution: τ^{-1} is not a function; this is merely notation for the set of all vectors in V that are mapped by τ to \mathbf{w} .] Then $\tau^{-1}(\mathbf{w})$ is a subspace of W if and only if $\mathbf{w} = \mathbf{0}$. If $K = \ker(\tau)$, then

$$\tau(\mathbf{v}_1) = \tau(\mathbf{v}_2) \iff \mathbf{v}_1 - \mathbf{v}_2 \in K \iff K + \mathbf{v}_1 = K + \mathbf{v}_2.$$

So $\tau^{-1}(\mathbf{w})$ is just the translation of K by any vector \mathbf{v} for which $\tau(\mathbf{v}) = \mathbf{w}$; this is akin to a coset in groups.

To clarify this abstract idea, consider two concrete examples.

Example 1.2.1 Let $\tau : \mathbb{R}^2 \rightarrow \mathbb{R}$ be the projection onto the x -axis; i.e., $\tau(x, y) = x$. Then the kernel of τ is just the y -axis ($\{(0, y) : y \in \mathbb{R}\}$) and $\tau^{-1}(5) = \{(5, y) : y \in \mathbb{R}\} = \ker(\tau) + (5, 0) = \ker(\tau) + (5, \pi) = \dots$, the vertical line through $(5, 0)$. Indeed, $\tau^{-1}(x)$ is just the vertical line through $(x, 0)$, and these vertical lines can be added in the natural way to get another vertical line: the sum of the vertical lines through $(x_1, 0)$ and $(x_2, 0)$ is the vertical line through $(x_1 + x_2, 0)$. Similarly, multiplying the vertical line through $(x, 0)$ by λ is just the vertical line through $(\lambda x, 0)$. Thus the set of vertical lines becomes a vector space in its own right and is called the quotient space via τ .

Example 1.2.2 Let V be the vector space of all real-valued twice differentiable real functions, and K be the subspace of all solutions of the homogeneous differential equation $f'' + t^2 f' + f = 0$. If $f_0'' + t^2 f_0' + f_0 = g$, then $f_0 + K$ is the set of all solutions to $f'' + t^2 f' + f = g$; i.e., the set of all solutions of the inhomogeneous differential equation is just the set obtained by adding a particular solution, f_0 , of the inhomogeneous differential equation to an arbitrary solution of the homogeneous differential equation. Similarly, the solution of an inhomogeneous system of linear equations is just a particular solution added to an arbitrary solution of the associated homogeneous system.

More generally, let K be a subspace of a vector space V over F . Define

$$\mathbf{v}_1 \sim \mathbf{v}_2 \quad \text{iff} \quad \mathbf{v}_1 - \mathbf{v}_2 \in K.$$

This defines an equivalence relation on V . Moreover, if $\mathbf{v}_1 \sim \mathbf{v}_2$ and $\mathbf{w}_1 \sim \mathbf{w}_2$, then it is easily checked that $\mathbf{v}_1 + \mathbf{w}_1 \sim \mathbf{v}_2 + \mathbf{w}_2$ and $\lambda \mathbf{v}_1 \sim \lambda \mathbf{v}_2$. So the set of equivalence classes forms a vector space over F . This is called the quotient space V/K of V over K . The map $\rho_K : V \rightarrow V/K$ given by $\rho_K(\mathbf{v}) = [\mathbf{v}]$ is linear and is called the quotient map. It is easily seen that $\ker(\rho_K) = K$. (Establish all the claims in this paragraph.)

Theorem 1.A The First Isomorphism Theorem *Let $\tau \in \mathcal{L}(V, W)$ and $K = \ker(\tau)$. Then the map $\phi : V/K \rightarrow \text{Im}(\tau)$ given by $\phi([\mathbf{v}]) = \tau(\mathbf{v})$ is a well-defined isomorphism between V/K and $\text{Im}(\tau)$. Moreover, $\phi \circ \rho_K = \tau$.*

Proof:

$$\mathbf{u} \sim \mathbf{v} \iff \mathbf{u} - \mathbf{v} \in K = \ker(\tau) \iff \tau(\mathbf{u} - \mathbf{v}) = \mathbf{0} \iff \tau(\mathbf{u}) = \tau(\mathbf{v}).$$

By \implies , ϕ is well-defined; and by \impliedby it is injective. If $\mathbf{w} \in \text{Im}(\tau)$, say $\mathbf{w} = \tau(\mathbf{v})$, then $\phi(\rho_K(\mathbf{v})) = \phi([\mathbf{v}]) = \tau(\mathbf{v}) = \mathbf{w}$; hence ϕ is surjective. An easy check shows that ϕ is linear, whence an isomorphism with the desired property. //

1.3 Spanning and Linear Dependence

Notation: Let V be a vector space over F and S be a subset of V . Let $\langle S \rangle = \{ \sum_{i=1}^n \lambda_i \mathbf{s}_i : n \in \mathbb{N} \cup \{0\}, \mathbf{s}_1, \dots, \mathbf{s}_n \in S, \lambda_1, \dots, \lambda_n \in F \}$. Then $\langle S \rangle$ is a subspace of V containing S . If W is any subspace of V containing S , then W contains $\langle S \rangle$ (why?); i.e., S generates or spans $\langle S \rangle$.

Definition 1.3.1 *Let S be a subset of a vector space V . Then S spans V if $\langle S \rangle = V$.*

Proposition 1.3.2 *Let S be a subset of a vector space V and $\mathbf{s}_0 \in S$. If $\mathbf{s}_0 \in \langle S \setminus \{\mathbf{s}_0\} \rangle$, then $\langle S \rangle = \langle S \setminus \{\mathbf{s}_0\} \rangle$.*

Proof: Let $\mathbf{s}_0 = \sum_{i=1}^n \mu_i \mathbf{s}_i$ with $\mathbf{s}_1, \dots, \mathbf{s}_n \in S \setminus \{\mathbf{s}_0\}$. Let $\mathbf{u} \in \langle S \rangle$, say $\mathbf{u} = \lambda_0 \mathbf{s}_0 + \sum_{i=1}^n \lambda_i \mathbf{s}_i$. Then $\mathbf{u} = \sum_{i=1}^n (\lambda_0 \mu_i + \lambda_i) \mathbf{s}_i \in \langle S \setminus \{\mathbf{s}_0\} \rangle$. //

Definition 1.3.3 *A subset S is said to be linearly independent if $s \notin \langle S \setminus \{s\} \rangle$ for all $s \in S$; i.e., no proper subset of S spans $\langle S \rangle$.*

We say that S is linearly dependent if S is not linearly independent.

Proposition 1.3.4 S is linearly independent iff for all distinct $\mathbf{s}_1, \dots, \mathbf{s}_n \in S$, $\sum_{j=1}^n \lambda_j \mathbf{s}_j = \mathbf{0} \implies \lambda_1 = \dots = \lambda_n = 0$.

Proof: $\mathbf{s}_0 \in \langle S \setminus \{\mathbf{s}_0\} \rangle$ iff there are distinct $\mathbf{s}_1, \dots, \mathbf{s}_n \in S \setminus \{\mathbf{s}_0\}$ with $\mathbf{s}_0 = \sum_{i=1}^n \lambda_i \mathbf{s}_i$ iff $\sum_{i=0}^n \lambda_i \mathbf{s}_i = \mathbf{0}$ where $\lambda_0 = -1 \neq 0$. //

Definition 1.3.5 A linearly independent spanning set of a vector space is called a basis.

Note: Bases are not assumed to be finite unless this is explicitly stated.

Corollary 1.3.6 If S is finite and spans a vector space V , then some subset of S is a basis of V .

Proposition 1.3.7 \mathcal{B} is a basis for vector space V iff every element of V can be written uniquely as a linear combination of elements of \mathcal{B} .

Proof: Suppose that \mathcal{B} is a basis of V . If \mathbf{v} can be written in two distinct ways, then subtracting gives $\mathbf{0} = \sum_{j=1}^m \lambda_j \mathbf{b}_j$ with not all λ_j equal to 0. This contradicts the linear independence of \mathcal{B} .

Conversely, spanning is obvious, and as $\mathbf{0} = \sum_{j=1}^n 0\mathbf{b}_j$ for all $\mathbf{b}_1, \dots, \mathbf{b}_n \in \mathcal{B}$, the uniqueness implies that \mathcal{B} is linearly independent. //

Linear maps are completely determined by specifying their actions on a basis:

Proposition 1.3.8 Let \mathcal{B} be a basis for a vector space V .

(i) If $\tau_1, \tau_2 \in \mathcal{L}(V, W)$ with $\tau_1(\mathbf{b}) = \tau_2(\mathbf{b})$ for all $\mathbf{b} \in \mathcal{B}$, then $\tau_1 = \tau_2$.

(ii) If $\tau : \mathcal{B} \rightarrow W$ is an arbitrary map, then there is a unique extension of τ to a linear map from V into W .

Proof: If $\mathbf{v} \in V$, then $\mathbf{v} = \sum_{j=1}^m \lambda_j \mathbf{b}_j$ uniquely. Any $\sigma \in \mathcal{L}(V, W)$ satisfies $\sigma(\mathbf{v}) = \sum_{j=1}^m \lambda_j \sigma(\mathbf{b}_j)$. This gives (i) and (ii). //

Corollary 1.3.9 If V is a vector space with a finite basis \mathcal{B} , then $V \cong F^n$ where n is the number of elements of \mathcal{B} .

Proof: If $\mathbf{e}_1, \dots, \mathbf{e}_n$ is the standard basis for F^n and $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, then $\tau(\mathbf{b}_j) = \mathbf{e}_j$ ($j = 1, \dots, n$) extends uniquely to a linear map τ^* from V into F^n by (ii). Each element of F^n has form $\sum_{j=1}^n \lambda_j \mathbf{e}_j = \tau^*(\sum_{j=1}^n \lambda_j \mathbf{b}_j)$, whence τ^* is surjective. The same calculation shows that τ^* is injective. //

This strongly suggests that any two bases have the same size. We will obtain this result from the following important Lemma.

Lemma 1.3.10 The Steinitz Exchange Lemma *If $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ is a basis of a vector space V and $\mathbf{v} = \sum_{j=1}^n \lambda_j \mathbf{b}_j$ with $\lambda_1 \neq 0$, then $\{\mathbf{v}, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ is a basis of V .*

Proof: $\{\mathbf{v}, \mathbf{b}_2, \dots, \mathbf{b}_n\} \supseteq \mathcal{B}$ and so spans V . Since $\mathbf{b}_1 \in \langle \mathbf{v}, \mathbf{b}_2, \dots, \mathbf{b}_n \rangle$, we get $\{\mathbf{v}, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ spans V by Proposition 1.3.2. If $\mu \mathbf{v} + \sum_{j=2}^n \mu_j \mathbf{b}_j = \mathbf{0}$, then $\mu \lambda_1 \mathbf{b}_1 + \sum_{j=2}^n (\mu_j + \lambda_j \mu) \mathbf{b}_j = \mathbf{0}$. Since \mathcal{B} is a basis and $\lambda_1 \neq 0$ we deduce that $\mu = 0$ and $\mu_j = 0$ ($j = 2, \dots, n$). Thus $\{\mathbf{v}, \mathbf{b}_2, \dots, \mathbf{b}_n\}$ is linearly independent (and we already know that it spans); hence it is a basis of V . //

Corollary 1.3.11 *If $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ is a basis for a vector space V and $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ is a linearly independent set, then $n \geq m$ and there is a basis $\mathcal{C} \supseteq \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ with $|\mathcal{B} \cap (\mathcal{C} \setminus \{\mathbf{u}_1, \dots, \mathbf{u}_m\})| = n - m$.*

Proof: By the Steinitz Exchange Lemma and relabelling, we have that $\{\mathbf{u}_1, \mathbf{b}_1, \dots, \mathbf{b}_{n-1}\}$ is a basis of V . Since $\mathbf{u}_2 \notin \langle \mathbf{u}_1 \rangle$, we relabel and obtain $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{b}_1, \dots, \mathbf{b}_{n-2}\}$ is a basis of V likewise. Since $\mathbf{u}_3 \notin \langle \mathbf{u}_1, \mathbf{u}_2 \rangle$, we relabel and get $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{b}_1, \dots, \mathbf{b}_{n-3}\}$ is also a basis of V ; etc.. //

Definition 1.3.12 *A vector space V is said to be finite dimensional if it has a finite spanning set.*

By Corollary 1.3.11, we immediately obtain

Corollary 1.3.13 *If V is a finite dimensional vector space and $L \subseteq V$ is linearly independent, then V has a basis containing L .*

and

Corollary 1.3.14 *If V is a finite dimensional vector space and \mathcal{B} is any basis of V , then every basis of V has $|\mathcal{B}|$ elements.*

This is what we sought.

Definition 1.3.15 *If V is a finite dimensional vector space, then its dimension is just the number of elements in any basis.*

By Corollary 1.3.9 F^0, F^1, F^2, \dots exhaust the list of all finite dimensional vector spaces over F (to within isomorphism). So the spaces of column vectors that you learnt about in Algebra & Geometry are all the finite dimensional vector spaces!!

Corollary 1.3.16 *Let V be a vector space over F with $\dim(V) = n$.*

(a) *If $S \subseteq V$ is linearly independent and $|S| = n$, then S is a basis for V .*

(b) *If $S \subseteq V$ spans V and $|S| = n$, then S is a basis for V .*

Corollary 1.3.17 *If U is a subspace of a finite dimensional vector space V , then U is finite dimensional and $\dim(U) \leq \dim(V)$. Moreover, under these hypotheses, $U = V$ iff $\dim(U) = \dim(V)$.*

Note that Corollary 1.3.17 fails if V is not finite dimensional: Consider the vector space of all real sequences that are eventually 0. This has basis $\{\mathbf{e}_j : j = 1, 2, \dots\}$ where $\mathbf{e}_1 = (1, 0, 0, \dots)$, $\mathbf{e}_2 = (0, 1, 0, 0, \dots)$, etc.. Then any basis has countably infinite size (prove this) and $U = \langle \mathbf{e}_2, \mathbf{e}_4, \dots, \mathbf{e}_{2n}, \dots \rangle$ is a proper subspace of V , but any basis of U is also countably infinite.

1.4 Direct Sums

Let U_1 and U_2 be subspaces of a vector space V . Let

$$U_1 + U_2 = \{\mathbf{u}_1 + \mathbf{u}_2 : \mathbf{u}_j \in U_j, \quad j = 1, 2\}.$$

Then $U_1 + U_2$ is a subspace of V . Indeed $U_1 + U_2 = \langle U_1 \cup U_2 \rangle$.

Proposition 1.4.1 *If U_1 and U_2 are finite dimensional, then*

$$\dim(U_1 + U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2).$$

Proof: Since $U_1 \cap U_2 \subseteq U_1$, it is finite dimensional by Corollary 1.3.17. Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for $U_1 \cap U_2$. By Corollary 1.3.13, we can extend this to bases \mathcal{B}_j of U_j ($j = 1, 2$). An easy exercise (that frequently appears on the Tripos — so do it!) shows that $\mathcal{B}_1 \cup \mathcal{B}_2$ spans $U_1 + U_2$ and is linearly independent (where $\mathbf{b}_1, \dots, \mathbf{b}_n$ are included only once). //

Definition 1.4.2 If $U_1 \cap U_2 = \{\mathbf{0}\}$, then we write $U_1 \oplus U_2$ for $U_1 + U_2$. A vector space V is said to be the direct sum of subspaces U_1 and U_2 if $V = U_1 \oplus U_2$.

Proposition 1.4.3 V is a direct sum of subspaces U_1 and U_2 iff every $\mathbf{v} \in V$ can be written uniquely in the form $\mathbf{u}_1 + \mathbf{u}_2$ where $\mathbf{u}_j \in U_j$ ($j = 1, 2$).

Proof: If $V = U_1 \oplus U_2$ and $\mathbf{u}_1 + \mathbf{u}_2 = \mathbf{w}_1 + \mathbf{w}_2$, then $\mathbf{u}_1 - \mathbf{w}_1 = \mathbf{u}_2 - \mathbf{w}_2 \in U_1 \cap U_2 = \{\mathbf{0}\}$. Hence $\mathbf{u}_j = \mathbf{w}_j$ ($j = 1, 2$).

Conversely, the condition clearly implies that $V = U_1 + U_2$ and that $U_1 \cap U_2 = \{\mathbf{0}\}$. //

Definition 1.4.4 If U and W are subspaces of V and $V = U \oplus W$, then we call W a complementary subspace of U .

Example 1.4.5 Let $V = \mathbb{R}^2$ and $U = \{(x, 0) : x \in \mathbb{R}\}$. Then U is a subspace of V and $W_1 = \{(0, y) : y \in \mathbb{R}\}$ and $W_2 = \{(z, z) : z \in \mathbb{R}\}$ are both complementary subspaces of U in V . Hence complementary subspaces are not necessarily unique.

Proposition 1.4.6 Each subspace of a finite dimensional subspace has a complementary subspace.

Proof: Let \mathcal{B} be a basis of U . By Corollary 1.3.13, \mathcal{B} is contained in a basis \mathcal{C} of V . Then $\langle \mathcal{C} \setminus \mathcal{B} \rangle$ is a complementary subspace of U . //

Proposition 1.4.6 is also true for arbitrary vector spaces; one uses Zorn's Lemma in the infinite dimensional case.

Returning to quotient spaces: Let K be a subspace of V and W be a complementary subspace of K . Define $\phi : W \rightarrow V/K$ by $\phi(\mathbf{w}) = K + \mathbf{w} = \rho_K(\mathbf{w})$; i.e., $\phi = \rho_K|_W$, so ϕ is linear. Then $\mathbf{w} \in \ker(\phi)$ iff $\mathbf{w} = \mathbf{0}$, whence ϕ is injective. Moreover $K + \mathbf{v} = K + (\mathbf{k} + \mathbf{w}) = K + \mathbf{w} = \phi(\mathbf{w})$. Thus ϕ is surjective. Consequently, via ϕ , $W \cong V/K$:

Corollary 1.4.7 $W \cong (U \oplus W)/U$ for all subspaces U, W of V with $U \cap W = \{\mathbf{0}\}$.

Definition 1.4.8 V is a direct sum of subspaces U_1, \dots, U_m if every element of V can be written uniquely in the form $\mathbf{u}_1 + \dots + \mathbf{u}_m$ where each $\mathbf{u}_j \in U_j$.

We write $V = \bigoplus_{j=1}^m U_j$ or $U_1 \oplus \dots \oplus U_m$ in this case.

Now do Exercise 10.

1.5 Change of Bases

To recap: Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ and $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ be bases for a vector spaces V and W over the same field F . Let $\tau \in \mathcal{L}(V, W)$. Since $\tau(\mathbf{b}_j) \in W$ ($j = 1, \dots, m$) there are elements $t_{i,j} \in F$ such that $\tau(\mathbf{b}_j) = \sum_{i=1}^n t_{i,j} \mathbf{c}_i$. Thus we obtained an $n \times m$ matrix T associated with τ and we saw that composition of linear maps had the same action as matrix multiplication when the vector spaces V and W were renamed F^m and F^n respectively. With our previous notation from Proposition 1.1.14 if $\alpha : F^m \cong V$ with $\alpha(\mathbf{e}_j) = \mathbf{b}_j$ ($j = 1, \dots, m$) and $\beta : F^n \cong W$ with $\beta(\mathbf{f}_i) = \mathbf{c}_i$, then $\rho = \beta^{-1}\tau\alpha : F^m \rightarrow F^n$ does indeed correspond to the matrix T as is easily checked. In this sense, we will regard every linear map between finite dimensional vector spaces as corresponding to matrix multiplication and observe, as before, that composition of linear maps corresponds to matrix multiplication.

Now let $\mathcal{B}' = \{\mathbf{b}'_1, \dots, \mathbf{b}'_m\}$ and $\mathcal{C}' = \{\mathbf{c}'_1, \dots, \mathbf{c}'_n\}$ be new bases of V and W respectively. Let P and Q represent the bases changes; i.e.,

$$\mathbf{b}'_j = \sum_{\ell=1}^m p_{\ell,j} \mathbf{b}_\ell \quad \text{and} \quad \mathbf{c}'_i = \sum_{k=1}^n q_{k,i} \mathbf{c}_k.$$

Then $\tau(\mathbf{b}'_j) = \sum_{\ell=1}^m p_{\ell,j} \tau(\mathbf{b}_\ell) = \sum_{\ell,k} p_{\ell,j} t_{k,\ell} \mathbf{c}_k = \sum_{\ell,k,r} p_{\ell,j} t_{k,\ell} \hat{q}_{r,k} \mathbf{c}'_r$, where $\hat{q}_{r,k}$ is the (r, k) entry of the matrix Q^{-1} . Hence, for $j = 1, \dots, m$,

$$\tau(\mathbf{b}'_j) = \sum_{r=1}^n (Q^{-1}TP)_{r,j} \mathbf{c}'_r.$$

Consequently,

Proposition 1.5.1 *Let V and W be finite dimensional vector spaces over a field F . Let $\mathcal{B}, \mathcal{B}'$ be bases of V with $P(\mathcal{B}) = \mathcal{B}'$, and $\mathcal{C}, \mathcal{C}'$ be bases of W with $Q(\mathcal{C}) = \mathcal{C}'$. Let $\tau \in \mathcal{L}(V, W)$ be represented by matrix T with respect to \mathcal{B} and \mathcal{C} . Then τ is represented by $Q^{-1}TP$ with respect to \mathcal{B}' and \mathcal{C}' .*

In the special case that $V = W$, $\mathcal{C} = \mathcal{B}$ and $\mathcal{C}' = \mathcal{B}'$ we get that τ is represented by $P^{-1}TP$ with respect to the primed basis (where P is the change of basis and τ is represented by T with respect to the original basis).

We now want to choose the bases to represent τ as simply as possible:

Proposition 1.5.2 *Let $\tau \in \mathcal{L}(V, W)$ be linear, $\dim(V) = m$ & $\dim(W) = n$. Let $k = \dim(\ker(\tau))$ and $r = m - k$. Then there are bases $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ for V and $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ for W such that*

$$\tau(\mathbf{b}_i) = \begin{cases} \mathbf{c}_i & \text{if } i \leq r \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

So, with respect to these bases, τ is represented by the $n \times m$ matrix

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

Proof: Let $\{\mathbf{b}_{r+1}, \dots, \mathbf{b}_m\}$ be a basis for $\ker(\tau)$ and extend it to a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ for V . Let $\mathbf{c}_i = \tau(\mathbf{b}_i)$ for $i = 1, \dots, r$. It is enough to show that $\{\mathbf{c}_1, \dots, \mathbf{c}_r\}$ is linearly independent and spans $\text{Im}(\tau)$ (then we can extend it to a basis \mathcal{C} of W ; and τ has the desired form with respect to \mathcal{B} and \mathcal{C})

But $\sum_{j=1}^r \lambda_j \mathbf{c}_j = \mathbf{0}$ iff $\sum_{j=1}^r \lambda_j \mathbf{b}_j \in \ker(\tau)$ iff $\sum_{j=1}^r \lambda_j \mathbf{b}_j = \sum_{j=r+1}^m \lambda_j \mathbf{b}_j$. Since \mathcal{B} is a basis, all λ_j are 0. Hence $\{\mathbf{c}_1, \dots, \mathbf{c}_r\}$ is linearly independent. It is immediate that $\{\tau(\mathbf{b}_1), \dots, \tau(\mathbf{b}_r)\}$ spans $\text{Im}(\tau)$: If $\mathbf{w} \in \text{Im}(\tau)$, then $\mathbf{w} = \tau(\sum_{j=1}^m \lambda_j \mathbf{b}_j)$ for some $\lambda_1, \dots, \lambda_m \in F$. So $\mathbf{w} = \sum_{j=1}^m \lambda_j \tau(\mathbf{b}_j) = \sum_{j=1}^r \lambda_j \tau(\mathbf{b}_j)$ (since $\tau(\mathbf{b}_j) = \mathbf{0}$ if $j = r + 1, \dots, m$). The proposition follows. //

Definition 1.5.3 *If $\tau \in \mathcal{L}(V, W)$, then the rank of τ , $\text{rk}(\tau) = \dim(\text{Im}(\tau))$, and the nullity of τ , $n(\tau) = \dim(\ker(\tau))$.*

The above proof showed that $\{\tau(\mathbf{b}_1), \dots, \tau(\mathbf{b}_r)\}$ was linearly independent. Hence:

Corollary 1.5.4 *Let $\tau \in \mathcal{L}(V, W)$ and V be finite dimensional. Then*

$$rk(\tau) + n(\tau) = \dim(V).$$

Corollary 1.5.5 *If V is finite dimensional and $\tau \in \mathcal{L}(V, V)$, then τ is injective iff it is surjective.*

Corollary 1.5.6 *Let $\sigma \in \mathcal{L}(U, V)$ and $\tau \in \mathcal{L}(V, W)$ where U and V are finite dimensional. Then*

$$rk(\sigma) + rk(\tau) - \dim(V) \leq rk(\tau\sigma) \leq rk(\tau), rk(\sigma).$$

Proof: Clearly $Im(\tau\sigma) \subseteq Im(\tau)$. If $\rho = \tau|Im(\sigma)$, then by Corollary 1.5.4 we have $rk(\tau\sigma) = rk(\rho) = \dim(Im(\sigma)) - n(\rho)$; i.e., $rk(\tau\sigma) = rk(\sigma) - n(\rho)$. Since $ker(\tau) \supseteq ker(\rho)$ we use Corollary 1.5.4 again to deduce that $rk(\tau\sigma) \geq rk(\sigma) - n(\tau) = rk(\sigma) + rk(\tau) - \dim(V)$. //

Definition 1.5.7 *Let $\tau \in \mathcal{L}(F^m, F^n)$ be represented by a matrix T . So the columns of T are $\tau(\mathbf{e}_1), \dots, \tau(\mathbf{e}_m)$. Thus the image of τ is the subspace of F^n spanned by the columns of T . We call this the column space of T and define the column rank of T to be the dimension of this column space.*

We can analogously define the row rank of T to be the dimension of the row space of T .

We will show that the row rank and column rank are equal, and define the rank of T to be this number.

Note that the rank of τ will then be the same as the rank of the associated matrix T (as we'd expect).

Proposition 1.5.8 *row rank(T)=column rank(T).*

Proof: Let T^t be the transpose of T (so $(T^t)_{i,j} = T_{j,i}$.) Therefore row rank(T)=column rank(T^t). Let r be the column rank of T . Then by Proposition 1.5.2, there are bases \mathcal{B} of F^m and \mathcal{C} of F^n with respect to which T (τ) is represented by the matrix

$$D_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

Clearly, D_r^t and D_r each have column rank r , whence the row and column ranks of D_r are equal. By Proposition 1.5.1, this change of basis corresponds to $T = Q^{-1}D_rP$ where P and Q are invertible $m \times m$ and $n \times n$ matrices. Since $T^t = P^tD_r^t(Q^{-1})^t$, we have that T has column (row) rank equal to the column (row) rank of D_r . //

Note that $A\mathbf{x} = \mathbf{b}$ iff $\sum_{j=1}^n x_j\mathbf{c}_j = \mathbf{b}$ where $\mathbf{c}_1, \dots, \mathbf{c}_n$ are the columns of A . Thus $A\mathbf{x} = \mathbf{b}$ has a solution iff \mathbf{b} belongs to the column space of A .

1.6 Column Reduction

Let $B_{i,j}$ be the result of interchanging columns i and j of the identity matrix I_n ; so

$$B_{i,j}\mathbf{e}_k = \begin{cases} \mathbf{e}_k & \text{if } k \neq i, j \\ \mathbf{e}_j & \text{if } k = i \\ \mathbf{e}_i & \text{if } k = j \end{cases}$$

Thus $B_{i,j}B_{i,j} = I_n$ and $B_{i,j}^{-1}$ exists and is equal to $B_{i,j}$. It is easy to verify that if $A \in \mathcal{M}_{n \times n}(F)$, then $AB_{i,j}$ is the matrix obtained as the result of interchanging the i^{th} and j^{th} columns of A and leaving all the other columns of A unchanged — see Exercise 25.

Let $\lambda \neq 0$ be a scalar and $B_i(\lambda)$ be the matrix obtained by multiplying column i of I_n by λ . It is similarly easy to verify that $B_i(\lambda)B_i(1/\lambda) = I_n$ and that $AB_i(\lambda)$ is the $n \times n$ matrix A with column i multiplied by λ — again see Exercise 25.

Let μ be any scalar and $B_{i,j}(\mu)$ be the matrix obtained by taking I_n and adding μ times the j^{th} column to the i^{th} column to form a new i^{th} column. Then $B_{i,j}(\mu)B_{i,j}(-\mu) = I_n$ and that $AB_{i,j}(\mu)$ is the $n \times n$ matrix A with a new i^{th} column obtained by adding μ times the j^{th} column of A to the i^{th} column of A — again see Exercise 25.

Clearly the column space spanned by the columns of A is the same as those spanned by the columns of $AB_{i,j}$, $AB_i(\lambda)$ and $AB_{i,j}(\mu)$ (λ, μ scalars with $\lambda \neq 0$). Moreover, a tedious exercise in bookkeeping (see Proposition 1.6.1 below) shows that there is a sequence of multiplications on the right by these various “elementary” matrices ($B_{i,j}, B_i(\lambda), B_{i,j}(\mu)$) to obtain a matrix in “column echelon form”:

- (i) each non-zero column begins with a (leading) 1;

(ii) all other entries in the same row as a leading 1 are 0; and
 (iii) all columns are 0 or the first column has a leading 1 occurring earlier (higher) than all other leading 1's; if more than one column is non-zero, then the second column has a leading 1 higher than any leading 1's in subsequent columns; etc.

This is called column reduction.

Since each of the elementary matrices has an inverse the column rank of A is the number of leading 1's.

Proposition 1.6.1 *Any matrix can be column reduced to one in column echelon form.*

Proof: If $A \neq 0$, let $a_{i_0, j_0} \neq 0$ be such that $a_{i, j} = 0$ for all $i < i_0$. Then $AB_{j_0}(1/a_{i_0, j_0})B_{1, j_0}$ has 1 in the $(i_0, 1)$ place and 0 in the $(i, j)^{th}$ entry for all $i < i_0$. Let C be the resulting matrix. Then $CB_{j, 1}(-c_{i_0, j})$ has 0 in the (i_0, j) place if $j > 1$. Doing this for $j = 2, 3, \dots, n$ successively gives a matrix $A(1) = D$ with 0 in the (i_0, j) place for all $j > 1$. If there is $d_{i_1, j_1} \neq 0$ for some $j_1 > 1$, choose (i_1, j_1) so that i_1 is least (for all such $j_1 > 1$). [Otherwise, stop.] Just as we got C from A , we now perform $A(1)B_{j_1}(1/a_{i_1, j_1})B_{2, j_1}$ to get $C(1)$, say. Repeat the analogous algorithm which gave $A(1)$ from C (using $C(1)$ instead of C and (i_1, j) for $j > 2$ instead of (i_0, j) for $j > 1$). Let $A(2)$ be the resulting matrix. Then the process $A \mapsto A(1) \mapsto A(2) \mapsto \dots \mapsto A(\ell)$ reduces A to column echelon form. //

Multiplying on the left by an appropriate sequence elementary matrices can be put in row echelon form. Since this row reduction does not change the equations, row reduction can be used to solve $A\mathbf{x} = \mathbf{b}$ if \mathbf{b} belongs to the column space of A : each multiplication on the left by an elementary matrix must be performed on \mathbf{b} too

Example 1.6.2 Solve the system of equations:

$$3x + 4y = 7$$

$$x + y + z = 1$$

$$y - 2z = 3.$$

This is the matrix equation $A\mathbf{x} = \mathbf{b}$ where

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

$\mathbf{x}^t = (x, y, z)$ and $\mathbf{b}^t = (7, 1, 3)$. Now $A(1) = A^t B_{1,2}$ is the matrix

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

and $A(2) = A(1)B_{2,1}(-3)$ is the matrix

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

Then $A(3) = A(2)B_{1,2}(-1)B_{3,2}(-1)$ is the matrix

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

The matrix $A(4) = A(3)B_{1,3}(-4)B_{2,3}(3) = I_3$ which is in row echelon form. Hence $\mathbf{x}^t = \mathbf{b}^t B_{1,2} B_{2,1}(-3) B_{1,2}(-1) B_{3,2}(-1) B_{1,3}(-4) B_{2,3}(3) = (1, 1, -1)^t$.

1.7 Exercises

Note that supervisors are expected to select those questions they think most suitable for their students' needs and abilities. There are 5 sets of exercises for 4 supervisions. My own preference would be to use some of the exercises from Chapter 1 for the first supervision; from Chapter 2 for the second; Chapter 3 for the third; and Chapters 4 & 5 for the fourth. The first 11 questions below are really revision of material from the Algebra & Geometry course, and are only included to help students get started; they should be covered briefly or omitted altogether as supervisors think suitable for their particular supervisees.

-1. Let U be the subset of \mathbb{R}^3 consisting of all vectors \mathbf{x} satisfying the various conditions below. In which of these cases is U a vector space over \mathbb{R} ? (a) $x_1 > 0$. (b) either $x_1 = 0$ or $x_2 = 0$. (c) $x_1 + x_2 = 0$. (d) $x_1 + x_2 = 1$. (e) $x_1 + x_2 + x_3 = 0$ and $x_1 - x_3 = 0$.

-2. Let $F(\mathbb{R}, \mathbb{R})$ be the set of all functions $f : \mathbb{R} \rightarrow \mathbb{R}$. Explain how addition and scalar multiplication are defined for these functions and show that these operations make $F(\mathbb{R}, \mathbb{R})$ into a vector space. Which of the following sets of functions form a vector subspace of $F(\mathbb{R}, \mathbb{R})$? (a) The set C of continuous functions. (b) The set P of all polynomials (with real coefficients). (c) The set $\{f \in C : |f(t)| \leq 1 \text{ for all } t \in [0, 1]\}$. (d) The set $\{f \in C : f(t) \rightarrow 0 \text{ as } t \rightarrow \infty\}$. (e) The set $\{f \in C : |f(t)| \rightarrow \infty \text{ as } |t| \rightarrow \infty\}$. (f) The set $\{f \in C : f(t) \rightarrow 1 \text{ as } t \rightarrow \infty\}$. (g) The set of solutions of the differential equation $\ddot{x}(t) + (t^2 - 3)\dot{x}(t) + t^4x(t) = 0$. (h) The set of solutions of $\ddot{x}(t) + (t^2 - 3)\dot{x}(t) + t^4x(t) = \sin t$. (i) The set of solutions of $(\dot{x}(t))^2 - x(t) = 0$. (j) The set of solutions of $(\ddot{x}(t))^4 + (x(t))^2 = 0$.

-3. Show that the set of all real-valued sequences (x_n) form a vector space over \mathbb{R} . Which of the following subsets are vector subspaces? (a) x_n is bounded. (b) x_n is convergent. (c) $x_n \rightarrow 1$ as $n \rightarrow \infty$. (d) $x_n \rightarrow 0$ as $n \rightarrow \infty$. (e) $x_{n+2} = x_{n+1} + x_n$. (f) There exists m such that $x_n = 0$ for all $n > m$. (g) $\sum |x_n|$ is convergent. (h) $\sum x_n^2$ is convergent.

4. Let T, U, W be subspaces of V . Prove or give counter-examples to the following statements. (a) $T + (U \cap W) = (T + U) \cap (T + W)$. (b) $(T + U) \cap W = (T \cap W) + (U \cap W)$. (c) $(T + U) \cap W = (T \cap W) + (U \cap W)$ if $T \subset W$. (d) $T \cap (U + (T \cap W)) = (T \cap U) + (T \cap W)$.

-5. Which of the following are bases?

(a) For \mathbb{R}^3 : $(1, 1, 0)^t, (0, 1, 1)^t, (1, 0, 1)^t$.

(b) For \mathbb{R}^4 : $(1, 1, 0, 0)^t, (0, 1, 1, 0)^t, (0, 0, 1, 1)^t, (0, 0, 0, 1)^t$.

6. Let $\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_K\}$ and $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_K\}$ be two subsets of the vector space V each containing exactly K elements. Suppose that every vector \mathbf{u}_i is a linear combination of the vectors in \mathcal{V} and that every \mathbf{v}_j is a linear combination of the vectors in \mathcal{U} . Show that $\langle \mathcal{U} \rangle = \langle \mathcal{V} \rangle$ and that \mathcal{U} is linearly independent if, and only if, \mathcal{V} is linearly independent.

-7. Show that if U is a proper subspace of the finite dimensional vector space V , then $\dim(U) < \dim(V)$.

8. Let $\{\mathbf{x}_1, x_2, \dots, \mathbf{x}_r\}$ and $\{\mathbf{y}_1, y_2, \dots, \mathbf{y}_s\}$ be linearly independent subsets of a vector space V , and suppose $r \leq s$. Show that it is possible to choose distinct indices i_1, i_2, \dots, i_r from $\{1, 2, \dots, s\}$ such that, if we delete each \mathbf{y}_{i_j} from Y and replace it by \mathbf{x}_j , the resulting set is still linearly independent.

9. Let

$$U = \{x \in \mathbb{R}^5 : x_1 + x_3 + x_4 = 0, \quad x_1 + x_2 + \frac{1}{2}x_5 = 0\},$$

$$W = \{x \in \mathbb{R}^5 : x_1 + x_5 = 0, \quad x_2 = x_3 = x_4\}.$$

Find bases for U and W containing a basis for $U \cap W$ as a subset. Describe $U + W$ and show that it is given by $\{x \in \mathbb{R}^5 : x_1 + 2x_2 + x_5 = x_3 + x_4\}$.

10. If U_1, \dots, U_r are subspaces of a vector space V , show that the following conditions are equivalent. (i) $\dim \sum_{i=1}^r U_i = \sum_{i=1}^r \dim U_i$; (ii) every element of $\sum_{i=1}^r U_i$ can be uniquely expressed as a sum $\sum_{i=1}^r \mathbf{u}_i$ with $\mathbf{u}_i \in U_i$; (iii) For each j , $U_j \cap \sum_{i \neq j} U_i = \{0\}$. Show that the conditions (i) to (iii) are **not** equivalent to (iv) For each $i \neq j$, $U_i \cap U_j = \{0\}$.

-11. Let P denote the space of all polynomial functions $\mathbb{R} \rightarrow \mathbb{R}$. Which of the following define linear maps $P \rightarrow P$? (a) $D(p)(t) = p'(t)$. (b) $S(p)(t) = p(t^2 + 1)$. (c) $T(p)(t) = p(t)^2 + 1$. (d) $E(p)(t) = p(e^t)$. (e) $J(p)(t) = \int_0^t p(s) ds$. (f) $K(p)(t) = 1 + \int_0^t p(s) ds$. (g) $L(p)(t) = p(0) + \int_0^t p(s) ds$. (h) $M(p)(t) = p(t^2) - tp(t)$. (i) $R(p)$ is the remainder when the polynomial p is divided by the fixed polynomial $t^2 + 1$. (j) $Q(p)$ is the quotient when the polynomial p is divided by the fixed polynomial $t^2 + 1$.

12. For each part of the previous question where the answer is 'yes', find the rank and nullity of the linear map $P_5 \rightarrow P$ (where P_5 denotes the space of polynomials of degree at most 5) obtained by restricting the given linear map to the vector subspace P_5 of P .

13. If α and β are linear maps from U to V , show that $\alpha + \beta$ is linear and that $Im(\alpha + \beta) \subseteq Im(\alpha) + Im(\beta)$ & $ker(\alpha + \beta) \supseteq ker(\alpha) \cap ker(\beta)$.

Show by example that each inclusion may be strict.

14. For each of the following pairs of vector spaces (V, W) over \mathbb{R} , either give an isomorphism $V \rightarrow W$ or show that no such isomorphism can exist. (Here P , as before, denotes the space of polynomial functions $\mathbb{R} \rightarrow \mathbb{R}$, and $C[a, b]$ denotes the space of continuous functions defined on the closed interval $[a, b] \subset \mathbb{R}$.) (a) $V = \mathbb{R}^4$, $W = \{\mathbf{x} \in \mathbb{R}^5 : x_1 + x_2 + x_3 + x_4 + x_5 = 0\}$. (b) $V = \mathbb{R}^5$, $W = \{p \in P : \deg p \leq 5\}$. (c) $V = C[0, 1]$, $W = C[-1, 1]$. (d) $V = C[0, 1]$, $W = \{f \in C[0, 1] : f(0) = 0, f \text{ continuously differentiable}\}$. (e) $V = \mathbb{R}^2$, $W = \{\text{solutions of } \ddot{x}(t) + x(t) = 0\}$. (f) $V = \mathbb{R}^4$, $W = C[0, 1]$. (g) $V = P$, $W = \mathbb{R}^{\mathbb{N}}$, where \mathbb{N} is the natural numbers.

+15. Show that no finite dimensional vector space over \mathbb{R} (or \mathbb{C}) can be written as the union of a finite set of proper subspaces.

What happens if the vector space is not finite dimensional?

+16. (a) The linear map $\alpha : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{2n}$ takes a real $n \times n$ matrix A to $(r_1, \dots, r_n, c_1, \dots, c_n)$ where r_i is the sum of the entries in the i^{th} row of A and c_i is the sum of the entries in the i^{th} column of A . Find the rank of α .

(b) An $n \times n$ magic square is an $n \times n$ matrix of real numbers such that the sum of the entries in each row, in each column, and along either of the main diagonals yields the same answer. Express the set M of magic squares as the kernel of a linear map $\mathbb{R}^{n^2} \rightarrow \mathbb{R}^{2n+1}$, and deduce that it is a vector space. What is its dimension? Find a basis for the space of 3×3 magic squares.

-17. If $V = U \oplus W$, show that the map

$$\pi : V \rightarrow V ; \mathbf{u} + \mathbf{w} \mapsto \mathbf{u} \quad \text{for } \mathbf{u} \in U, \mathbf{w} \in W$$

is a linear map with $\pi^2 = \pi$.

A linear map $\pi : V \rightarrow V$ is a *projection* if $\pi^2 = \pi$. Show that for any projection π the space V is the direct sum of $ker(\pi)$ and $Im(\pi)$.

18. Let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r\}$ be a subset of a vector space V . Show that \mathcal{X} is linearly independent if and only if, for any vector space W and

$\mathcal{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_r\}$ a subset of W , there is a linear map $\alpha : V \rightarrow W$ with $\alpha(\mathbf{x}_i) = \mathbf{y}_i$ for $i = 1, 2, \dots, r$.

Show that \mathcal{X} is a basis if, and only if, there always exists a unique such α .

19. Let $\alpha : V \rightarrow W$ be a linear map between finite-dimensional vector spaces. Show that α is surjective if, and only if, there exists a linear map $\beta : W \rightarrow V$ such that $\alpha\beta$ is the identity on W , and that α is injective if, and only if, there exists $\beta : W \rightarrow V$ such that $\beta\alpha$ is the identity on V .

If $V = W$, show that α is injective if, and only if, it is surjective.

Is this still true if $V = W$ is infinite dimensional?

20. Let V' and W' be subspaces of vector spaces V and W respectively. Show that $T = \{\alpha \in \mathcal{L}(V, W) : \alpha(\mathbf{x}) \in W' \text{ for all } \mathbf{x} \in V'\}$ is a subspace of $\mathcal{L}(V, W)$. Calculate the dimension of T when V and W are finite-dimensional.

21. Let V and W be vector spaces and $\alpha : V \rightarrow W$ be linear. Let $N = \ker(\alpha)$ and B be a basis of N . If $C \supseteq B$ is a basis for V , prove that $\text{Im}(\alpha)$ is isomorphic to $\langle C \setminus B \rangle$ (B and C not necessarily finite).

22. Let $\tau : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be the linear map given by $\tau : \mathbf{x} \mapsto A\mathbf{x}$ where

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

Find the matrix representing τ relative to the basis $(1, 1, 1)^t, (1, 1, 0)^t$ & $(1, 0, 0)^t$ for both the domain and the range.

Find two different bases, one for the domain and the other for the range, so that the matrix representing τ is I_3 .

23. Let $\alpha : U \rightarrow V$ and $\beta : V \rightarrow W$ be maps between finite dimensional vector spaces, and suppose that $\ker(\beta) = \text{Im}(\alpha)$. Show that bases may be chosen for U, V and W with respect to which α and β have matrices

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 \\ 0 & I_{n-r} \end{pmatrix}$$

respectively, where $\dim(V) = n$, $r = rk(\alpha)$.

24. Let $\tau : U \rightarrow V$ be a linear map between two finite dimensional vector spaces and let W be a vector subspace of U . Show that the restriction of τ to W is a linear map $\tau_1 : W \rightarrow V$ which satisfies

$$rk(\tau) \geq rk(\tau_1) \geq rk(\tau) - \dim(U) + \dim(W).$$

Give examples to show that either of the two inequalities can be an equality.

25. Let $B_{i,j}$ be the result of interchanging the i^{th} and j^{th} columns of the identity matrix I_n . Verify that $B_{i,j}^2 = I_n$ and that $AB_{i,j}$ is the $n \times n$ matrix A with columns i and j interchanged. Let $\lambda \neq 0$ be a scalar and $B_i(\lambda)$ be the matrix obtained by multiplying column i of I_n by λ . Verify that $B_i(\lambda)B_i(1/\lambda) = I_n$ and that $AB_i(\lambda)$ is the $n \times n$ matrix A with column i multiplied by λ .

Let λ be any scalar and $B_{i,j}(\lambda)$ be the matrix obtained by taking I_n and adding λ times the j^{th} column to the i^{th} column to form a new i^{th} column. Verify that $B_{i,j}(\lambda)B_{i,j}(-\lambda) = I_n$ and that $AB_{i,j}(\lambda)$ is the $n \times n$ matrix A with a new i^{th} column obtained by adding λ times the j^{th} column of A to the i^{th} column of A .

26. Let V be the vector space of all complex sequences (z_n) which satisfy the difference equation

$$z_{n+2} = 3z_{n+1} - 2z_n \quad \text{for } n = 1, 2, \dots$$

Find a basis for V and determine its dimension. Show that the “shift” operator which sends a sequence (z_1, z_2, z_3, \dots) to (z_2, z_3, z_4, \dots) is a linear map from V to itself. Find the matrix which represents this map relative to your basis. Show that there is a basis for V relative to which the map is represented by a diagonal matrix. What happens if we replace the difference equation by $z_{n+2} = 2z_{n+1} - z_n$?

27. Find the reduced column echelon form of the matrix:

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

and hence describe the space spanned by its columns.

28. Show that the rank of a diagonal square matrix is equal to the number of non-zero entries on the main diagonal. Is the same true for triangular matrices?

Tripes questions on the material in this chapter: 98114, 98206, 99105, 99206, 00206.

[Key: 98206 = Part IB Tripes 1998 paper 2 question 06.]

Chapter 2

Endomorphisms

We now restrict attention to linear maps from a vector space into itself. We can obtain far deeper results about the form of such maps when the vector space is finite dimensional so we will assume throughout this chapter that *all spaces are finite dimensional*.

2.1 Eigenvalues and Eigenvectors

Definition 2.1.1 *Let V be a vector space. Any element of $\mathcal{L}(V, V)$ is called an endomorphism of V .*

To recap: Let V be finite dimensional with basis \mathcal{B} and τ be an endomorphism of V . Then τ can be associated with a matrix T where $\tau(\mathbf{b}_j) = \sum_{i=1}^n t_{i,j} \mathbf{b}_i$ ($j = 1, \dots, n$) where $n = \dim(V)$. If we change bases to \mathcal{C} with $\mathbf{c}_k = \sum_{i=1}^n p_{i,k} \mathbf{b}_i$, then Proposition 1.5.1 gives $\tau(\mathbf{c}_k) = \sum_{i=1}^n (P^{-1}TP)_{i,k} \mathbf{c}_i$ ($k = 1, \dots, n$).

So if τ is represented by matrix T with respect to one basis, then the conjugates of T represent τ with respect to all other bases.

GOAL: Choose a basis to make τ transparent.

In Proposition 1.5.2 we saw that we could choose (possibly different) bases \mathcal{B} and \mathcal{C} of V so that τ is represented by

$$D_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

That is, $Q^{-1}TP = D_r$, where $r = rk(\tau)$.

Can we do this with $Q = P$?
 If $P^{-1}TP = D_r$, then

$$T = \begin{pmatrix} P_1 I_r P_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} = D_r;$$

so only in very special cases when τ is the identity on a subspace of V complementary to $\ker(\tau)$.

A more realistic goal might be to find P so that $P^{-1}TP$ is diagonal, for then τ would just be a sequence of dilations and reflections.

As you already know from Algebra & Geometry, this cannot always be achieved in \mathbb{R}^2 , but we wish to determine for *which* endomorphisms it can be done (for an arbitrary vector space of dimension n).

Definition 2.1.2 Let V be a finite dimensional vector space and τ be an endomorphism of V . $\lambda \in F$ is said to be an eigenvalue of τ if $\tau(\mathbf{v}_0) = \lambda\mathbf{v}_0$ for some $\mathbf{v}_0 \neq \mathbf{0}$. Such a vector \mathbf{v}_0 is called an eigenvector of τ (with eigenvalue λ).

If λ is an eigenvalue of τ , then the set of all eigenvectors of τ (which have eigenvalue λ) together with $\mathbf{0}$ forms a subspace of V (Verify this) that is called an eigenspace of τ with eigenvalue λ . The subspace will be denoted by $\mathcal{E}_\tau(\lambda)$, or simply $\mathcal{E}(\lambda)$ if τ is clear from context.

If λ, μ are distinct eigenvalues of an endomorphism τ , then $\mathcal{E}(\lambda) \cap \mathcal{E}(\mu) = \{\mathbf{0}\}$ (for if $\mathbf{v} \in \mathcal{E}(\lambda) \cap \mathcal{E}(\mu)$, then $\lambda\mathbf{v} = \tau(\mathbf{v}) = \mu\mathbf{v}$; since $\lambda \neq \mu$, we have $\mathbf{v} = \mathbf{0}$).

If $P^{-1}TP$ is diagonal, say $\text{diag}(\lambda_1, \dots, \lambda_n)$, with respect to basis \mathcal{B} , then each $\mathbf{b}_j \in \mathcal{E}(\lambda_j)$, $j = 1, \dots, n$. So we have a basis of eigenvectors. Moreover, if $\{\lambda_1, \dots, \lambda_n\} = \{\mu_1, \dots, \mu_\ell\}$ with $\mu_i \neq \mu_k$ if $i \neq k$, then $V = \mathcal{E}(\mu_1) \oplus \dots \oplus \mathcal{E}(\mu_\ell)$ (since clearly $\mathcal{E}(\mu_i) \cap \mathcal{E}(\mu_k) = \{\mathbf{0}\}$ if $i \neq k$).

Conversely, let τ be represented by a matrix T with respect to a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$. If there is a basis \mathcal{B} of V comprising eigenvectors of τ (with eigenvalues $\lambda_1, \dots, \lambda_n$), then let P be such that $P(\mathbf{e}_i) = \mathbf{b}_i$. Now $P^{-1}TP = \text{diag}(\lambda_1, \dots, \lambda_n)$ represents τ with respect to \mathcal{B} . Thus

Proposition 2.1.3 T/τ is diagonalisable iff there is a basis \mathcal{B} of V comprising eigenvectors of τ . The diagonal matrix is just the set of eigenvalues of τ (appearing $\dim(\mathcal{E}(\lambda))$ times) and is unique to within the order on the diagonal.

Corollary 2.1.4 *Let μ_1, \dots, μ_ℓ be the distinct eigenvalues of an endomorphism τ of V . Let $\mathcal{E} = \langle \bigcup_{i=1}^{\ell} \mathcal{E}(\mu_i) \rangle$. Then $\mathcal{E} = \mathcal{E}(\mu_1) \oplus \dots \oplus \mathcal{E}(\mu_\ell)$, and τ is diagonalisable iff $\mathcal{E} = V$.*

Thus τ is diagonalisable iff $V = \mathcal{E}(\mu_1) \oplus \dots \oplus \mathcal{E}(\mu_\ell)$ for some distinct μ_1, \dots, μ_ℓ .

Definition 2.1.5 *If μ is an eigenvalue of τ , then the geometric multiplicity of μ is $\dim(\mathcal{E}(\mu))$.*

Note that each $\mathcal{E}(\mu_j)$ is mapped by τ to itself and the restriction of τ to this subspace is just dilation by μ_j (i.e., $\mu_j I_{d_j}$, where d_j is the geometric multiplicity of μ_j). Moreover, τ is diagonalisable iff the sum of the geometric multiplicities equals the dimension of V .

We now seek other equivalent conditions for a matrix/linear transformation to be diagonalisable.

2.2 The Minimal Polynomial

We now come to the main tool in the course.

Let σ, τ be endomorphisms of V . As before, we define $\sigma\tau$ to be the composition $\sigma \circ \tau$, an endomorphism of V . We write i_V for the identity endomorphism of V .

Proposition 2.2.1 *Let V be a vector space of dimension n . Then $\mathcal{M}_{n \times n}(F)$ has dimension n^2 . Thus $\mathcal{L}(V, V)$ has dimension n^2 .*

Proof: Let $E_{i,j}$ be the matrix with (k, m) entry $\delta_{ik}\delta_{jm}$. Then, as is easily verified, $\{E_{i,j} : 1 \leq i, j \leq n\}$ spans $\mathcal{M}_{n \times n}(F)$ and is linearly independent. By Proposition 1.1.14, $\mathcal{L}(V, V) \cong \mathcal{M}_{n \times n}(F)$, so the result follows. //

If $p(X) = a_0 + a_1X + \dots + a_mX^m$ is a polynomial with coefficients in F , we let $p(\tau) = a_0i_V + a_1\tau + \dots + a_m\tau^m \in \mathcal{L}(V, V)$ and $p(T) = a_0I + a_1T + \dots + a_mT^m$ where T is an $n \times n$ matrix ($n = \dim(V)$).

Since τ commutes with $p(\tau)$ under composition, any two polynomials in τ commute under composition.

Proposition 2.2.2 *Let V be a vector space over F with $\dim(V) = n$ and τ be an endomorphism of V . Then $p(\tau) = 0$ for some non-zero polynomial $p(X) \in F[X]$.*

Proof: By Proposition 2.2.1, $\mathcal{L}(V, V)$ has dimension n^2 ; whence the subset $\{i_V, \tau, \tau^2, \dots, \tau^{n^2}\}$ with $n^2 + 1$ elements must be linearly dependent. So there are $a_0, \dots, a_{n^2} \in F$, not all 0, with $\sum_{j=0}^{n^2} a_j \tau^j = 0$. Hence $p(\tau) = 0$ where $p(X) = \sum_{j=0}^{n^2} a_j X^j \in F[X]$. //

Among the non-zero polynomials $p(X) \in F[X]$ which have τ as a root, choose one with minimal degree. Divide by the coefficient of the highest term and let $m_\tau(X)$ be the resulting *monic* polynomial (leading coefficient is 1). If $p(\tau) = 0$, then by the Division Theorem for polynomials over the field F , we get $p(X) = m_\tau(X)q(X) + r(X)$ for some $q(X), r(X) \in F[X]$ with $r(X) = 0$ or $\deg(r(X)) < \deg(m_\tau(X))$. Since $p(\tau) = m_\tau(\tau) = 0$, we have $r(\tau) = 0$. By the minimality of the degree of m_τ , $r(X) = 0$; i.e., $m_\tau(X) | p(X)$. Thus:

Proposition 2.2.3 $p(\tau) = 0$ iff $m_\tau(X) | p(X)$.

Definition 2.2.4 *The unique monic polynomial $m_\tau(X)$ is called the minimal polynomial for τ .*

Note that $p(Q^{-1}TQ) = Q^{-1}p(T)Q$; so the minimal polynomial of the matrix representing an endomorphism is independent of the basis chosen.

Proposition 2.2.5 *Let V be a finite dimensional vector space and τ be an endomorphism of V . Then the eigenvalues of τ are precisely the roots of the minimal polynomial for τ .*

Proof: If λ is an eigenvalue of τ with eigenvector \mathbf{w} , then we have $\tau(\mathbf{w}) = \lambda\mathbf{w}$, $(\tau^2)(\mathbf{w}) = \tau(\lambda\mathbf{w}) = \lambda^2\mathbf{w}$, and more generally, $(\tau^k)(\mathbf{w}) = \lambda^k\mathbf{w}$ for all positive integers k (by induction on k). Hence $\mathbf{0} = m_\tau(\tau)(\mathbf{w}) = m_\tau(\lambda)\mathbf{w}$. Since $\mathbf{w} \neq \mathbf{0}$, it follows that $m_\tau(\lambda) = 0$; i.e., λ is a root of m_τ .

Conversely, by the Division Theorem, $m_\tau(X) = (X - \lambda)p(X) + r$ for some polynomial $p(X)$ and $r \in F$. If $m_\tau(\lambda) = 0$, then $r = 0$; so

$m_\tau(X) = (X - \lambda)p(X)$. Now $p(X)$ is a proper factor of $m_\tau(X)$. By the minimality of $m_\tau(X)$, there must be $\mathbf{v} \in V$ such that $\mathbf{w} \equiv p(\tau)(\mathbf{v}) \neq \mathbf{0}$. Since $(\tau - \lambda i_V)\mathbf{w} = (\tau - \lambda i_V)p(\tau)(\mathbf{v}) = m_\tau(\tau)(\mathbf{v}) = \mathbf{0}$, we get that λ is an eigenvalue of τ (with \mathbf{w} as corresponding eigenvector). //

Suppose that τ is diagonalisable. Then $V = \mathcal{E}(\mu_1) \oplus \dots \oplus \mathcal{E}(\mu_\ell)$ with μ_1, \dots, μ_ℓ distinct. But $(\tau - \mu_j i_V)\mathbf{w} = \mathbf{0}$ for all $\mathbf{w} \in \mathcal{E}(\mu_j)$. If $\mathbf{v} \in V$, then $\mathbf{v} = \sum_{j=1}^{\ell} \mathbf{w}_j$ with $\mathbf{w}_j \in \mathcal{E}(\mu_j)$ for $j = 1, \dots, \ell$. Therefore, $\prod_{j=1}^{\ell} (\tau - \mu_j i_V)\mathbf{v} = \prod_{j=1}^{\ell} (\tau - \mu_j i_V)(\sum_{k=1}^{\ell} \mathbf{w}_k) = \sum_{k=1}^{\ell} \prod_{j=1}^{\ell} (\tau - \mu_j i_V)(\mathbf{w}_k) = \sum_{k=1}^{\ell} [\prod_{j \neq k} (\tau - \mu_j i_V)](\tau - \mu_k i_V)(\mathbf{w}_k) = \mathbf{0}$ for all $\mathbf{v} \in V$.

Hence $m_\tau(X) \mid \prod_{j=1}^{\ell} (X - \mu_j)$, and equality follows from Proposition 2.2.5. Thus we have established

Proposition 2.2.6 *Let V be a finite dimensional vector space and τ be a diagonalisable endomorphism of V . If μ_1, \dots, μ_ℓ are the distinct eigenvalues of τ , then $m_\tau(X) = \prod_{j=1}^{\ell} (X - \mu_j)$. In particular, the minimal polynomial is a product of polynomials of degree 1 none of which are repeated.*

Example 2.2.7 Let

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Then $(X - 1)^2$ is the minimal polynomial for T . By Proposition 2.2.6, T is not diagonalisable (even over \mathbb{C}).

Our next goal is to establish the converse of Proposition 2.2.6, that **diagonalisability of τ is completely equivalent to a factorisation property of the minimal polynomial for τ** .

Proposition 2.2.8 *Let V be a finite dimensional vector space and τ be an endomorphism of V . Then τ is diagonalisable iff $m_\tau(X)$ is a product of polynomials of degree 1 none of which are repeated.*

To prove Proposition 2.2.8 we need a fact about polynomials:

Proposition 2.2.9 *If F is a field and $f(X), g(X) \in F[X]$ are not zero and have no common divisors of positive degree, then there are $h(X), k(X) \in F[X]$ such that $f(X)h(X) + g(X)k(X) = 1$.*

Proof: By induction on $\min\{\deg(f), \deg(g)\}$. Without loss of generality, $\deg(f) \leq \deg(g)$. If $\deg(f) = 0$, let $k(X) = 0$ and $h(X) = 1/f$; so assume that $\deg(f) > 0$. Then $g(X) = f(X)q(X) + r(X)$ where $\deg(r) < \deg(f)$. Note that $r \neq 0$ since $f(X)$ does not divide $g(X)$ by hypothesis. Now $r(X)$ and $f(X)$ share no common non-constant factor (otherwise it would also divide $g(X)$). By the inductive hypothesis, $1 = r(X)k_1(X) + f(X)k_2(X)$ for some $k_1(X), k_2(X) \in F[X]$. Substituting back for $r(X)$ gives $1 = f(X)[k_2(X) - q(X)k_1(X)] + g(X)k_1(X)$. //

Note that if $f(X) \in F[X]$, τ is an endomorphism of a finite dimensional vector space V , and $U = \ker(f(\tau))$, then $\tau(U) \subseteq U$: if $\mathbf{u} \in U$, then $f(\tau)(\tau(\mathbf{u})) = \tau(f(\tau)(\mathbf{u})) = \tau(\mathbf{0}) = \mathbf{0}$.

Let V be a finite dimensional vector space, τ an endomorphism of V and λ an eigenvalue of τ . Then $m_\tau(X) = (X - \lambda)^m p(X)$ for some polynomial $p(X)$ sharing no common factor with $X - \lambda$ (i.e., $p(\lambda) \neq 0$).

Let $U = \ker((\tau - \lambda i_V)^m) \supseteq \mathcal{E}(\lambda)$. So $\tau(U) \subseteq U$. If $W = \ker(p(\tau))$, then $\tau(W) \subseteq W$.

Proposition 2.2.10 *With the above notation, $V = U \oplus W$.*

Proof: By Proposition 2.2.9, there are $h(X), k(X) \in F[X]$ such that $1 = h(X)p(X) + k(X)(X - \lambda)^m$. Let $\mathbf{v} \in U \cap W$. Then $p(\tau)(\mathbf{v}) = \mathbf{0} = (\tau - \lambda i_V)^m(\mathbf{v})$. Hence $\mathbf{v} = 1(\mathbf{v}) = h(\tau)p(\tau)(\mathbf{v}) + k(\tau)(\tau - \lambda i_V)^m(\mathbf{v}) = \mathbf{0}$. Thus $U \cap W = \{\mathbf{0}\}$.

Since any $\mathbf{v} \in V$ satisfies $1(\mathbf{v}) = h(\tau)p(\tau)(\mathbf{v}) + k(\tau)(\tau - \lambda i_V)^m(\mathbf{v})$, and the first summand belongs to U (because $p(X)(X - \lambda)^m = m_\tau(X)$) and the second to W (same reason), we obtain the desired result. //

Let $V = U \oplus W$ with $\tau(U) \subseteq U$ and $\tau(W) \subseteq W$. If $\alpha = \tau_U$ and $\beta = \tau_W$, then α and β are linear and $\tau(\mathbf{u} + \mathbf{w}) = \alpha(\mathbf{u}) + \beta(\mathbf{w})$ for all $\mathbf{u} \in U$ and $\mathbf{w} \in W$. We will therefore write $\tau = \alpha \oplus \beta$. It is immediately seen that $\tau^2 = \alpha^2 \oplus \beta^2$, etc., whence $p(\tau) = p(\alpha) \oplus p(\beta)$ for any polynomial p . If α is represented by a matrix A (with respect to a basis of U) and β is represented by a matrix B (with respect to a basis of W), then τ is represented by a matrix T (with respect to the resulting basis of $U \oplus W$) where

$$T = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

In this case we call T the direct sum of the matrices A and B and write $T = A \oplus B$.

Proof of Proposition 2.2.8: By induction on the degree of $m_\tau(X)$. Assume that $m_\tau(X)$ is a product of distinct linear terms. If $m_\tau(X) = X - \mu$, then $\tau = \mu i_V$ which is diagonal with respect to every basis. If $m_\tau(X) = \prod_{j=1}^{\ell+1} (X - \mu_j)$, then by Proposition 2.2.10, $V = U \oplus W$ where $U = \ker(\prod_{j=1}^{\ell} (\tau - \mu_j i_V))$ and $W = \ker(\tau - \mu_{\ell+1} i_V)$. By induction, $\alpha = \tau_U$ is diagonalisable (A) and $\beta = \tau_W = \mu_{\ell+1} i_W$ is too ($B = \mu I$). Hence so is τ ($T = A \oplus B$). //

Proposition 2.2.11 (Existence of Eigenvalues over \mathbb{C}) *Let V be a vector space over \mathbb{C} with $\dim(V) = n \geq 1$. Any endomorphism τ of V has an eigenvalue.*

Proof: Since every non-constant polynomial over \mathbb{C} is a product of polynomials of degree 1, the result follows from Proposition 2.2.5. //

Corollary 2.2.12 *Let τ be an endomorphism of a finite dimensional vector space V over \mathbb{C} with μ_1, \dots, μ_ℓ as its distinct eigenvalues. Then $m_\tau(X) = \prod_{j=1}^{\ell} (X - \mu_j)^{d_j}$ for some positive integers d_1, \dots, d_ℓ and $V = U_1 \oplus \dots \oplus U_\ell$ where $U_j = \ker((\tau - \mu_j i_V)^{d_j})$ for $j = 1, \dots, \ell$. Moreover $\tau(U_j) \subseteq U_j$ and $\tau = \alpha_1 \oplus \dots \oplus \alpha_\ell$ where $\alpha_j = \tau_{U_j}$ for $j = 1, \dots, \ell$ and has minimal polynomial $(X - \mu_j)^{d_j}$.*

Consider the orthogonal matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Viewed as a matrix over \mathbb{R} , there are no eigenvalues if $\theta \notin \pi\mathbb{Z}$ since each vector $\mathbf{v} \neq \mathbf{0}$ is rotated by θ and so does not belong to the subspace generated by \mathbf{v} . Hence the matrix is not diagonalisable over \mathbb{R} . However, over \mathbb{C} , there are two eigenvalues $e^{i\theta}$ and $e^{-i\theta}$ and the matrix is diagonalisable (do this).

Note that the minimal polynomial over \mathbb{R} is $X^2 - 2X\cos\theta + 1$ and this has no roots in \mathbb{R} ; over \mathbb{C} , the minimal polynomial is the same and can be written as $(X - e^{i\theta})(X - e^{-i\theta})$.

More generally, let V be a vector space over \mathbb{R} and $\tau \in \mathcal{L}(V, V)$. Let $m_\tau(X) = a_0 + a_1X + \dots + a_{n-1}X^{n-1} + X^n \in \mathbb{R}[X] \subseteq \mathbb{C}[X]$. Then $m_\tau(X)$ factors into linear pieces in $\mathbb{C}[X]$ by Gauss' result; say $m_\tau(X) = \prod_j (X - \lambda_j)$. Since $m_\tau(\lambda_j) = 0$ for any j , we have $a_0 + a_1\lambda_j + \dots + a_{n-1}\lambda_j^{n-1} + \lambda_j^n = 0$. Taking complex conjugates of both sides and using $\bar{a}_k = a_k$ for all k (the a_k are real), we get that $m_\tau(\bar{\lambda}_j) = 0$. But $(X - \bar{\lambda}_j)(X - \lambda_j) = X^2 - 2\operatorname{Re}(\lambda_j)X + |\lambda_j|^2 \in \mathbb{R}[X]$. Thus

Proposition 2.2.13 *Let V be a finite dimensional vector space over \mathbb{R} and τ be an endomorphism of V . Then $m_\tau(X) = \prod_s (X - \nu_s) \cdot \prod_m (X^2 - 2\operatorname{Re}(\lambda_m)X + |\lambda_m|^2)$ where ν_s are the real eigenvalues of τ and $\lambda_m \in \mathbb{C} \setminus \mathbb{R}$ are the complex non-real eigenvalues of τ .*

So the eigenvalues of τ (for V) are precisely the ν_s 's, and $\lambda \in \mathbb{C}$ is a root of $m_\tau(X)$ iff $\bar{\lambda}$ is.

Definition 2.2.14 *An $n \times n$ matrix T is said to be upper triangular if every entry below the diagonal is 0; i.e., $t_{i,j} = 0$ if $i > j$.*

Although not every linear transformation is diagonalisable (even over \mathbb{C}) (please see Example 2.2.7), we do have:

Proposition 2.2.15 *Let V be a finite dimensional vector space over \mathbb{C} and $\tau \in \mathcal{L}(V, V)$. Then τ can be represented by an upper triangular matrix with respect to some basis of V .*

Equivalently, if T is an $n \times n$ complex matrix, then there is an invertible $n \times n$ complex matrix P such that $P^{-1}TP$ is upper triangular.

Proof. This proof is typical of many that are used in the course. Induction on $n = \dim(V)$. If $n = 1$, then $\tau = \lambda i_V$ and we're done. So assume that the result is true for any endomorphism of a vector space W of dimension $< n$. By Proposition 2.2.11, τ has an eigenvalue λ_1 . Let $\mathbf{b}_1 \in \mathcal{E}(\lambda_1)$, and W be a complementary subspace to $\langle \mathbf{b}_1 \rangle$ in V . Then $\tau(\mathbf{w}) = \phi(\mathbf{w})\mathbf{b}_1 + \psi(\mathbf{w})$ for some $\phi : W \rightarrow \mathbb{C}$ and $\psi : W \rightarrow W$. Since τ is linear and $V = \langle \mathbf{b}_1 \rangle \oplus W$, both ϕ and ψ are linear. Hence ψ is an endomorphism of W with $\dim(W) < n$. By induction, there is a basis $\mathcal{B} = \{\mathbf{b}_2, \dots, \mathbf{b}_n\}$ of W with ψ represented by an upper triangular matrix M . Then, with respect to $\{\mathbf{b}_1\} \cup \mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, τ is represented by the upper triangular matrix

$$T = \begin{pmatrix} \lambda_1 & \phi(\mathcal{B}) \\ 0 & M \end{pmatrix}$$

//

The complex numbers are special in that every non-zero polynomial is a product of linear factors. We can generalise Proposition 2.2.15 to get a condition for a representation as an upper triangular matrix in terms of the minimal polynomial even when the field is not \mathbb{C} . The condition is quite similar to the condition for diagonalisability (please see Proposition 2.2.8).

Proposition 2.2.16 *Let V be a finite dimensional vector space over a field F and τ be an endomorphism of V . Then τ can be represented by an upper triangular matrix iff $m_\tau(X)$ is a product of (not necessarily distinct) linear factors.*

Hence an $n \times n$ matrix T is conjugate to an upper triangular matrix iff $m_T(X)$ is a product of linear factors.

To prove Proposition 2.2.16, we need a lemma whose proof requires a new idea.

Let V be a vector space of dimension n and τ be an endomorphism of V . Let W be a proper subspace of V with $\tau(W) \subseteq W$. Let $\mathbf{v} \in V \setminus W$. Since $m_\tau(\tau)(\mathbf{v}) = \mathbf{0} \in W$, there is a monic polynomial $m_{\tau, \mathbf{v}, W}(X)$ of minimal degree such that $(m_{\tau, \mathbf{v}, W}(\tau))(\mathbf{v}) \in W$. As before, $p(\tau)(\mathbf{v}) \in W$ iff $m_{\tau, \mathbf{v}, W}(X) | p(X)$. Hence if W_1 is a subspace of W and $\mathbf{v} \in V \setminus W$, then $m_{\tau, \mathbf{v}, W}(X) | m_{\tau, \mathbf{v}, W_1}(X)$. In the special case that $W_1 = \{0\}$, we get the obvious fact that $m_{\tau, \mathbf{v}, W}(X) | m_{\tau, \mathbf{v}, \{0\}}(X) | m_\tau(X)$.

Lemma 2.2.17 *Let W be a proper subspace of V and $\tau \in \mathcal{L}(V, V)$ with $\tau(W) \subseteq W$. If $m_\tau(X) = \prod_{j=1}^{\ell} (X - \mu_j)^{d_j}$ with μ_1, \dots, μ_ℓ distinct, then there is $\mathbf{v} \in V \setminus W$ and $J \in \{1, \dots, \ell\}$ such that $(\tau - \mu_J i_V)\mathbf{v} \in W$.*

Proof: Let $\mathbf{u} \in V \setminus W$. Then $m_{\mathbf{u}, W}(X) = \prod_{j=1}^{\ell} (X - \mu_j)^{k_j}$ for some k_1, \dots, k_ℓ , since $m_{\mathbf{u}, W}(X)$ divides $m_\tau(X)$. Moreover, since $\mathbf{u} \notin W$, some $k_J > 0$. Let $q_J(X) = \prod_{j \neq J} (X - \mu_j)^{k_j} \cdot (X - \mu_J)^{k_J - 1}$, a polynomial. Then $\mathbf{v} = q_J(\tau)(\mathbf{u}) \notin W$ (by the minimality of $m_{\mathbf{u}, W}$) but $(\tau - \mu_J i_V)(\mathbf{v}) = m_{\mathbf{u}, W}(\tau)(\mathbf{u}) \in W$. //

Proof of Proposition 2.2.16: If τ can be represented by an upper triangular matrix S (with respect to some basis), then $\prod_{j=1}^n (S - s_{j,j} I_n)$ has 0 on the diagonal and below. Hence $\prod_{j=1}^n (S - s_{j,j} I_n)^n = 0$ (verify

this), whence $m_\tau(X) \mid \prod_{j=1}^n (X - s_{j,j})^n$. Thus $m_\tau(X)$ is a product of linear factors.

Conversely, if $m_\tau(X) = \prod_{j=1}^m (X - \mu_j)^{d_j}$, let $W_0 = \{\mathbf{0}\}$. By Lemma 2.2.17, there is $\mathbf{b}_1 \neq \mathbf{0}$ such that $(\tau - \mu_{\ell_1} i_V)(\mathbf{b}_1) = \mathbf{0}$ for some ℓ_1 . Thus $\tau(\mathbf{b}_1) = \mu_{\ell_1} \mathbf{b}_1$. Hence $W_1 = \langle \mathbf{b}_1 \rangle$ satisfies the hypotheses of Lemma 2.2.17. Applying the lemma gives $\mathbf{b}_2 \notin W_1$ with $(\tau - \mu_{\ell_2} i_V)(\mathbf{b}_2) \in W_1$. That is, $\tau(\mathbf{b}_2) = s_{1,2} \mathbf{b}_1 + \mu_{\ell_2} \mathbf{b}_2$. Continue the process with $W_2 = \langle \mathbf{b}_1, \mathbf{b}_2 \rangle$, etc. With respect to the constructed basis, τ is represented by the upper triangular matrix S where

$$S_{i,j} = \begin{cases} \mu_{\ell_i} & \text{if } i = j \\ s_{i,j} & \text{if } i < j \\ 0 & \text{if } i > j. \end{cases}$$

//

Caution: Example 2.2.7 can easily be extended to give an endomorphism τ of \mathbb{C}^3 that is not diagonalisable but whose restriction to a specific one dimensional subspace is diagonalisable. Equally, it is possible to give a diagonalisable endomorphism τ of \mathbb{C}^2 with $\mathbb{C}^2 = \mathcal{E}_\tau(\mu_1) \oplus \mathcal{E}_\tau(\mu_2)$ and a one dimensional subspace W of \mathbb{C}^2 such that $W \neq (W \cap \mathcal{E}_\tau(\mu_1)) \oplus (W \cap \mathcal{E}_\tau(\mu_2))$: let $\tau(\mathbf{e}_1) = \mathbf{e}_1$, $\tau(\mathbf{e}_2) = -\mathbf{e}_2$ and $W = \{(x, x) : x \in \mathbb{C}\}$.

However, if $\tau(W) \subseteq W$, the situation is more propitious.

Notation: If W is a subspace of a vector space V and $\tau \in \mathcal{L}(V, U)$, then τ_W is the restriction of τ to W . So if $\tau(W) \subseteq W$, then τ_W is an endomorphism of W .

Proposition 2.2.18 *Let V be a finite dimensional vector space and W be a subspace of V . Let τ be an endomorphism of V such that $\tau(W) \subseteq W$. Then τ_W is an endomorphism of W and $m_{\tau_W}(X) \mid m_\tau(X)$. Hence τ_W is diagonalisable/upper triangularisable if τ is. In particular, if $\tau(W) \subseteq W$ and $V = \mathcal{E}_\tau(\mu_1) \oplus \dots \oplus \mathcal{E}_\tau(\mu_\ell)$, then $W = (W \cap \mathcal{E}_{\tau_W}(\mu_1)) \oplus \dots \oplus (W \cap \mathcal{E}_{\tau_W}(\mu_\ell))$.*

Proof: A fortiori, $m_\tau(\tau_W)(\mathbf{w}) = m_\tau(\tau)(\mathbf{w}) = \mathbf{0}$ for all $\mathbf{w} \in W$, whence $m_{\tau_W}(X) \mid m_\tau(X)$. The Proposition now follows using Propositions 2.2.8 and 2.2.16. //

Suppose that we are given a set of diagonalisable matrices. When can we simultaneously diagonalise them?

Note that $P^{-1}SP$ and $P^{-1}TP$ commute iff S and T commute. Since diagonal matrices commute we have that if S and T are simultaneously diagonalisable, then they must commute. We now prove the converse; this requires a better understanding of how to divide a vector space into subspaces that are respected by a given linear transformation.

Proposition 2.2.19 *Let \mathcal{T} be a commuting family of diagonalisable endomorphisms of a finite dimensional vector space V . Then there is a basis of V with respect to which each $\tau \in \mathcal{T}$ is represented by a diagonal matrix. Equivalently, let \mathcal{M} be a commuting family of $n \times n$ diagonalisable matrices. Then there is an invertible $n \times n$ matrix P such that $P^{-1}MP$ is diagonal for all $M \in \mathcal{M}$.*

To prove the Proposition we need a trivial but important lemma:

Lemma 2.2.20 *Let V be a vector space and $\sigma, \tau \in \mathcal{L}(V, V)$ commute. If μ is an eigenvalue of τ , then $\sigma(\mathcal{E}_\tau(\mu)) \subseteq \mathcal{E}_\tau(\mu)$.*

Proof: Let $\mathbf{v} \in \mathcal{E}_\tau(\mu)$. Then $\tau(\sigma(\mathbf{v})) = \sigma(\tau(\mathbf{v})) = \sigma(\mu\mathbf{v}) = \mu\sigma(\mathbf{v})$ //

Proof of Proposition 2.2.19: Since $\mathcal{M}_{n \times n}(F)$ has dimension n^2 , there is a finite subset M_1, \dots, M_k of \mathcal{M} such that $\mathcal{M} \subseteq \langle M_1, \dots, M_k \rangle$. If M_1, \dots, M_k are simultaneously diagonalisable, then so is any linear combination; i.e., so is \mathcal{M} . It is therefore enough to show that $\{M_1, \dots, M_k\}$ can be simultaneously diagonalised. This we do by induction on k . The result is trivially true if $k = 1$, so we proceed to the induction step assuming the result for k . Let $\tau_1, \dots, \tau_{k+1}$ be the endomorphisms corresponding to M_1, \dots, M_{k+1} respectively. Now

$$V = \mathcal{E}_{\tau_{k+1}}(\lambda_1) \oplus \dots \oplus \mathcal{E}_{\tau_{k+1}}(\lambda_s)$$

for some $\lambda_1, \dots, \lambda_s \in F$ (since τ_{k+1} is diagonalisable). Let $W_j = \mathcal{E}(\lambda_j) = \mathcal{E}_{\tau_{k+1}}(\lambda_j)$ ($j = 1, \dots, s$). Then $\tau_{i,j} = (\tau_i)_{W_j}$ is an endomorphism of W_j by Lemma 2.2.20 for all i, j . Furthermore, each $\tau_{i,j}$ is diagonalisable by Proposition 2.2.18. Hence, by induction on k , we can find bases $\mathcal{B}_1, \dots, \mathcal{B}_s$ of W_1, \dots, W_s respectively such that $\tau_{i,j}$ is diagonal on W_j (with respect to \mathcal{B}_j) for $j = 1, \dots, s$ and $i = 1, \dots, k$. Since $\tau_{k+1,j}$ is λ_j times the identity, it is diagonal on W_j with respect to *any* basis. Thus, for $\mathcal{B} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_s$, each of $\tau_1, \dots, \tau_{k+1}$ is represented by a diagonal matrix (and so the same

is true of M_1, \dots, M_{k+1} where P is the change of basis from the original to \mathcal{B}). //

One can analogously prove:

Proposition 2.2.21 *Let \mathcal{T} be a commuting family of upper triangularisable endomorphisms of a finite dimensional vector space V . Then there is a basis of V with respect to which each $\tau \in \mathcal{T}$ is represented by an upper triangular matrix. Equivalently, let \mathcal{M} be a commuting family of $n \times n$ upper triangularisable matrices. Then there is an invertible $n \times n$ matrix P such that $P^{-1}MP$ is upper triangular for all $M \in \mathcal{M}$.*

Corollary 2.2.22 *let \mathcal{M} be a commuting family of complex $n \times n$ matrices. Then there is an invertible $n \times n$ matrix P such that $P^{-1}MP$ is upper triangular for all $M \in \mathcal{M}$.*

To prove Proposition 2.2.21, the required lemma is

Lemma 2.2.23 *Let V be a finite dimensional vector space and $\sigma_1, \dots, \sigma_m$ be linearly independent pairwise commuting upper triangularisable endomorphisms of V . Let W be a proper subspace of V with $\sigma_j(W) \subseteq W$ for $j = 1, \dots, m$. Then there is $\mathbf{v} \in V \setminus W$ such that $\sigma_j(\mathbf{v}) \in \langle \mathbf{v}, W \rangle$ for all $j = 1, \dots, m$.*

The proof of Lemma 2.2.23 the subsequent deduction of Proposition 2.2.21 are left to the reader.

2.3 summary

An endomorphism τ of a finite dimensional vector space has a minimal polynomial $m_\tau(X)$.

- (1) The roots of $m_\tau(X)$ are precisely the eigenvalues of τ .
- (2) τ is diagonalisable iff $m_\tau(X)$ is a product of distinct linear factors.
- (3) τ is upper triangularisable iff $m_\tau(X)$ is a product of linear factors.

2.4 Exercises

1. Find the eigenvalues of the following complex matrices:

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

In each case, give a basis for the eigenspaces. If the matrix is diagonalizable, find a conjugate diagonal matrix. If it is not diagonal, find a conjugate upper-triangular matrix.

2. Show that $m_A(X) = m_{A^t}(X)$ for every $n \times n$ matrix A .
3. Let A be an $n \times n$ matrix in which the sum of each row is 1. Show that 1 is an eigenvalue of A . Show that 1 is also an eigenvalue of the transposed matrix A^t .
4. Let A be an $n \times n$ matrix all the entries of which are real. Show that the minimal polynomial (over the complex numbers) has real coefficients.
5. Prove that any two real matrices that are conjugate over \mathbb{C} are conjugate over \mathbb{R} .
6. Let A and B be $n \times n$ matrices over \mathbb{C} .
- If A and B have the same eigenvalues, are they conjugate?
 - If A and B have the same minimal polynomial, are they conjugate?
 - If A and B are diagonalisable, then are $A + B$ and AB ? What if A and B commute?
- Give proofs or counterexamples.
7. Let A be an $n \times n$ complex diagonalisable matrix. If all eigenvalues of A are real, find a positive integer m such that $A + mI$ is diagonalisable with all diagonal entries positive.

8. Although this question has been answered in the chapter, it is a good idea to do it now to ensure the ideas have sunk in — especially as it sometimes occurs as a Tripos question!

Let α and β be endomorphisms of the finite dimensional complex vector space V which commute, so $\alpha \circ \beta = \beta \circ \alpha$.

- (a) Show β maps each eigenspace $\mathcal{E}(\lambda) = \ker(\alpha - \lambda I)$ for α into itself.
 (b) Suppose that α and β are both diagonalizable. Show that there is a basis for V so that both α and β are represented by diagonal matrices. (Consider β restricted to $\mathcal{E}(\lambda)$.)

9. Let $p(X) = X^3 - 5X^2 + 6X$. For each polynomial f , let $R(f)$ be the remainder when $X \cdot f(X)$ is divided by p . Show that R is a linear map from the vector space of all polynomials into the vector subspace \mathcal{P}_2 of polynomials with degree at most 2. Consider the endomorphism ρ of \mathcal{P}_2 defined by $\rho : f \mapsto R(f)$. Find the eigenvalues and eigenvectors of ρ . Find the matrix of ρ relative to some basis for \mathcal{P}_2 . Find the minimal polynomial for ρ .

Repeat the above exercise when p is the polynomial $p(X) = X^3 - 2X^2 + X$.

-10. Prove that, if τ is represented by an upper triangular $n \times n$ matrix T , then each of the diagonal entries $t_{i,i}$ is an eigenvalue of τ and every eigenvalue arises in this way.

11. Let T be an upper triangular $n \times n$ with $t_{i,i} \neq t_{j,j}$ whenever $i \neq j$. Prove that T is diagonalisable. What happens if $t_{i,i} = t_{j,j}$ for some $i \neq j$? Try to give a necessary and sufficient condition for a non-diagonal upper-triangular matrix to be diagonalisable.

12. Show that the trace of the conjugate of a matrix is the same as the trace of the original matrix. Hence define the trace of an endomorphism of a finite dimensional vector space.

Show that there are no endomorphisms α, β of a finite dimensional vector space V with $\alpha \circ \beta - \beta \circ \alpha = I$.

Find endomorphisms of an infinite dimensional vector space V which do satisfy $\alpha \circ \beta - \beta \circ \alpha = I$.

13. Let $\mathcal{E}nd(V)$ denote the vector space of all endomorphisms of the finite dimensional complex vector space V . Then $\mathcal{E}nd(V)$ is finite dimensional. Show that, for $\alpha \in \mathcal{E}nd(V)$, the map $\Phi : \tau \mapsto \alpha \circ \tau$ is an endomorphism of $\mathcal{E}nd(V)$. If λ is an eigenvalue of α with geometric multiplicity k , show that λ is also an eigenvalue of Φ and find its geometric multiplicity. Is every eigenvalue of Φ an eigenvalue of α ?

Tripes questions on the material in this chapter: 98105, 98307, 99114(a), 99215, 99317, 00105, 00114, 00416.

Chapter 3

Jordan Normal Form

3.1 Nilpotent Matrices

Example 3.1.1 Let V be the vector space of all polynomials with complex coefficients of degree less than n ; so $\{1, X, X^2, \dots, X^{n-1}\}$ is a basis for V . Let $D \in \mathcal{L}(V, V)$ be differentiation; i.e., $D(X^j) = jX^{j-1}$ ($j = 0, 1, \dots, n-1$). Since $D(p(X))$ has smaller degree than $p(X)$ and is 0 only if $\deg(p) < 1$, it follows that D has no non-zero eigenvalues and no non-constant eigenvectors. Since $D(c) = 0 = 0c$ for any $c \in \mathbb{C}$, the only eigenvalue is 0 and $\mathcal{E}_D(0) = \langle 1 \rangle$. With respect to the above basis, D is represented by the matrix whose (i, j) entry is $i\delta_{i,j-1}$; i.e., an upper-triangular matrix all of whose entries are 0 except right above the diagonal where the entries are $1, 2, \dots, n-1$:

$$D = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 2 & 0 & \dots & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & n-1 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

Note that $D^n(p(X)) = 0$ for all polynomials p of degree strictly less than n . Thus the minimal polynomial for D divides X^n and so is X^k for some $k \leq n$. Since $D^j(X^j) = j!$, we get $m_D(X) = X^n$.

If we use the basis $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ where $\mathbf{b}_{j+1} = ((n-1)!/j!)X^j$ ($j = 0, 1, \dots, n-1$), then $D(\mathbf{b}_{j+1}) = \mathbf{b}_j$ ($j = 1, \dots, n-1$), $D(\mathbf{b}_1) = 0$ and D is

represented by the $n \times n$ matrix

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

This is the Jordan form for D .

More generally:

Definition 3.1.2 An $n \times n$ matrix N is said to be nilpotent if $N^k = 0$ for some positive integer k .

An endomorphism τ of a finite dimensional vector space V is said to be nilpotent if $\tau^k = 0$ for some positive integer k .

Note that $(P^{-1}MP)^k = P^{-1}M^kP$; so M is nilpotent iff its conjugates are nilpotent. Thus nilpotency of a representing matrix for a linear transformation is independent of the basis chosen.

Now suppose that τ is a nilpotent endomorphism of an n -dimensional vector space V and that m is minimal such that $\tau^m = 0$. Then $(\exists \mathbf{b}_m \in V)(\tau^{m-1}(\mathbf{b}_m) \neq \mathbf{0})$. Let $\mathbf{b}_{m-j} = \tau^j(\mathbf{b}_m)$ ($j = 1, \dots, m-1$). Then $\tau(\mathbf{b}_{j+1}) = \mathbf{b}_j$ for $j = 1, \dots, m-1$. Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ and $W = \langle \mathcal{B} \rangle$. Then $\tau(W) \subseteq W$ and the minimal polynomial for τ_W is also X^m .

Further \mathcal{B} is linearly independent: if $\sum_{j=1}^m \lambda_j \mathbf{b}_j = \mathbf{0}$, then applying τ^{m-1} to each side gives $\lambda_m \mathbf{b}_1 = \mathbf{0}$. Hence $\lambda_m = 0$. Now apply τ^{m-2} to each side and get $\lambda_{m-1} = 0$; etc.. Thus \mathcal{B} is a basis of W and with respect to this basis, τ_W is represented by the $m \times m$ matrix

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

Definition 3.1.3 Assume the above notation. If $W = V$, then \mathbf{b}_m is called a cyclic vector for τ .

Note that a nilpotent endomorphism τ of an n -dimensional vector space V has a cyclic vector iff $m_\tau(X) = X^n$.

The above shows that Example 3.1.1 generalises:

Proposition 3.1.4 *Let V be an n -dimensional vector space and τ be a nilpotent endomorphism of V with minimal polynomial X^n . Then there is a basis for V with respect to which τ is represented by the $n \times n$ matrix*

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

Definition 3.1.5 *Let τ be a nilpotent endomorphism. The least positive integer m such that $\tau^m = 0$ is called the order of τ .*

Proposition 3.1.6 *Let V be a vector space of dimension n and $\tau \in \mathcal{L}(V, V)$ be nilpotent of order m . Then there are subspaces W_1, \dots, W_m of V such that*

- (i) $V = W_1 \oplus \dots \oplus W_m$,
- (ii) $\tau(W_j) \subseteq W_{j-1}$ for $j = 2, \dots, m$,
- (iii) $\tau(W_1) = \{\mathbf{0}\}$, and
- (iv) $\tau|_{W_j}$ is injective for each $j = 2, \dots, m$.

[In Example 3.1.1, take $W_j = \langle \mathbf{b}_j \rangle$.]

Proof: Let U_j be the subspace $\ker(\tau^j)$ of V ($j = 0, 1, \dots, m$). So $\{\mathbf{0}\} = U_0 \subseteq U_1 \subseteq \dots \subseteq U_m = V$. Define W_j inductively so that

- (a) $U_j = U_{j-1} \oplus W_j$ and
- (b) $\tau(W_{j+1}) \subseteq W_j$.

First let W_m be any complement of U_{m-1} in $U_m = V$. Then (a) & (b) hold. Assume that W_m, \dots, W_{k+1} have been chosen to satisfy (a) and (b). Note that $\tau(W_{k+1}) \subseteq \tau(U_{k+1}) = U_k$ and $U_{k-1} \subseteq U_k$. If $\mathbf{v} \in U_{k-1} \cap \tau(W_{k+1})$, then $\mathbf{v} = \tau(\mathbf{w})$ for some $\mathbf{w} \in W_{k+1}$. Hence $\tau^k(\mathbf{w}) = \tau^{k-1}(\mathbf{v}) = \mathbf{0}$, whence $\mathbf{w} \in U_k \cap W_{k+1} = \{\mathbf{0}\}$ by (a) (induction hypothesis). Thus $\tau(W_{k+1}) \oplus U_{k-1} = \langle \tau(W_{k+1}), U_{k-1} \rangle \subseteq U_k$, and we can take Z_k to be a complementary subspace thereof in U_k . Let $W_k = \tau(W_{k+1}) \oplus Z_k$. Then (a) and (b) hold. Proceed by induction. Since $W_k \cap U_{k-1} = \{\mathbf{0}\} = U_0$ and

$U_{k-1} = \ker(\tau^{k-1})$, we get τ_{W_k} is injective and $V = U_m = U_{m-1} \oplus W_m = U_{m-2} \oplus W_{m-1} \oplus W_m = \dots = W_1 \oplus \dots \oplus W_m$. //

Corollary 3.1.7 *Let τ be a nilpotent endomorphism of order k of a finite dimensional vector space V . Then τ can be represented by a matrix of the form*

$$\begin{pmatrix} J^{(m_1)} & 0 & \dots & \dots & 0 \\ 0 & J^{(m_2)} & 0 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & \dots & 0 & J^{(m_\ell)} \end{pmatrix}$$

where $J^{(m_i)}$ is an $m_i \times m_i$ matrix of the form

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

Proof: For visual reasons, write $W^{(j)}$ for W_j in the previous proposition. Choose a basis $\mathbf{e}_1^{(k)}, \dots, \mathbf{e}_{N(k)}^{(k)}$ for $W^{(k)}$. Since τ maps $W^{(k)}$ injectively into $W^{(k-1)}$, the vectors $\tau(\mathbf{e}_1^{(k)}), \dots, \tau(\mathbf{e}_{N(k)}^{(k)})$ are linearly independent in $W^{(k-1)}$; label these images $\mathbf{e}_1^{(k-1)}, \dots, \mathbf{e}_{N(k)}^{(k-1)}$, and extend to a basis $\dots, \mathbf{e}_{N(k-1)}^{(k-1)}$ of $W^{(k-1)}$. Continuing gives

$$\begin{array}{ll} \mathbf{e}_1^{(k)}, \dots, \mathbf{e}_{N(k)}^{(k)} & W^{(k)} \\ \mathbf{e}_1^{(k-1)}, \dots, \mathbf{e}_{N(k)}^{(k-1)}, \dots, \mathbf{e}_{N(k-1)}^{(k-1)} & W^{(k-1)} \\ \vdots & \\ \mathbf{e}_1^{(1)}, \dots, \mathbf{e}_{N(k)}^{(1)}, \dots, \mathbf{e}_{N(k-1)}^{(1)}, \dots, \mathbf{e}_{N(1)}^{(1)} & W^{(1)} \\ \mathbf{0} & \end{array}$$

with τ mapping each line to the next.

The vectors so described form a basis for V . With respect to $\mathbf{e}_1^{(1)}, \dots, \mathbf{e}_1^{(k)}, \mathbf{e}_2^{(1)}, \dots, \mathbf{e}_2^{(k)}, \dots$ going up each column (as far as possible, in turn), we have τ represented by $J^{(k)}$ copied $N(k)$ times, then $J^{(k-1)}$ copied $N(k-1) - N(k)$ times, ..., $J^{(1)}$ copied $N(1) - N(2)$ times. //

3.2 Jordan Canonical Form

Definition 3.2.1 A Jordan block is either $J^{(1)}(\lambda) = (\lambda)$ or $J^{(k)}(\lambda) = \lambda I_k + J^{(k)}(0)$ where

$$J^{(k)}(0) = J^{(k)} = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

a $k \times k$ matrix.

A matrix M is in Jordan canonical form if it has the form $J(\lambda_1) \oplus \dots \oplus J(\lambda_m)$ for some $\lambda_1, \dots, \lambda_m$ where each $J(\lambda_i) = J^{(k_{i,1})}(\lambda_i) \oplus \dots \oplus J^{(k_{i,t_i})}(\lambda_i)$ and $k_{i,1} \geq \dots \geq k_{i,t_i}$.

Theorem 3.A (JORDAN CANONICAL FORM) Let V be a finite dimensional vector space over \mathbb{C} and τ an endomorphism of V . Let $m_\tau(X) = \prod_{j=1}^{\ell} (X - \mu_j)^{d_j}$, with μ_1, \dots, μ_ℓ all distinct. Let $V(\mu_j) = \ker(\tau - \mu_j i_V)^{d_j}$ ($j = 1, \dots, \ell$). Then $V = V(\mu_1) \oplus \dots \oplus V(\mu_\ell)$ and a basis of V can be chosen so that τ_j , the restriction of τ to $V(\mu_j)$, is represented by $J(\mu_j) = J^{(k_{j,1})}(\mu_j) \oplus \dots \oplus J^{(k_{j,t_j})}(\mu_j)$ and $k_{j,1} \geq \dots \geq k_{j,t_j}$ ($j = 1, \dots, \ell$). Moreover, t_j is the geometric multiplicity of μ_j . The resulting direct sum of all these $J(\mu_1) \oplus \dots \oplus J(\mu_\ell)$ gives the Jordan canonical form for τ . So, given any $n \times n$ matrix T over \mathbb{C} , there is an invertible matrix P (whose columns are the basis elements) such that $P^{-1}TP$ is in Jordan canonical form.

Proof: By Corollary 2.2.12, $\tau = \tau_1 \oplus \dots \oplus \tau_\ell$ and each τ_j has a single eigenvalue μ_j (μ_1, \dots, μ_ℓ distinct). Then τ_j corresponds to $\mu_j I + N_j$ where N_j is nilpotent. By Corollary 3.1.7, N_j has form the direct sum of $J^{(k_{j,i})}(0)$, whence each μ_j has the prescribed form. The geometric multiplicity result is in the exercises. //

The Jordan Form is a powerful tool in the complex case: prove any result about Jordan blocks, then Jordan forms, and then apply it to the original linear transformation (with respect to the appropriate basis), or conjugate to get the corresponding result for the original matrix. Once one has established the result in this way, it is usually worthwhile to try

the problem again without employing such heavy equipment; one usually learns a lot more!

Of course, the Jordan method does not apply in the real case.

3.3 Differential Equations

Example 3.3.1 Let $C^n(\mathbb{R})$ denote the vector space over \mathbb{R} of all n times continuously differentiable functions from \mathbb{R} into \mathbb{C} . Then $D : C^n(\mathbb{R}) \rightarrow C^{n-1}(\mathbb{R})$ where $D(f) = f'$.

Consider the differential equation

$$f^{(n)}(t) + a_{n-1}f^{(n-1)}(t) + \dots + a_1f'(t) + a_0f(t) = 0 \quad (*)$$

where $a_0, \dots, a_{n-1} \in \mathbb{C}$.

Then $e^{\lambda t}$ is a solution where λ is a root of $p(X) = \sum_{j=0}^n a_j X^j$ (where $a_n = 1$). Now $p(D) : C^n(\mathbb{R}) \rightarrow C(\mathbb{R})$ and $p(D)(f) = \sum_{j=0}^n a_j f^{(j)}(t)$. So if $V = \ker(p(D)) = \{f \in C^n(\mathbb{R}) : f \text{ satisfies } (*)\}$, then V is a subspace of $C^n(\mathbb{R})$ and $\tau = D_V$ is an endomorphism of V .

For all $y = (y_0, \dots, y_{n-1}) \in \mathbb{C}^n$, there is a unique $f \in V$ with $f^{(j)}(0) = y_j$ ($j = 0, \dots, n-1$). Let this solution be $\sigma(y)$. Then $\sigma : \mathbb{C}^n \rightarrow V$ is an injective linear map. Its inverse is $f \mapsto (f(0), \dots, f^{(n-1)}(0))$. So $\mathbb{C}^n \cong V$ and $\dim(V) = n$. If $\{\mathbf{e}_0, \dots, \mathbf{e}_{n-1}\}$ is the standard basis for \mathbb{C}^n and $g_k = \sigma(\mathbf{e}_k) \in V$ ($k = 0, \dots, n-1$), then $g_k^{(j)}(0) = \delta_{j,k}$ (the Krönercker delta). By (*), $g_k^{(n)}(0) = -\sum_{j=0}^{n-1} a_j g_k^{(j)}(0) = -a_k$, whence $D(g_k) = g_{k-1} - a_k g_{n-1}$ and $D(g_0) = -a_0 g_{n-1}$. With respect to this basis, τ is represented by the matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ -a_0 & -a_1 & \cdot & \cdot & \dots & -a_{n-1} \end{pmatrix}$$

This is known as the rational canonical form for τ . It is the matrix for $\sigma^{-1}\tau\sigma \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^n)$.

We next show that $p(X)$ is the minimal polynomial for τ . Since $f \in V$ implies $p(D)(f) = 0$, we have $m_\tau(X) | p(X)$. If $q(X) \neq 0$ and $q(\tau) = 0$

with $\deg(q) = m < n$, then $f^{(m)}(0) + q_{m-1}f^{(m-1)}(0) + \dots + q_0f(0) = 0$ for all $f \in V$. This is obviously false (putting $f = g_m$ gives $1 + 0 + 0 + \dots + 0 = 0$). Hence $m_\tau(X) = p(X)$.

Now $0 = (D - \lambda i_V)f = f' - \lambda f$ if $f \in V$ is an eigenvector of τ with eigenvalue λ . Thus $f(t) = ce^{\lambda t}$ for some $c \in \mathbb{C}$. So $f_\lambda(t) = e^{\lambda t}$ is a basis for $\mathcal{E}_\tau(\lambda)$, and λ has geometric multiplicity 1. If $p(X)$ had a repeated root λ , then $f_{\lambda,1} = te^{\lambda t} \in V$ and $(\tau - \lambda i_V)f_{\lambda,1} = f_\lambda$. Indeed, $\{f_{\lambda,1}, f_\lambda\}$ spans the subspace of all functions $h(t) \in V$ satisfying $(D - \lambda i_V)^2 h = 0$. More generally, if λ is a root of $m_\tau(X)$ of multiplicity k , then $\{t^m e^{\lambda t} : m = 0, \dots, k-1\}$ spans the vector subspace of V of all h such that $(D - \lambda i_V)^k h = 0$. By the Jordan Normal Form Theorem 3.A, we have that $V = V(\mu_1) \oplus \dots \oplus V(\mu_\ell)$ where μ_1, \dots, μ_ℓ are all distinct, $p(X) = (X - \mu_1)^{d_1} \dots (X - \mu_\ell)^{d_\ell}$ and $V(\mu_j) = \ker(\tau - \mu_j i_V)^{d_j}$ ($j = 1, \dots, \ell$). But $e^{\mu_j t} \in V(\mu_j)$ and if $q(t)e^{\mu_j t} \in V(\mu_j)$, then $e^{\mu_j t} q^{(d_j)}(t) = 0$ since $(D - \mu_j i_V)q(t)e^{\mu_j t} = [q'(t) + q(t)\mu_j - \mu_j q(t)]e^{\mu_j t} = q'(t)e^{\mu_j t}$. Hence $q(t)$ is a polynomial of degree strictly less than d_j . If $f_{k,j} : t \mapsto \frac{t^k}{k!} e^{\mu_j t}$ ($k = 0, \dots, d_j - 1$) then $\{f_{k,j} : k = 0, \dots, d_j - 1\}$ is a basis of $V(\mu_j)$ with $f_{0,j}$ an eigenvector. Since $\tau(V(\mu_j)) \subseteq V(\mu_j)$, the matrix for τ_{V_j} with respect to this basis is

$$J^{(d_j)}(\mu_j) = \begin{pmatrix} \mu_j & 1 & 0 & 0 & \dots & 0 \\ 0 & \mu_j & 1 & 0 & \dots & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \mu_j & 1 \\ 0 & 0 & 0 & 0 & \dots & \mu_j \end{pmatrix}$$

Then $J^{(d_1)}(\mu_1) \oplus \dots \oplus J^{(d_\ell)}(\mu_\ell)$ is the Jordan form for $\tau = D_V$.

3.4 Exercises

-1. Show that none of the following matrices are conjugate:

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

Is the matrix

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

conjugate to any of them? If so, which? [No computations are necessary.]

2. Show that any $n \times n$ matrix over \mathbb{C} with all its eigenvalues real is conjugate to an $n \times n$ matrix over \mathbb{R} .

3. Let V be a complex vector space with dimension n and let τ be an endomorphism of V with $\tau^{n-1} \neq 0$ but $\tau^n = 0$. Show that there is a vector $\mathbf{x} \in V$ for which

$$\mathbf{x}, \tau(\mathbf{x}), \tau^2(\mathbf{x}), \dots, \tau^{n-1}(\mathbf{x})$$

is a basis for V . What is the matrix of τ relative to this basis?

Show that if another endomorphism α of V commutes with τ then $\alpha = p(\tau)$ for some polynomial p . (Consider $\alpha(\mathbf{x})$.)

4. Show that the dimension of the eigenspace for λ is the number of Jordan blocks of the form $\lambda I + J$ in a matrix of Jordan canonical form.

5. Show that two endomorphisms σ, τ of a finite dimensional complex vector space V are conjugate if, and only if, they are represented by the same Jordan canonical forms.

6. Let $\tau \in \mathcal{L}(V, V)$ be an endomorphism of a finite dimensional complex vector space V . Show that, if λ is an eigenvalue of τ then λ^2 is an eigenvalue of τ^2 . Show further that every eigenvalue of τ^2 arises in this way. Give an example to show that the dimensions of the eigenspaces $\ker(\tau - \lambda I)$ and $\ker(\tau^2 - \lambda^2 I)$ may differ.

7. Let T be a $K \times K$ Jordan block matrix with eigenvalue λ . (So $T = \lambda I + J^{(K)}$.) Find the Jordan canonical form representing T^2 .

8. Consider the linear difference equation

$$x_{k+N} + a_{N-1}x_{k+N-1} + \dots + a_1x_{k+1} + a_0x_k = 0 \quad \text{for } k = 1, 2, 3, \dots \quad (*)$$

(a) Show that the set V of all solutions to (*) form a vector subspace of the space S of all complex sequences $\mathbf{x} = (x_k)_{k=1}^{\infty}$. Show further that the map

$$V \rightarrow \mathbb{C}^N ; \mathbf{x} \mapsto (x_1, x_2, \dots, x_N)$$

is an isomorphism.

(b) Show that, if $\mathbf{x} = (x_1, x_2, x_3, \dots)$ is in V , then the shifted sequence $\sigma(\mathbf{x}) = (x_2, x_3, x_4, \dots)$ is also in V and the map σ is an endomorphism of V . Prove that the minimal polynomial for σ is $p(X) = X^N + a_{N-1}X^{N-1} + \dots + a_1X + a_0$.

(c) Suppose that $p(X) = (X - \lambda)^m$. Show that the sequences $\mathbf{y}(r)$ with

$$y(r)_k = \binom{k}{r} \lambda^{k-r}$$

for $r = 0, 1, \dots, m-1$ form a basis for V . Find the matrix of σ relative to this basis.

(d) Suppose that $p(X) = (X - \lambda)^m q(X)$ with $q(\lambda) \neq 0$. Show that

$$V(\lambda) = \{\mathbf{x} \in S : (\sigma - \lambda I)^m \mathbf{x} = \mathbf{0}\} \quad \text{and} \quad W = \{\mathbf{x} \in S : q(\sigma)\mathbf{x} = \mathbf{0}\}$$

are vector subspaces of V with dimensions m and $N - m$ respectively. Show that $V = V(\lambda) \oplus W$.

Deduce that V is the direct sum of the subspaces $V(\lambda)$ for λ a zero of p and that σ maps each of these subspaces into itself.

Tripos questions: 98215, 98307, 98317, 99114, 99307, 00105, 00114, 00415.

Chapter 4

Determinants

4.1 The Desired Properties

In the Algebra & Geometry course we met the definition of the determinant of a small square matrix. We wish to extend the definition so that the function $\det : \mathcal{M}_{n \times n}(F) \rightarrow F$ satisfies:

(1) $\det(P^{-1}MP) = \det(M)$;

(2) $\det(M) \neq 0 \iff M^{-1} \text{ exists} \iff (M\mathbf{x} = \mathbf{0} \rightarrow \mathbf{x} = \mathbf{0})$;

(3)
$$\det \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \det(A)\det(C)$$
;

(4) $\det(I_n) = 1$;

(5) If N is the result of multiplication of any row or column of M by μ , then $\det(N) = \mu \det(M)$;

(6) $\det(XI_n - M)$ is a monic polynomial in X of degree n ;

(7) If F_1 is a field containing F , then the F -determinant function is the restriction of the F_1 -determinant function.

We will obtain all our results from these key 7 properties and then show that there is a function that satisfies them all. Although this is all that is necessary, we will explicitly construct such a function and shew that it is unique.

4.2 The Characteristic Polynomial

For any endomorphism τ of a finite dimensional vector space, we may define $\det(\tau) = \det(T)$ where T is a matrix representation of τ with respect to some basis. By (1), $\det(\tau)$ is independent of the basis chosen.

Definition 4.2.1 Let τ be an endomorphism of a vector space V with $\dim(V) = n$. The characteristic polynomial of τ is $\det(Xi_V - \tau) = \det(XI_n - T)$ where T represents τ .

We write $\chi_\tau(X)$ for the characteristic polynomial for τ .

Proposition 4.2.2 Let V be a finite dimensional vector space and τ be an endomorphism of V . The roots of the characteristic polynomial for τ are just the eigenvalues of τ .

Proof: By (2)

$$\det(\lambda i_V - \tau) = 0 \iff (\exists \mathbf{v} \neq \mathbf{0})(\lambda i_V - \tau)\mathbf{v} = \mathbf{0} \iff (\exists \mathbf{v} \neq \mathbf{0})(\lambda \mathbf{v} = \tau(\mathbf{v})) \iff \lambda \text{ is an eigenvalue of } \tau. //$$

Definition 4.2.3 The multiplicity of λ as a root of the characteristic polynomial for τ is called the algebraic multiplicity of the eigenvalue λ of τ .

So the algebraic multiplicity of λ is k iff $\chi_\tau(X) = (X - \lambda)^k p(X)$ where $p(\lambda) \neq 0$.

Proposition 4.2.4 Let V be a finite dimensional vector space and τ be an endomorphism of V . Then for any eigenvalue λ of τ ,

$$\text{geom.mult.}(\lambda) \leq \text{alg.mult.}(\lambda).$$

Proof: If d is the geometric multiplicity of λ , then $\dim(\mathcal{E}_\tau(\lambda)) = d$. Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_d\}$ be a basis of $\mathcal{E}(\lambda)$ and extend it to a basis $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ of V . Let τ be represented by the matrix T with respect to this basis; so

$$T = \begin{pmatrix} I_d & B \\ 0 & C \end{pmatrix}$$

By (3), (4) and (5) we have $\det(Xi_V - \tau) = \det((X - \lambda)I_d) \det(XI_{n-d} - C) = (X - \lambda)^d \det(XI_{n-d} - C)$. Hence $\chi_\tau(X) = (X - \lambda)^d q(X)$ for some polynomial $q(X)$ of degree $n - d$ (which may or may not have λ as a root). //

Corollary 4.2.5 *An endomorphism τ of a finite dimensional vector space is diagonalisable iff for each eigenvalue the algebraic multiplicity and geometric multiplicity are equal.*

Proof: Since the sum of the algebraic multiplicities of eigenvalues is at most $n = \dim(V) = \deg(\chi_\tau(X))$ and τ is diagonalisable iff the sum of the geometric multiplicities is n , the result follows at once from Proposition 4.2.4. //

As is standard, the trace of an $n \times n$ matrix A is defined as the sum of the diagonal entries: $\text{tr}(A) = \sum_{i=1}^n a_{i,i}$. Then $\text{tr}(P^{-1}AP) = \sum_{i,j,k} \hat{p}_{i,j} a_{j,k} p_{k,i} = \sum_{j,k} a_{j,k} \sum_i p_{k,i} \hat{p}_{i,j} = \sum_{j,k} a_{j,k} \delta_{k,j} = \sum_j a_{j,j} = \text{tr}(A)$, where $\hat{p}_{i,j}$ is the (i, j) entry of the matrix P^{-1} . Thus we can define

Definition 4.2.6 *The trace of an endomorphism τ of an n -dimensional vector space is the trace of some/any matrix T representing τ with respect to some basis.*

Over \mathbb{C} , every polynomial is a product of linear factors whence every linear transformation of a finite dimensional vector space over \mathbb{C} has an upper triangular representation by Proposition 2.2.16. Moreover, by (3), (4) and (5), if T is upper triangular, then

$$\det(T) = \det \begin{pmatrix} t_{1,1} & B \\ 0 & T_{1,1} \end{pmatrix} = t_{1,1} \det(T_{1,1}) = \dots = \prod_{i=1}^n t_{i,i}$$

since $T_{1,1}, \dots$ are upper triangular. Now T is upper triangular iff $XI_n - T$ is upper triangular. Thus for any endomorphism τ of an n -dimensional vector space over \mathbb{C} , $\chi_\tau(X) = \prod_{i=1}^n (X - t_{i,i})$ where $t_{1,1}, \dots, t_{n,n}$ are the eigenvalues of τ . Note that

- (a) the algebraic multiplicity of λ is just $|\{j \in \{1, \dots, n\} : t_{j,j} = \lambda\}|$,
- (b) $\text{tr}(\tau)$ is the sum of the eigenvalues in \mathbb{C} of τ (counted to algebraic multiplicity), and
- (c) $\chi_\tau(X) = X^n - \text{tr}(\tau)X^{n-1} + \dots$

Theorem 4.A (The Cayley-Hamilton Theorem). *Let τ be an endomorphism of a finite dimensional vector space over a subfield F of \mathbb{C} . Then $\chi_\tau(\tau) = 0$.*

Proof: We first consider the special case that the vector space is over the field \mathbb{C} .

Choose a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ with respect to which τ is represented by an upper triangular matrix T (use Proposition 2.2.15). Let $U_j = \langle \mathbf{b}_1, \dots, \mathbf{b}_j \rangle$ ($j = 1, \dots, n$). Then $\tau(U_j) \subseteq U_j$ and $\tau(\mathbf{b}_j) - t_{j,j}\mathbf{b}_j \in U_{j-1}$ for $j = 1, \dots, n$. Thus $(\tau - t_{j,j}i_V)(U_j) \subseteq U_{j-1}$ for $j = 1, \dots, n$. Now

$$\begin{aligned} \chi_\tau(\tau)(V) &= \chi_\tau(\tau)(U_n) = \left(\prod_{j=1}^{n-1} (\tau - t_{j,j}i_V) \right) (\tau - t_{n,n}i_V)(U_n) \subseteq \\ &\prod_{j=1}^{n-1} (\tau - t_{j,j}i_V)(U_{n-1}) = \left(\prod_{j=1}^{n-2} (\tau - t_{j,j}i_V) \right) (\tau - t_{(n-1),(n-1)}i_V)(U_{n-1}) \subseteq \\ &\dots \subseteq (\tau - t_{1,1}i_V)(U_1) \subseteq U_0 = \{\mathbf{0}\}. \end{aligned}$$

Hence $\chi_\tau(\tau)(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in V$.

For the general case, let \mathcal{B} be a basis of V and τ be represented by the matrix T with respect to this basis. Note that the entries of T belong to $F \subseteq \mathbb{C}$. Write $T_{\mathbb{C}}$ for the same matrix T viewed as a matrix over \mathbb{C} . We have $\chi_{T_{\mathbb{C}}}(T) = \chi_{T_{\mathbb{C}}}(T_{\mathbb{C}}) = 0$ by the first part. By (7), $\chi_T(X) = \chi_{T_{\mathbb{C}}}(X)$. Hence $\chi_\tau(\tau) = \chi_T(T) = \chi_{T_{\mathbb{C}}}(T) = 0$. //

The same theorem holds for arbitrary fields F ; we use the same proof with \overline{F} , the algebraic closure of F , in place of \mathbb{C} .

Corollary 4.2.7 *For any endomorphism τ of a finite dimensional vector space, $m_\tau(X) | \chi_\tau(X)$. Moreover, $m_\tau(X)$ and $\chi_\tau(X)$ have the same roots (though possibly occurring to different multiplicities).*

Proof: Since $\chi_\tau(\tau) = 0$, $\chi_\tau(X)$ is divisible by the minimal polynomial. The roots of the minimal polynomial are precisely the eigenvalues of τ (Proposition 2.2.11) so the result follows from Proposition 4.2.2. //

4.3 Volumes

Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^n$. We wish to define $\mathcal{V}(\mathbf{x}_1, \dots, \mathbf{x}_n)$, the volume of an n -dimensional parallelepiped $P(\mathbf{x}_1, \dots, \mathbf{x}_n) = \{\sum_{j=1}^n \lambda_j \mathbf{x}_j : 0 \leq \lambda_j \leq 1\}$. Fix $\mathbf{x}_2, \dots, \mathbf{x}_n$ and let \mathbf{x}_1 vary; say $\mathcal{V}_1(\mathbf{v}) = \mathcal{V}(\mathbf{v}, \mathbf{x}_2, \dots, \mathbf{x}_n)$. Let B be

the base of the parallelepiped $P(\mathbf{v}, \mathbf{x}_2, \dots, \mathbf{x}_n)$ and \mathbf{u} be a unit vector perpendicular to B . Then we should have that $\mathcal{V}_1(\mathbf{v})$ should equal the product of $\mathcal{V}(\mathbf{x}_2, \dots, \mathbf{x}_n)$ and the component of \mathbf{v} in the direction \mathbf{u} . Note that this gives a negative value if \mathbf{v} is in the opposite direction to \mathbf{u} .

Similarly, each of the maps $i = 1, \dots, n$

$$\mathbf{v} \mapsto \mathcal{V}_i(\mathbf{v}) = \mathcal{V}(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{v}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_n) \quad (\dagger)$$

should be linear.

Definition 4.3.1 *We say that \mathcal{V} is n -linear if each of the n maps $\mathbf{v} \mapsto \mathcal{V}(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{v}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_n)$ is linear ($i = 1, \dots, n$).*

If $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ is linearly dependent, then the set lies in an $(n - 1)$ -dimensional subspace of \mathbb{R}^n ; so we will want $\mathcal{V}(\mathbf{x}_1, \dots, \mathbf{x}_n) = 0$. In particular, if $\mathbf{x}_i = \mathbf{x}_j$ for some distinct i, j we want $\mathcal{V}(\mathbf{x}_1, \dots, \mathbf{x}_n) = 0$. Hence $0 = \mathcal{V}(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n) = \mathcal{V}(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0 + 0 + \mathcal{V}(\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_3, \dots, \mathbf{x}_n)$ by n -linearity, whence $\mathcal{V}(\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_3, \dots, \mathbf{x}_n) = -\mathcal{V}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n)$. Thus interchanging the first two variables multiplies the volume by -1 . Similarly any transposition of two variables introduces a factor of -1 .

Now for any $\sigma \in \text{Sym}(n)$, σ is a product of transpositions. Moreover, it is a product of an even number of transpositions iff every expression for σ as a product of transpositions involves an even number of such. We define $sg(\sigma) = 1$ if σ can be written as a product of an even number of transpositions and $sg(\sigma) = -1$ otherwise. Note that $sg(\sigma^{-1}) = sg(\sigma)$ and $sg(\rho^{-1}\sigma\rho) = sg(\sigma)$.

By the above we have

$$\mathcal{V}(\mathbf{x}_{\sigma(1)}, \mathbf{x}_{\sigma(2)}, \dots, \mathbf{x}_{\sigma(n)}) = sg(\sigma)\mathcal{V}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \quad (\dagger\dagger).$$

Definition 4.3.2 *We say that $\mathcal{V} : \mathbb{R}^n \rightarrow \mathbb{R}$ is alternating if $\mathcal{V}(\mathbf{x}_{\sigma(1)}, \mathbf{x}_{\sigma(2)}, \dots, \mathbf{x}_{\sigma(n)}) = sg(\sigma)\mathcal{V}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ for all $\sigma \in \text{Sym}(n)$.*

We will show that apart from a scalar factor, (\dagger) and $(\dagger\dagger)$ completely determine the volume.

We first extend the previous two definitions.

Definition 4.3.3 Let U be an n -dimensional vector space over a field F and let $f : U^n \rightarrow F$. We say that f is n -linear if each of the maps ($i = 1, \dots, n$) $\mathbf{v} \mapsto f(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{v}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_n)$ is linear; and alternating if $f(\mathbf{x}_{\sigma(1)}, \mathbf{x}_{\sigma(2)}, \dots, \mathbf{x}_{\sigma(n)}) = \text{sg}(\sigma)f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ for all $\sigma \in \text{Sym}(n)$. If f is both n -linear and alternating then we will call it a volume form.

Recall that $\epsilon_{i,j,k} = 1$ if i, j, k is a cyclic permutation of $1, 2, 3$ (i.e., $(ijk) = (123)$ as cycles) and $\epsilon_{i,j,k} = -1$ if i, j, k is a cyclic permutation of $1, 3, 2$; $\epsilon_{i,j,k} = 0$ otherwise. The standard definition of a 3×3 determinant from Algebra & Geometry is $\det(A) = \sum_{i,j,k} \epsilon_{i,j,k} a_{i,1} a_{j,2} a_{k,3}$. For example the term $\epsilon_{2,1,3} a_{2,1} a_{1,2} a_{3,3}$ is just $\epsilon_{2,1,3} a_{\sigma(1),1} a_{\sigma(2),2} a_{\sigma(3),3}$ where $\sigma(1) = 2$, $\sigma(2) = 1$ and $\sigma(3) = 3$. So σ is the single transposition (12) and hence $\text{sg}(\sigma) = -1$. Hence $\epsilon_{2,1,3} a_{2,1} a_{1,2} a_{3,3} = \text{sg}(\sigma) a_{\sigma(1),1} a_{\sigma(2),2} a_{\sigma(3),3}$ for this σ . Therefore $\det(A) = \sum_{\sigma \in \text{Sym}(3)} \text{sg}(\sigma) a_{\sigma(1),1} a_{\sigma(2),2} a_{\sigma(3),3}$.

This leads to the following generalisation:

Proposition 4.3.4 Let $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for a vector space U over F . For each $\mathbf{x}_j \in U$, let $\mathbf{x}_j = \sum_{i=1}^n x_{i,j} \mathbf{b}_i$. Let $\Delta : U^n \rightarrow F$ be defined by $\Delta(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \sum_{\sigma \in \text{Sym}(n)} \text{sg}(\sigma) x_{\sigma(1),1} \dots x_{\sigma(n),n}$. Then Δ is a volume form with $\Delta(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n) = 1$.

Proof: Clearly Δ is n -linear. Fix $\rho \in \text{Sym}(n)$. Then $x_{\sigma(1),\rho(1)}, \dots, x_{\sigma(n),\rho(n)}$ is just a rearrangement of $x_{\theta(1),1}, \dots, x_{\theta(n),n}$ where $\theta = \sigma\rho^{-1}$. As σ runs through $\text{Sym}(n)$ so does θ . Hence $\Delta(\mathbf{x}_{\rho(1)}, \mathbf{x}_{\rho(2)}, \dots, \mathbf{x}_{\rho(n)}) = \sum_{\sigma \in \text{Sym}(n)} \text{sg}(\sigma) a_{\sigma(1),\rho(1)} \dots a_{\sigma(n),\rho(n)} = \sum_{\theta \in \text{Sym}(n)} \text{sg}(\theta\rho) a_{\theta(1),1} \dots a_{\theta(n),n} = \text{sg}(\rho) \Delta(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$. Thus Δ is an alternating form whence a volume form. By definition, $\Delta(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n) = \text{sg}(\text{id}) 1 \dots 1 = 1$. //

Proposition 4.3.5 If f is a volume form on V , then $f = \lambda\Delta$ for some $\lambda \in F$. That is, Δ is unique to within a scalar (the volume of a given parallelepiped).

Proof: $f(\mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_s x_{s(1),1} \dots x_{s(n),n} f(\mathbf{b}_{s(1)}, \dots, \mathbf{b}_{s(n)})$ where s in the sum ranges over all functions from $\{1, \dots, n\}$ into itself. Since f is alternating $f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n) = -f(\mathbf{x}_2, \mathbf{x}_1, \mathbf{x}_3, \dots, \mathbf{x}_n)$. Hence $f(\mathbf{x}_1, \dots, \mathbf{x}_n) = 0$ if $\mathbf{x}_1 = \mathbf{x}_2$. Similarly, if $\mathbf{x}_i = \mathbf{x}_j$ for some $i \neq j$. Hence we may restrict the range of s in the sum to the set of injective maps from $\{1, \dots, n\}$ into itself; i.e., the set of permutations of $\{1, \dots, n\}$. Further $f(\mathbf{b}_{s(1)}, \dots, \mathbf{b}_{s(n)}) = \text{sg}(s) f(\mathbf{b}_1, \dots, \mathbf{b}_n)$ so the result follows with $\lambda = f(\mathbf{b}_1, \dots, \mathbf{b}_n)$. //

Proposition 4.3.6 *Let $f \neq 0$ be a volume form for an n -dimensional vector space U . Then $f(\mathbf{c}_1, \dots, \mathbf{c}_n) \neq 0$ iff $\{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ is a basis of U .*

Proof: If $\{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ is a basis of U , then by Proposition 4.3.4 there is a volume form Δ with $\Delta(\mathbf{c}_1, \dots, \mathbf{c}_n) = 1$. By Proposition 4.3.5, $f = \lambda\Delta$ for some non-zero scalar λ . Hence $f(\mathbf{c}_1, \dots, \mathbf{c}_n) = \lambda \neq 0$.

Conversely, suppose that $f(\mathbf{c}_1, \dots, \mathbf{c}_n) \neq 0$. As before, the function f takes the value 0 if two coordinates are the same. Consequently, if $\sum_{j=1}^n \mu_j \mathbf{c}_j = \mathbf{0}$, then $0 = f(\mathbf{0}, \mathbf{c}_2, \dots, \mathbf{c}_n) = f(\sum_j \mu_j \mathbf{c}_j, \mathbf{c}_2, \dots, \mathbf{c}_n) = \sum_{j=1}^n \mu_j f(\mathbf{c}_j, \mathbf{c}_2, \dots, \mathbf{c}_n) = \mu_1 f(\mathbf{c}_1, \dots, \mathbf{c}_n)$. Thus $\mu_1 = 0$. Similarly we get $\mu_2, \dots, \mu_n = 0$; so $\{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ is linearly independent, whence a basis. //

Proposition 4.3.7 *Let U be an n -dimensional vector space over a field F and τ be an endomorphism of U . Then there is a constant $\delta \in F$ such that $f(\tau(\mathbf{x}_1), \dots, \tau(\mathbf{x}_n)) = \delta f(\mathbf{x}_1, \dots, \mathbf{x}_n)$ for all volume forms f on U and all $\mathbf{x}_1, \dots, \mathbf{x}_n \in U$.*

Proof: By Proposition 4.3.4, there is a volume form $\Delta \neq 0$ on U . Define $g : (\mathbf{x}_1, \dots, \mathbf{x}_n) \mapsto \Delta(\tau(\mathbf{x}_1), \dots, \tau(\mathbf{x}_n))$. Then g is n -linear and alternating. (Verify this.) By Proposition 4.3.5 there is $\delta \in F$ with $g = \delta\Delta$. Let f be a volume form on U ; so $f = \lambda\Delta$ by Proposition 4.3.5. Then $f(\tau(\mathbf{x}_1), \dots, \tau(\mathbf{x}_n)) = \lambda\Delta(\tau(\mathbf{x}_1), \dots, \tau(\mathbf{x}_n)) = \lambda g(\mathbf{x}_1, \dots, \mathbf{x}_n) = \lambda\delta\Delta(\mathbf{x}_1, \dots, \mathbf{x}_n) = \delta f(\mathbf{x}_1, \dots, \mathbf{x}_n)$. //

Definition 4.3.8 *Let $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for a vector space U over a field F . Let Δ be a volume form on U with $\Delta(\mathbf{b}_1, \dots, \mathbf{b}_n) = 1$. Define $\det(\tau) = \Delta(\tau(\mathbf{b}_1), \dots, \tau(\mathbf{b}_n)) = \delta$. By Proposition 4.3.7, this definition is independent of the basis chosen for U . We call Δ the determinant.*

4.4 Properties of the Determinant

First observe that $\det(i_U) = 1$ by the definition so the linear transformation version of Property (4) is immediate.

Proposition 4.4.1 *Let U be a finite dimensional vector space and α & β be endomorphisms of U . Then $\det(\alpha\beta) = \det(\alpha)\det(\beta)$.*

If α^{-1} exists, then $\det(\alpha) \neq 0$ and $\det(\alpha^{-1}) = 1/\det(\alpha)$.

Proof: Let $f \neq 0$ be a volume form on U . Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in U$ such that $f(\mathbf{x}_1, \dots, \mathbf{x}_n) \neq 0$. Then $\det(\alpha\beta)f(\mathbf{x}_1, \dots, \mathbf{x}_n) = f(\alpha\beta(\mathbf{x}_1), \dots, \alpha\beta(\mathbf{x}_n)) = \det(\alpha)f(\beta(\mathbf{x}_1), \dots, \beta(\mathbf{x}_n)) = \det(\alpha)\det(\beta)f(\mathbf{x}_1, \dots, \mathbf{x}_n)$. Hence $\det(\alpha\beta) = \det(\alpha)\det(\beta)$.

If α^{-1} exists, then as $\det(\alpha^{-1}\alpha) = \det(i_U) = 1$, the first part shows that $\det(\alpha^{-1})\det(\alpha) = 1$. //

Property (1) for linear transformations follows at once from the proposition.

We now give the converse to the last part.

Proposition 4.4.2 *If τ is an endomorphism of a finite dimensional vector space, the τ is invertible if $\det(\tau) \neq 0$.*

Proof: Let $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for a vector space U over F . By Proposition 4.3.4 there is a volume form Δ with $\Delta(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n) = 1$. Then $\Delta(\tau(\mathbf{b}_1), \dots, \tau(\mathbf{b}_n)) = \det(\tau)\Delta(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n) = \det(\tau) \neq 0$. Hence $\{\tau(\mathbf{b}_1), \dots, \tau(\mathbf{b}_n)\}$ is a basis of U by Proposition 4.3.6. Thus τ is surjective and injective. //

Let T be an $n \times n$ matrix over F and define $\tau : F^n \rightarrow F^n$ by $\tau(\mathbf{x}) = T\mathbf{x}$. Then $\Delta(\tau(\mathbf{e}_1), \dots, \tau(\mathbf{e}_n)) = \det(\tau)$. Now $\tau(\mathbf{e}_j) = T\mathbf{e}_j = \sum_i t_{i,j}\mathbf{e}_i$, so

Definition 4.4.3 $\det(\tau) = \sum_{\sigma \in \text{Sym}(n)} \text{sg}(\sigma)t_{\sigma(1),1} \dots t_{\sigma(n),n}$ which we define to be $\det(T)$.

Since $\tau_1 \circ \tau_2$ is represented by the matrix product T_1T_2 we can use Proposition 4.4.1 to get $\det(AB) = \det(A)\det(B)$, $\det(I_n) = 1$ and T^{-1} exists iff $\det(T) \neq 0$. Thus Properties (1) and (4) hold for matrices, too, and Property (5) is immediate from the definition.

If T is an upper triangular matrix (so $t_{i,j} = 0$ for $i > j$), then $t_{\sigma(1),1} \dots t_{\sigma(n),n} = 0$ if $\sigma(j) > j$ for some j . Hence $\det(T) = \sum_{\sigma \in \text{Sym}(n)} \text{sg}(\sigma)t_{\sigma(1),1} \dots t_{\sigma(n),n} = t_{1,1} \dots t_{n,n}$, the product of the diagonal elements of T .

More generally, if T is a matrix with $t_{i,j} = 0$ for all $i > r$ and $j \leq r$, then $t_{\sigma(1),1} \dots t_{\sigma(n),n} = 0$ if $\sigma(j) > r$ for some $j \leq r$. Hence

$$\det(T) = \sum_{\sigma \in \text{Sym}(n)} \text{sg}(\sigma)t_{\sigma(1),1} \dots t_{\sigma(n),n} = \sum_{\sigma \in \text{Sym}(r), \rho \in \text{Sym}(n-r)} \text{sg}(\sigma)\text{sg}(\rho)t_{\sigma(1),1} \dots t_{\sigma(r),r} t_{r+\rho(1),r+1} \dots t_{r+\rho(n-r),n} =$$

$\det(A)\det(C)$ where

$$T = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}.$$

This gives Property (3).

Since $\det(T_C) = \det(T)$ by definition, we have established that the determinant function defined satisfies all the required properties. We now show that its value for any endomorphism agrees with that of any matrix that represents it (with respect to any basis).

Proposition 4.4.4 *Let V be an n -dimensional vector space and τ be an endomorphism of V . If the matrix T represents τ with respect to some basis \mathcal{B} for V , then $\det(\tau) = \det(T)$.*

Proof: Let $\alpha : F^n \cong V$ and $\alpha(\mathbf{x}) = \sum_j x_j \mathbf{b}_j$. Then $\alpha(T\mathbf{x}) = \tau(\alpha(\mathbf{x}))$. For $\mathbf{y}_1, \dots, \mathbf{y}_n \in V$, define $f(\mathbf{y}_1, \dots, \mathbf{y}_n) = \Delta(\alpha^{-1}(\mathbf{y}_1), \dots, \alpha^{-1}(\mathbf{y}_n))$. Then f is a volume form and for $\mathbf{x}_1, \dots, \mathbf{x}_n \in F^n$, $\det(\tau)\Delta(\mathbf{x}_1, \dots, \mathbf{x}_n) = \det(\tau)f(\alpha(\mathbf{x}_1), \dots, \alpha(\mathbf{x}_n)) = f(\tau\alpha(\mathbf{x}_1), \dots, \tau\alpha(\mathbf{x}_n)) = f(\alpha(T\mathbf{x}_1), \dots, \alpha(T\mathbf{x}_n)) = \Delta(T\mathbf{x}_1, \dots, T\mathbf{x}_n) = \det(T)\Delta(\mathbf{x}_1, \dots, \mathbf{x}_n)$. Hence $\det(\tau) = \det(T)$. //

Definition 4.4.5 *If T is a matrix, let T^t be the matrix whose (i, j) entry is $t_{j,i}$. Then T^t is called the transpose matrix of T .*

Proposition 4.4.6 *If T is any $n \times n$ matrix, then $\det(T^t) = \det(T)$.*

Proof: If $\rho = \sigma^{-1}$, then $sg(\rho) = sg(\sigma)$, and $t_{\sigma(1),1} \dots t_{\sigma(n),n}$ & $t_{1,\rho(1)} \dots t_{n,\rho(n)}$ are the same products (with factors written in a possibly different order).

Thus $\det(T) = \sum_{\sigma \in \text{Sym}(n)} sg(\sigma) t_{\sigma(1),1} \dots t_{\sigma(n),n} = \sum_{\rho \in \text{Sym}(n)} t_{1,\rho(1)} \dots t_{n,\rho(n)} = \det(T^t)$. //

4.5 Cofactors and the Adjugate Matrix

Let T be an $n \times n$ matrix. Let \mathbf{c}_j be the j^{th} column of T ; so $\mathbf{c}_j = (t_{1,j}, \dots, t_{n,j})^t$ and $\det(T) = \Delta(\mathbf{c}_1, \dots, \mathbf{c}_n)$. Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be the standard basis for F^n , and $a_{j,i} = \Delta(\mathbf{c}_1, \dots, \mathbf{c}_{j-1}, \mathbf{e}_i, \mathbf{c}_{j+1}, \dots, \mathbf{c}_n)$.

Since we can subtract any multiple of \mathbf{e}_i from any of the other columns without changing the value of the determinant, we have that $a_{j,i}$ is the determinant of the matrix obtained by putting all the i^{th} row and j^{th} column of T equal to 0 except the (i, j) entry which is set at 1.

Definition 4.5.1 $a_{i,j}$ is called the (j, i) -cofactor of T . The matrix with (i, j) entry the (j, i) cofactor $(a_{i,j})$ is called the adjugate of T and is denoted by $\text{adj}(T)$.

Proposition 4.5.2 Let T be an $n \times n$ matrix. Then $\text{adj}(T)T = \det(T)I_n = T\text{adj}(T)$. Hence $T^{-1} = \text{adj}(T)/\det(T)$ if $\det(T) \neq 0$.

Proof. Let $A = \text{adj}(T)$. Then $\Delta(\mathbf{c}_1, \dots, \mathbf{c}_{j-1}, \mathbf{y}, \mathbf{c}_{j+1}, \dots, \mathbf{c}_n) = \sum_{i=1}^n a_{j,i}y_i$. So if $\mathbf{y} = \mathbf{c}_k$, we get $\sum_{i=1}^n a_{j,i}t_{i,k} = \Delta(\mathbf{c}_1, \dots, \mathbf{c}_{j-1}, \mathbf{c}_k, \mathbf{c}_{j+1}, \dots, \mathbf{c}_n) = (\det(T))\delta_{j,k}$. Thus $AT = \det(T)I_n$. Hence $\text{adj}(T^t)T^t = \det(T^t)I_n$. By Proposition 4.4.6 this gives $\text{adj}(T)^t T^t = \det(T)I_n$. The proof is completed by taking transposes of each side and recalling that T^{-1} exists iff $\det(T) \neq 0$. //

4.6 Exercises

1. Is it true that square matrices with the same size, rank, determinant and trace are conjugate? Is it true for 2×2 matrices?

2. Let A and B be $n \times n$ matrices over the field F . Show that the $(2n \times 2n)$ matrix

$$C = \begin{pmatrix} I & B \\ -A & 0 \end{pmatrix} \text{ can be transformed into } D = \begin{pmatrix} I & B \\ 0 & AB \end{pmatrix}$$

by elementary row operations. By considering the determinants of C and D , obtain another proof that $\det(AB) = \det(A)\det(B)$.

3. Find the characteristic and minimal polynomials and the eigenvalues of the matrix

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

Also find the algebraic and geometric multiplicities of these eigenvalues.

4. Suppose that $\alpha \in \mathcal{L}(V, V)$ is invertible. Describe the characteristic and minimal polynomials and the eigenvalues of α^{-1} in terms of those of α . Hence find the characteristic and minimal polynomials and eigenvalues of A^{-1} in terms of those of an $n \times n$ invertible matrix A .

5. Let $f(x) = a_0 + a_1x + \dots + a_nx^n$, with $a_i \in \mathbb{C}$, and let C be the circulant matrix

$$\begin{pmatrix} a_0 & a_1 & a_2 & \dots & a_n \\ a_n & a_0 & a_1 & \dots & a_{n-1} \\ a_{n-1} & a_n & a_0 & \dots & a_{n-2} \\ \vdots & & & \ddots & \vdots \\ a_1 & a_2 & a_3 & \dots & a_0 \end{pmatrix}$$

Show that

$$\det C = \prod_{j=0}^{n-1} f(\zeta^j)$$

where $\zeta = \exp(2\pi i/(n+1))$.

6. Let A, B be $n \times n$ matrices, where $n \geq 2$. Show that, if A and B are non-singular, then $\text{adj}(AB) = \text{adj}(B)\text{adj}(A)$ and $\det(\text{adj} A) = (\det A)^{n-1}$ and $\text{adj}(\text{adj} A) = (\det A)^{n-2}A$.

What happens if A is singular?

Show that the rank of the matrix $\text{adj} A$ is

$$\text{rk}(\text{adj}(A)) = \begin{cases} n & \text{if } \text{rk}(A) = n \\ 1 & \text{if } \text{rk}(A) = n - 1 \\ 0 & \text{if } \text{rk}(A) \leq n - 2 \end{cases}$$

7. Let A be the $n \times n$ complex matrix with

$$a_{ij} = \begin{cases} \lambda & \text{if } i = j \\ 1 & \text{if } i + 1 = j \\ 0 & \text{otherwise} \end{cases}$$

Find all of the eigenvalues of A and their geometric multiplicities. Find the characteristic polynomial of A and hence the algebraic multiplicity of each eigenvalue. Find the minimal polynomial for A .

Suppose that λ_k ($k = 1, 2, \dots, K$) are K distinct complex numbers and a_k, g_k are natural numbers with $1 \leq g_k \leq a_k$ for each k . Construct a

matrix A whose eigenvalues are precisely λ_k for $k = 1, 2, \dots, K$ and for which λ_k has geometric multiplicity g_k and algebraic multiplicity a_k .

8. Let

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

Verify directly that the Cayley - Hamilton theorem holds for A . Hence compute $A^7 - 2A^5 + 2A^4 - 2A^2 + 2A + I$ and A^{-1} .

Tripes questions: 98307, 98415, 99415, 00317.

Chapter 5

The Dual Space

5.1 The Dual and Double Dual

Definition 5.1.1 Let V be a vector space over a field F . The dual space of V is the space $\mathcal{L}(V, F)$ and is denoted by V^* . The double dual of V is the space $V^{**} = \mathcal{L}(V^*, F) = \mathcal{L}(\mathcal{L}(V, F), F)$.

If V is finite dimensional, then $\dim(V^*) = \dim(V)\dim(F) = \dim(V)$; thus $V \cong V^* \cong V^{**}$ in this case.

We wish to examine this isomorphism more closely. Recall that F^n is the vector space of all column vectors and write $(F^n)^t$ for the space of all row vectors. If $\mathbf{a} \in (F^n)^t$, then $\phi_{\mathbf{a}} : \mathbf{x} \mapsto \mathbf{a}\mathbf{x}$ is a linear map from F^n into F . If $\phi \in \mathcal{L}(F^n, F)$, let $\mathbf{a} = (\phi(\mathbf{e}_1), \dots, \phi(\mathbf{e}_n)) \in (F^n)^t$, then $\mathbf{a}\mathbf{x} = \sum_{j=1}^n \phi(\mathbf{e}_j)x_j = \phi(\sum_{j=1}^n x_j\mathbf{e}_j) = \phi(\mathbf{x})$. Thus $\phi = \phi_{\mathbf{a}}$. Consequently, the map $\mathbf{a} \mapsto \phi_{\mathbf{a}}$ is a surjective map of $(F^n)^t$ onto $(F^n)^*$. It is linear and has kernel $\{\mathbf{0}\}$ so $(F^n)^t \cong (F^n)^*$. From now on we will therefore identify the dual space of F^n with the space of all row vectors.

Example 5.1.2 Let \mathcal{P} be the vector space $\mathbb{R}[X]$ of all real polynomials and $\phi \in \mathcal{L}(\mathcal{P}, \mathbb{R})$. For each non-negative integer j let $a_j = \phi(X^j)$. This gives a real sequence $a \in F^{\mathbb{N}}$. Since ϕ is linear, $\phi(p_0 + p_1X + \dots + p_nX^n) = \sum_{j=0}^n a_j p_j$. Note that the isomorphism (between \mathcal{P} and the subspace of all sequences that are eventually zero) induces an isomorphism between \mathcal{P}^* and the space of all real sequences. Now $\dim(\mathcal{P}) = \aleph_0$ but \mathcal{P}^* has no countable basis (see Exercise 1.15). Hence $\mathcal{P} \not\cong \mathcal{P}^*$. If we restrict to \mathcal{P}_n ,

the subspace of all polynomials of degree at most n , then the sequences \underline{a} with $a_j = \phi(X^j)$ ($j = 0, \dots, n$) give the dual space.

Alternatively, we can think of the dual \mathcal{P}_n^* slightly differently. Let $p \in \mathcal{P}_n$. For any sequence (a_n) , the map $p \mapsto a_0p(0) + a_1p'(0) + \dots + a_np^{(n)}(0)$ is linear and maps $p_0 + p_1X + \dots + p_nX^n$ to $p_0a_0 + p_1a_1 + 2p_2a_2 + \dots + n!p_na_n$, and every linear map from \mathcal{P}_n to \mathbb{R} arises in this way.

Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for a vector space V and $\beta_j \in V^*$ be given by $\beta_j(\mathbf{b}_i) = \delta_{i,j}$ ($1 \leq i, j \leq n$). So if $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{b}_i$, then $\beta_j(\mathbf{v}) = \sum_{i=1}^n v_i \beta_j(\mathbf{b}_i) = \sum_{i=1}^n v_i \delta_{i,j} = v_j$; i.e., $\mathbf{v} = \sum_{i=1}^n \beta_i(\mathbf{v}) \mathbf{b}_i$.

Proposition 5.1.3 *Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for a vector space V and $\beta_j \in V^*$ be given by $\beta_j(\mathbf{b}_i) = \delta_{i,j}$. Then $\{\beta_1, \dots, \beta_n\}$ is a basis for V^* .*

Proof: If $\sum_j \lambda_j \beta_j = 0$, then $\lambda_i = \sum_j \lambda_j \delta_{i,j} = \sum_j \lambda_j \beta_j(\mathbf{b}_i) = 0$ for all i . Hence $\{\beta_1, \dots, \beta_n\}$ is linearly independent. Since $\dim(V^*) = n$, the result follows. Alternatively, if $\phi \in V^*$, then $\phi - \sum_{j=1}^n \phi(\mathbf{b}_j) \beta_j$ is 0 on \mathcal{B} and hence is the 0 map; i.e., $\phi = \sum_{j=1}^n \phi(\mathbf{b}_j) \beta_j$. //

Definition 5.1.4 *If $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ is a basis for a vector space V and $\beta_j \in V^*$ is given by $\beta_j(\mathbf{b}_i) = \delta_{i,j}$, then $\{\beta_1, \dots, \beta_n\}$ is called the dual basis for V^* and is denoted by \mathcal{B}^* .*

Unfortunately, the linear map between V and V^* given by $\sum_j x_j \mathbf{b}_j \mapsto \sum_j x_j \beta_j$ is dependent on the basis \mathcal{B} chosen for V . A different basis leads to a different map. In this sense, the isomorphism is not intrinsic but basis dependent.

In contrast, consider $V^{**} = (V^*)^*$.

Proposition 5.1.5 *Let V be a finite dimensional vector space and $\mathbf{v} \in V$. Define $\epsilon_{\mathbf{v}} : V^* \rightarrow F$ by $\epsilon_{\mathbf{v}}(\phi) = \phi(\mathbf{v})$ ($\phi \in V^*$). Then the map $\epsilon : \mathbf{v} \mapsto \epsilon_{\mathbf{v}}$ is an isomorphism between V and V^{**} .*

Proof: Since ϕ is linear, $\epsilon_{\mathbf{u}+\mathbf{v}}(\phi) = \phi(\mathbf{u} + \mathbf{v}) = \phi(\mathbf{u}) + \phi(\mathbf{v}) = \epsilon_{\mathbf{u}}(\phi) + \epsilon_{\mathbf{v}}(\phi) = (\epsilon_{\mathbf{u}} + \epsilon_{\mathbf{v}})(\phi)$. Similarly, $\epsilon_{\lambda\mathbf{v}}(\phi) = \lambda\epsilon_{\mathbf{v}}(\phi)$. Thus $\epsilon_{\mathbf{u}+\mathbf{v}} = \epsilon_{\mathbf{u}} + \epsilon_{\mathbf{v}}$ and $\epsilon_{\lambda\mathbf{u}} = \lambda\epsilon_{\mathbf{u}}$; so ϵ is linear. If $\mathbf{v} \neq \mathbf{0}$, then there is a basis of V containing \mathbf{v} and a dual basis for V^* . If ψ is the dual to \mathbf{v} , then $\epsilon_{\mathbf{v}}(\psi) = \psi(\mathbf{v}) = 1$, so $\epsilon_{\mathbf{v}} \neq 0$. Hence ϵ is injective and so surjective. //

Notice that ϵ is not dependent on the choice of basis for V . In this sense we speak of the isomorphism between V and V^{**} as being “natural” (in contrast to the “unnaturalness” of the isomorphism given between V and V^*).

Also observe that apart from the surjectivity, the above proof gives a natural injective linear map between V and V^{**} ; i.e., if V is infinite dimensional, then ϵ defined above is an isomorphism between V and a subspace of V^{**} (which may be proper).

Proposition 5.1.6 *Let U and V be finite dimensional vector spaces and $\tau \in \mathcal{L}(U, V)$. Then the map $\tau^* : \phi \mapsto \phi \circ \tau$ belongs to $\mathcal{L}(V^*, U^*)$ and induces an isomorphism $*$: $\mathcal{L}(U, V) \cong \mathcal{L}(V^*, U^*)$.*

Proof: Easy verification. //

Definition 5.1.7 *With the above notation, τ^* is called the dual of τ .*

Note that if $\sigma \in \mathcal{L}(V, W)$, then $(\sigma \circ \tau)^*(\psi) = \psi \circ (\sigma \circ \tau) = \tau^*(\psi \circ \sigma) = \tau^*(\sigma^*(\psi))$ for all $\psi \in W^*$. That is, $(\sigma\tau)^* = \tau^*\sigma^*$.

Proposition 5.1.8 *Let U and V be finite dimensional vector spaces with bases \mathcal{A} and \mathcal{B} and let $\tau \in \mathcal{L}(U, V)$. If τ is represented by T with respect to these bases, then τ^* is represented by T^t with respect to the dual bases.*

Proof: Let $\mathcal{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ and $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$. Then $\tau(\mathbf{a}_j) = \sum_k t_{k,j} \mathbf{b}_k$. If $\tau^*(\beta_i) = \sum_k s_{k,i} \alpha_k$, then $\beta_i(\tau(\mathbf{a}_j)) = \beta_i(\sum_k t_{k,j} \mathbf{b}_k) = \sum_k t_{k,j} \beta_i(\mathbf{b}_k) = \sum_k t_{k,j} \delta_{i,k} = t_{i,j}$. Since $\beta_i(\tau(\mathbf{a}_j)) = (\tau^*(\beta_i))(\mathbf{a}_j) = \sum_k s_{k,i} \alpha_k(\mathbf{a}_j) = \sum_k s_{k,i} \delta_{k,j} = s_{j,i}$, we get $t_{i,j} = s_{j,i}$; i.e., $S = T^t$. //

Thus the matrix equivalent of $(\sigma\tau)^* = \tau^*\sigma^*$ is the old chestnut $(ST)^t = T^t S^t$.

If we choose bases for U and V so that τ is represented by

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

then τ^* is represented by the transpose of this.

If $\tau \in \mathcal{L}(U, V)$, then $\tau^{**} \in \mathcal{L}(U^{**}, V^{**})$ and $\tau^{**}(\epsilon_{\mathbf{u}}(\phi)) = \epsilon_{\mathbf{u}}(\tau^*(\phi)) = \epsilon_{\mathbf{u}}(\phi \circ \tau) = \phi(\tau(\mathbf{u})) = \epsilon_{\tau(\mathbf{u})}(\phi)$; i.e., $\tau^{**} \circ \epsilon = \epsilon \circ \tau$. So if we use ϵ to identify U with U^{**} and V with V^{**} , then we can/should identify τ^{**} with τ .

Proposition 5.1.9 *Let V be a finite dimensional vector space and τ be an endomorphism of V . Then $\text{tr}(\tau^*) = \text{tr}(\tau)$, $\det(\tau^*) = \det(\tau)$ and the eigenvalues of τ^* are those of τ with same multiplicities.*

Proof: Since the transpose operation preserves trace and determinant, the first two parts are immediate. Also $(\lambda i_V - \tau)^* = \lambda i_V^* - \tau^*$ which corresponds to $(\lambda I - T)^t = \lambda I^t - T^t$. Hence τ^* and τ have the same characteristic polynomial and hence the same eigenvalues with the same algebraic multiplicity. Since the row rank of a matrix equals its column rank, the geometric multiplicity of any eigenvalue in τ^* is the same as in τ . //

Definition 5.1.10 *Let W be a subspace of V . Then $W^\circ = \{\phi \in V^* : (\forall \mathbf{w} \in W)(\phi(\mathbf{w}) = 0)\}$ is called the annihilator of W .*

Let Y be a subspace of V^ . Then $Y_\circ = \{\mathbf{v} \in V : (\forall \phi \in Y)(\phi(\mathbf{v}) = 0)\}$ is called the annihilator of Y .*

Proposition 5.1.11 *Let V be a finite dimensional vector space and W and Y be subspaces of V and V^* respectively. Then W° is a subspace of V^* and Y_\circ is a subspace of V . Moreover, $\dim(W) + \dim(W^\circ) = \dim(V) = \dim(Y) + \dim(Y_\circ)$.*

Proof: That W° and Y_\circ are subspaces are routine verifications. Let $\{\mathbf{b}_1, \dots, \mathbf{b}_k\}$ be a basis of W which we extend to a basis of V . Let $\{\beta_1, \dots, \beta_n\}$ be the dual basis. Then $\{\beta_{k+1}, \dots, \beta_n\} \subseteq W^\circ$ and $\sum_{j=1}^n \lambda_j \beta_j \in W^\circ \iff \sum_{j=1}^n \lambda_j \beta_j(\mathbf{b}_i) = 0$ for $i = 1, \dots, k \iff \lambda_i = 0$ for $i = 1, \dots, k$. Hence $\{\beta_{k+1}, \dots, \beta_n\}$ spans W° and is linearly independent. Thus we get $\dim(W) + \dim(W^\circ) = \dim(V)$.

Note that $\mathbf{v} \in Y_\circ \iff \phi(\mathbf{v}) = 0$ all $\phi \in Y \iff \epsilon_{\mathbf{v}}(\phi) = 0$ all $\phi \in Y \iff \epsilon_{\mathbf{v}} \in Y^\circ \subseteq V^{**}$. So $\dim(Y_\circ) = \dim(Y^\circ) = \dim(V^*) - \dim(Y) = \dim(V) - \dim(Y)$ by the first part. //

Corollary 5.1.12 *For V a finite dimensional vector space and W and Y subspaces of V and V^* respectively, $(W^\circ)_\circ = W$ and $(Y_\circ)^\circ = Y$.*

Proof: For $\mathbf{w} \in W$, $\phi \in W^\circ$ implies $\phi(\mathbf{w}) = 0$, whence $\mathbf{w} \in (W^\circ)_\circ$. So W is a subspace of $(W^\circ)_\circ$. But by Proposition 5.1.11, W and $(W^\circ)_\circ$ have the same dimension. Hence they are equal. The other equality follows similarly. //

Proposition 5.1.13 *Let U and V be finite dimensional vector spaces and $\tau \in \mathcal{L}(U, V)$. Then $\ker(\tau^*) = (\text{Im}(\tau))^\circ$ and $\text{Im}(\tau^*) = (\ker(\tau))^\circ$.*

Proof: $\phi \in \ker(\tau^*) \iff 0 = \tau^*(\phi) = \phi \circ \tau \iff \phi(\tau(\mathbf{u})) = 0 \ (\forall \mathbf{u} \in U) \iff \phi \in (\text{Im}(\tau))^\circ$.

Similarly, $\mathbf{v} \in \ker(\tau) \iff \mathbf{v} \in (\text{Im}(\tau^*))^\circ$.
Thus $\ker(\tau)^\circ = ((\text{Im}(\tau^*))^\circ)^\circ = \text{Im}(\tau^*)$. //

Corollary 5.1.14 *If U and V are finite dimensional vector spaces and $\tau \in \mathcal{L}(U, V)$, then $\text{rank}(\tau) = \text{rank}(\tau^*)$.*

Proof: Since $\text{rank}(\tau) + \dim(\ker(\tau)) = \dim(U)$, the result follows at once from Propositions 5.1.11 and 5.1.13. //

5.2 Exercises

1. If A and B are $n \times m$ and $m \times n$ matrices over the field F , let $\tau_A(B) = \text{tr}(AB)$. Show that, for each A , τ_A is a linear map $\mathcal{M}_{m \times n}(F) \rightarrow F$. Show further that the mapping $A \mapsto \tau_A$ is a linear map $\mathcal{M}_{m \times n}(F) \rightarrow \mathcal{M}_{m \times n}(F)^*$, and that it is an isomorphism.

2. Let \mathbf{x} be a non-zero vector in the finite dimensional vector space V . Show that there is a linear functional $f \in V^*$ such that $f(\mathbf{x}) \neq 0$. Deduce that if $\mathbf{x} \neq \mathbf{y}$ are vectors in V , then there is a linear functional $f \in V^*$ such that $f(\mathbf{x}) \neq f(\mathbf{y})$.

3. Show that the dual of the space P of real polynomials is isomorphic to the space $\mathbb{R}^{\mathbb{N}}$ of all sequences of real numbers, via the mapping which sends a linear form $\xi : P \rightarrow \mathbb{R}$ to the sequence $(\xi(1), \xi(t), \xi(t^2), \dots)$. In terms of this identification, describe the effect on a sequence (a_0, a_1, a_2, \dots) of the linear maps dual to each of the following linear maps $P \rightarrow P$: (a) The map D defined by $D(p)(t) = p'(t)$. (b) The map S defined by $S(p)(t) = p(t^2)$. (c) The map E defined by $E(p)(t) = p(t-1)$. (d) The composite DS . (e) The composite SD . Verify that $(DS)^* = S^*D^*$ and $(SD)^* = D^*S^*$.

4. Let $\tau : V \rightarrow V$ be an endomorphism of a finite dimensional complex vector space and let $\tau^* : V^* \rightarrow V^*$ be its dual.

Show that a complex number λ is an eigenvalue for τ if, and only if, it is an eigenvalue for τ^* . How are the algebraic and geometric multiplicities of λ for τ and τ^* related? How are the minimal and characteristic polynomials for τ and τ^* related?

5. Let \mathcal{F} be a subset of the dual V^* of a finite dimensional vector space V . Show that \mathcal{F} spans V^* if, and only if,

$$f(\mathbf{v}) = 0 \text{ for all } f \in \mathcal{F} \iff \mathbf{v} = \mathbf{0} .$$

Tripes questions: 98406, 99406, 00215.