

Results in Linear Mathematics (P1)

T.W.Körner

May 19, 2003

Small print The syllabus for the course is defined by the Faculty Board Schedules (which are minimal for lecturing and maximal for examining). I have starred certain results which seem to me to go beyond a strict interpretation of the syllabus. However whilst it would not, in my opinion, be fair to set such results as bookwork they could well appear as problems. I should **very much** appreciate being told of any corrections or possible improvements. This document is written in L^AT_EX and stored in the file labeled `~twk/1B/V1.tex` on emu in (I hope) read permitted form. My e-mail address is `twk`.

1 Vector Spaces

Convention 1.1 *We shall write \mathbb{F} to mean \mathbb{R} or \mathbb{C} .*

Definition 1.2 *We call $(V, +, \cdot, \mathbb{F})$ a vector space over \mathbb{F} if, whenever $u, v, w \in V$ and $\lambda, \mu \in \mathbb{F}$, then $u + v \in V$, $\lambda u \in V$ and*

(i) $(V, +)$ is an Abelian group (so in particular $u + (v + w) = (u + v) + w$, $u + v = v + u$).

(ii) $\lambda(\mu u) = (\lambda\mu)u$.

(iii) $(\lambda + \mu)u = \lambda u + \mu u$.

(iv) $\lambda(u + v) = \lambda u + \lambda v$.

(v) $1u = u$.

Lemma 1.3 *(i) The zero $\underline{0}$ of $(V, +)$ satisfies $0u = \underline{0}$ for all $u \in V$.*

(ii) The additive inverse $-u$ of $u \in V$ satisfies $-u = (-1)u$.

We call $\underline{0}$ the zero vector and write it as 0. (Our general policy of dropping ‘boldface’ \mathbf{u} and ‘underline’ \underline{u} in favour of the simple u will not usually lead to ambiguity but if it does we simply revert to the less simple convention.)

Theorem 1.4 *If X is any set then the set \mathbb{F}^X of functions $f : X \rightarrow \mathbb{F}$ is a vector space if we define ‘vector addition’ and ‘multiplication by a scalar’ by*

$$(f + g)(x) = f(x) + g(x), \text{ and } (\lambda f)(x) = \lambda f(x)$$

for all $x \in X$ where $f, g \in \mathbb{F}^X$ and $\lambda \in \mathbb{F}$.

Definition 1.5 *If V is a vector space we say that $U \subseteq V$ is a subspace of V if $0 \in U$ and*

$$(\lambda, \mu \in \mathbb{F}, u, v \in U) \Rightarrow \lambda u + \mu v \in U.$$

Lemma 1.6 *If U is a subspace of a vector space V then U is itself a vector space.*

It is usually easier to use Theorem 1.4 (or its generalisation Theorem 2.7 below) together with Lemma 1.6 to prove that something is a vector space than to verify the axioms in Definition 1.2.

Example 1.7 *The space $C([0, 1])$ of continuous functions $f : [0, 1] \rightarrow \mathbb{F}$, the space \mathcal{P} of real polynomials $P : \mathbb{R} \rightarrow \mathbb{R}$, the classical spaces \mathbb{F}^n and the set J of $n \times n$ real matrices all of whose rows and columns add up to the same number can all be made into vector spaces in a natural way.*

Definition 1.8 (i) *Vectors e_1, e_2, \dots, e_n span a vector space E if given any $e \in E$ we can find $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{F}$ such that*

$$e = \lambda_1 e_1 + \lambda_2 e_2 + \dots + \lambda_n e_n.$$

(ii) *Vectors e_1, e_2, \dots, e_n in a vector space E are linearly independent if the only solution of*

$$0 = \lambda_1 e_1 + \lambda_2 e_2 + \dots + \lambda_n e_n$$

(with $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{F}$) is $\lambda_1 = \lambda_2 = \dots = \lambda_n = 0$.

(iii) *Vectors e_1, e_2, \dots, e_n form a (finite) basis of a vector space E if they span E and are linearly independent.*

We shall not be interested in ‘infinite bases’ (few people are, in an algebraic context) so we shall write ‘basis’ rather than ‘(finite) basis’ from now on.

Lemma 1.9 *Vectors e_1, e_2, \dots, e_n form a basis of a vector space E if and only if the equation*

$$e = \lambda_1 e_1 + \lambda_2 e_2 + \dots + \lambda_n e_n$$

(with $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{F}$) has one and only one solution for each $e \in E$.

Definition 1.10 A vector space E is said to be finite dimensional if it has a finite spanning set.

Lemma 1.11 (i) If vectors e_1, e_2, \dots, e_n span a vector space E then some sub-collection forms a basis.

(ii) Every finite dimensional vector space has a basis.

The key result in the study of finite dimensional vector spaces is the Steinitz Replacement Lemma.

Theorem 1.12 Let E be a vector space. If

(A) e_1, e_2, \dots, e_n span E , and

(B) f_1, f_2, \dots, f_m are linearly independent in E ,

then $n \geq m$ and (possibly after reordering the e_j) $f_1, f_2, \dots, f_m, e_{m+1}, e_{m+2}, \dots, e_n$ span E .

Corollary 1.13 Every finite dimensional space E has an associated dimension N such that

(i) Every basis of E has N elements.

(ii) Every linearly independent collection of vectors in E has at most N elements.

(iii) Every spanning collection of vectors in E has at least N elements.

Corollary 1.14 (i) Any subspace of a finite dimensional space is finite dimensional (and the dimension of the subspace is no greater than the dimension of the space).

(ii) Any linearly independent collection of vectors in a finite dimensional space can be extended to a basis.

Example 1.15 In Example 1.7 the space $C([0, 1])$ of continuous functions is infinite dimensional and the space J of $n \times n$ magic squares is finite dimensional.

Definition 1.16 If V and W are subspaces of a vector space U we write

$$V + W = \{v + w : v \in V, w \in W\}$$

Lemma 1.17 If V and W are subspaces of a vector space U then $V + W$ and $V \cap W$ are subspaces of U . Further, if $V + W$ is finite dimensional,

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

Definition 1.18 Subspaces E_1, E_2, \dots, E_m of a vector space E are said to have E as direct sum if and only if the equation

$$e = e_1 + e_2 + \dots + e_m$$

(with $e_j \in E_j$ for $1 \leq j \leq m$) has one and only one solution for each $e \in E$. We then write

$$E = E_1 \oplus E_2 \oplus \dots \oplus E_m.$$

Lemma 1.19 Subspaces E_1, E_2, \dots, E_m of a finite dimensional vector space E have E as direct sum if and only if the combination of bases of E_1, E_2, \dots, E_m gives a basis of E .

Lemma 1.20 If V and W are subspaces of a vector space U then $V \oplus W = U$ if and only if $V + W = U$ and $V \cap W = \{0\}$.

The reader is warned that Lemma 1.20 does not generalise in the obvious way to direct sums of more than two subspaces.

Example 1.21 In \mathbb{R}^2 if we set

$$U = \{(x, 0) : x \in \mathbb{R}\}$$

$$V = \{(0, y) : y \in \mathbb{R}\}$$

$$W = \{(t, t) : t \in \mathbb{R}\}$$

then $U \cap V = V \cap W = W \cap U = \{0\}$ and $U + V + W = \mathbb{R}^2$ but \mathbb{R}^2 is not the direct sum of U, V and W .

Definition 1.22 If V and W are subspaces of a vector space U and $V \oplus W = U$ then W is called a complementary subspace of V in U .

The reader is warned that this definition is a strong competitor for the title of 'Definition most frequently mangled by undergraduates'. She is also warned that except in the trivial cases $V = U$ and $V = \{0\}$ the complementary subspace of V in U IS NOT UNIQUE! In Example 1.21 both U and V are complementary subspaces of W in \mathbb{R}^2 .

Example 1.23 Consider the vector space $C(\mathbb{R})$ of continuous functions $f : \mathbb{R} \rightarrow \mathbb{R}$. Let

$$E = \{f \in C(\mathbb{R}) : f(x) = f(-x) \text{ for all } x \in \mathbb{R}\}$$

$$F = \{f \in C(\mathbb{R}) : f(x) = -f(-x) \text{ for all } x \in \mathbb{R}\}$$

$$G = \{f \in C(\mathbb{R}) : f(x) = 0 \text{ for all } x \leq 0\}.$$

Then both F and G are complements of E in $C(\mathbb{R})$.

The course now contains some remarks on quotient groups. These are starred, but only to prevent the examiners going overboard, the actual ideas are very easy.

The development is parallel to, but easier than the development of quotient groups in course C1. Suppose that U is a subspace of a vector space V . We observe that U is a subgroup of $(V, +)$ the the vector space V considered as an Abelian group under addition. We take over from group theory the idea of a coset

$$v + U = \{v + u : u \in U\}$$

and observe that the first part of the proof of Lagrange's theorem shows that the cosets form a disjoint cover of V .

Lemma 1.24 *Let U be a subspace of a vector space V . Then*

(i) $\bigcup_{v \in V} (v + U) = V$.

(ii) *If $v, w \in V$ then either $(v + U) \cap (w + U) = \emptyset$ or $v + U = w + U$.*

The remarkable thing is that we can define addition and scalar multiplication of these cosets in a natural way. (Of course, we can deal with addition by noting that any subgroup of an Abelian group is normal and quoting course C1 but it is just as easy to do things directly.)

Lemma 1.25 *If U is a subspace of a vector space V over \mathbb{F} and*

$$v_1 + U = v_2 + U, \quad w_1 + U = w_2 + U, \quad \lambda \in \mathbb{F}$$

then

$$(v_1 + w_1) + U = (v_2 + w_2) + U, \quad \lambda v_1 + U = \lambda v_2 + U.$$

Definition 1.26 *If U is a subspace of a vector space V over \mathbb{F} we write V/U for the set of cosets of U and define addition and scalar multiplication on V/U by*

$$(v + U) + (w + U) = (v + w) + U, \quad \lambda(v + U) = \lambda v + U.$$

Note that $0(v + U) = 0v + U = U$.

Lemma 1.27 *If U is a subspace of a vector space V over \mathbb{F} then V/U with addition and scalar multiplication as in the previous definition is a vector space over \mathbb{F} .*

We call V/U a quotient space (or a quotient vector space).

The reader will natural expect us to produce an isomorphism theorem. We shall do so in Theorem 2.26 but first we need to discuss linear maps.

2 Linear Maps

Definition 2.1 If U and V are vector spaces over \mathbb{F} then the map $\alpha : U \rightarrow V$ is said to be linear if

$$\alpha(\lambda_1 u_1 + \lambda_2 u_2) = \lambda_1 \alpha(u_1) + \lambda_2 \alpha(u_2)$$

for all $u_j \in U_j$, $\lambda_j \in \mathbb{F}$.

If abstract algebraists were the only people to use vector spaces then ‘linear maps’ would be called ‘vector space homomorphisms’. The following definitions may have been mentioned in previous courses.

Definition 2.2 Let $f : X \rightarrow Y$.

(i) We say that f is injective if $f(x_1) = f(x_2)$ implies $x_1 = x_2$.

(ii) We say that f is surjective if, given $y \in Y$, we can find an $x \in X$ such that $f(x) = y$.

(iii) We say that f is bijective if it is both injective and surjective.

If $f : X \rightarrow Y$ is bijective then there is a unique function $f^{-1} : Y \rightarrow X$ (the inverse of f) such that $f^{-1}(f(x)) = x$ for all $x \in X$ and $f(f^{-1}(y)) = y$ for all $y \in Y$.

Definition 2.3 If U and V are vector spaces over \mathbb{F} the linear map $\alpha : U \rightarrow V$ is said to be an isomorphism if it is a bijection.

Lemma 2.4 If $\alpha : U \rightarrow V$ is an isomorphism then $\alpha^{-1} : V \rightarrow U$ is also linear (and so also an isomorphism).

Lemma 2.5 If U and V are vector spaces over \mathbb{F} and $\alpha : U \rightarrow V$ is linear then

(i) $\alpha(U) = \{\alpha(u) : u \in U\}$ is a subspace of V .

(ii) $\alpha^{-1}(0) = \{u \in U : \alpha(u) = 0\}$ is a subspace of U .

(iii) α is injective if and only if $\alpha^{-1}(0) = \{0\}$.

We call $\alpha^{-1}(0)$ the null space of α and $\alpha(U)$ the range space of α . (Note that, unless α is bijective, there is no inverse function α^{-1} .)

Theorem 2.6 (The Classification Theorem For Finite Dimensional Vector Spaces) Every vector space over \mathbb{F} of dimension N is isomorphic to \mathbb{F}^N .

We now give the promised generalisation of Theorem 1.4

Theorem 2.7 *If X is any set and U any vector space over \mathbb{F} then the set U^X of functions $f : X \rightarrow U$ is a vector space over \mathbb{F} if we define ‘vector addition’ and ‘multiplication by a scalar’ by*

$$(f + g)(x) = f(x) + g(x), \text{ and } (\lambda f)(x) = \lambda f(x)$$

for all $x \in X$ where $f, g \in U^X$ and $\lambda \in \mathbb{F}$.

Theorem 2.8 *If U and V are vector spaces over \mathbb{F} then the set $L(U, V)$ of linear maps forms a vector space if we define ‘vector addition’ and ‘multiplication by a scalar’ by*

$$(\alpha + \beta)(u) = \alpha(u) + \beta(u), \text{ and } (\lambda\alpha)(u) = \lambda\alpha(u)$$

for all $u \in U$ where $\alpha, \beta \in L(U, V)$ and $\lambda \in \mathbb{F}$.

The operation of composition interacts with the operation of addition just defined in a suggestive way.

Theorem 2.9 *If U, V and W are vector spaces over \mathbb{F} and $\alpha, \beta \in L(U, V)$, $\gamma, \delta \in L(V, W)$, $\epsilon \in L(W, X)$ then*

$$\epsilon(\gamma\alpha) = (\epsilon\gamma)\alpha, \quad \gamma(\alpha + \beta) = \gamma\alpha + \gamma\beta, \quad (\gamma + \delta)\alpha = \gamma\alpha + \delta\alpha.$$

Theorems 2.7 and 2.9 apply in particular to $L(U, U)$.

Theorem 2.10 *If U is a vector space over \mathbb{F} then $L(U, U)$ is a vector space over \mathbb{F} which obeys the additional laws*

$$\alpha(\beta\gamma) = (\alpha\beta)\gamma, \quad \alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma, \quad (\alpha + \beta)\gamma = \alpha\gamma + \beta\gamma, \quad \iota\alpha = \alpha\iota,$$

where ι is the identity map.

(A vector space which obeys these laws is called an ‘algebra’ but the definition is not part of this course.)

Theorem 2.11 *If U is a vector space over \mathbb{F} then the set of bijective (‘invertible’) maps in $L(U, U)$ form a group $GL(U)$ under composition with unit the identity map ι .*

$GL(U)$ is called ‘the general linear group’ and its elements are called automorphisms.

Theorem 2.10 forms a link with an older but not unsuccessful tradition of formal symbolic manipulation.

Example 2.12 (i) Consider the vector space $C_{\mathbb{C}}^{\infty}(\mathbb{R})$ of infinitely differentiable functions $f : \mathbb{R} \rightarrow \mathbb{C}$. The map $D : C_{\mathbb{C}}^{\infty}(\mathbb{R}) \rightarrow C_{\mathbb{C}}^{\infty}(\mathbb{R})$ is a well defined linear map. If $a_0, a_1, \dots, a_n \in \mathbb{R}$ then, taking $D^0 = I$ the identity function

$$\left(\sum_{r=0}^n a_r D^r \right) f = \sum_{r=0}^n a_r f^{(r)}.$$

(ii)* If $a \in \mathbb{C}$ then $D - aI$ is surjective with a one dimensional nul space. The map $T_a : C_{\mathbb{C}}^{\infty}(\mathbb{R}) \rightarrow C_{\mathbb{C}}^{\infty}(\mathbb{R})$ defined by

$$(T_a f)(t) = e^{at} \int_0^t f(x) e^{-ax} dx$$

is linear and injective but not surjective. We have

$$(D - aI)T_a = I.$$

The general solution f of

$$(D - aI)f = g$$

(with $f, g \in C_{\mathbb{C}}^{\infty}(\mathbb{R})$) is $f = T_a g + h$ with $h \in (D - aI)^{-1}(0)$.

(iii)* If $a_j \in \mathbb{C}$ for $1 \leq j \leq n$ then $(D - a_1 I)(D - a_2 I) \dots (D - a_n I)$ is surjective with an n dimensional nul space H . The equation

$$(D - a_1 I)(D - a_2 I) \dots (D - a_n I)f = g$$

(with $f, g \in C_{\mathbb{C}}^{\infty}(\mathbb{R})$) always has a solution f_0 (the ‘particular integral’) and its general solution is $f = f_0 + h$ where h (the ‘complementary function’) lies in the n dimensional nul space H .

Although general theorems about linear maps are often best viewed geometrically or abstractly, particular computations require matrices.

Theorem 2.13 Suppose that U is a finite dimensional vector space over \mathbb{F} with basis u_1, u_2, \dots, u_m and V is a finite dimensional vector space over \mathbb{F} with basis v_1, v_2, \dots, v_n . If $\alpha : U \rightarrow V$ is linear then α has an associated $n \times m$ matrix $A = (a_{ij})$ with entries in \mathbb{F} given by

$$\alpha(u_j) = \sum_{i=1}^n a_{ij} v_i. \quad (*)$$

Automatically

$$\alpha\left(\sum_{j=1}^m x_j u_j\right) = \sum_{i=1}^n \left(\sum_{j=1}^m a_{ij} x_j\right) v_i. \quad (**)$$

Conversely if A is an $n \times m$ matrix with entries in \mathbb{F} then the formula $(**)$ defines a linear map $\alpha : U \rightarrow V$ which has A as associated matrix.

We say that A is the matrix of α with respect to the bases u_1, u_2, \dots, u_n of U and v_1, v_2, \dots, v_n of V . Equation (*) is conventional (in the same way that making clocks go clockwise is conventional) but it represents a universal convention which you should follow.

Theorem 2.14 *Suppose that U is a finite dimensional vector space over \mathbb{F} with basis u_1, u_2, \dots, u_m , V is a finite dimensional vector space over \mathbb{F} with basis v_1, v_2, \dots, v_n and W is a finite dimensional vector space over \mathbb{F} with basis w_1, w_2, \dots, w_p . If $\alpha, \beta \in L(U, V)$ have matrices A and B with respect to the bases u_1, u_2, \dots, u_m of U , v_1, v_2, \dots, v_n of V and $\gamma \in L(V, W)$ has matrix C with respect to the bases v_1, v_2, \dots, v_n of V and w_1, w_2, \dots, w_p of W and $\lambda \in \mathbb{F}$ then $\alpha + \beta$ has matrix $A + B$ and $\lambda\alpha$ has matrix λA with respect to the bases u_1, u_2, \dots, u_m of U and v_1, v_2, \dots, v_n of V and $\gamma\beta$ has matrix CB with respect to the bases u_1, u_2, \dots, u_m of U and w_1, w_2, \dots, w_p of W .*

Here $A + B$ is the matrix E given by

$$e_{ij} = a_{ij} + b_{ij} \quad \text{for } 1 \leq i \leq n, 1 \leq j \leq m,$$

λA is the matrix F given by

$$f_{ij} = \lambda a_{ij} \quad \text{for } 1 \leq i \leq n, 1 \leq j \leq m,$$

and CB is the matrix G given by

$$g_{rj} = \sum_{i=1}^n c_{ri} b_{ij} \quad \text{for } 1 \leq r \leq p, 1 \leq j \leq m$$

Notice that if, in Theorem 2.13, we write

$$\alpha \left(\sum_{j=1}^m x_j u_j \right) = \sum_{i=1}^n y_i v_i$$

and write \mathbf{x} for the column vector (i.e. $m \times 1$ matrix) (x_j) , \mathbf{y} for the column vector (y_i) then equation (**) becomes

$$\mathbf{y} = A\mathbf{x}.$$

The exact correspondence between linear maps of finite dimensional vector spaces and matrices means that we can rewrite theorems about maps as theorems about matrices. For example, restricting Theorem 2.10 to finite dimensional spaces, we obtain the following matricial translation.

Theorem 2.15 *If we use the standard matrix addition, multiplication and multiplication by a scalar the set $M_n(\mathbb{F})$ of $n \times n$ matrices with entries in \mathbb{F} is a vector space satisfying the further rules*

$$A(B + C) = AB + AC, \quad A(B + C) = AB + AC, \quad (A + B)C = AC + BC, \quad IA = AI = A,$$

where I is the identity matrix with (i, j) th entry δ_{ij} .

Conversely we can use results about matrices to obtain results on linear maps between finite dimensional spaces.

Theorem 2.16 *(i) The vector space $M_{nm}(\mathbb{F})$ of $n \times m$ matrices over \mathbb{F} has dimension nm .*

(ii) If U and V are vector spaces of dimension n and m then $L(U, V)$ has dimension nm .

Since Theorem 2.6 tells us that all finite dimensional vector spaces are isomorphic to \mathbb{F}^n for some n and since linear maps between such spaces can always be represented by matrices, it is clearly possible to do all questions involving finite dimensional vector spaces by taking bases and using co-ordinates. There are however several reasons for using co-ordinate free methods when possible.

- As Maxwell pointed out, co-ordinate free methods often give a better formulation of the underlying physical or geometric problem.
- For analysts and most physicists the study of finite dimensional is merely a prelude to the study of infinite dimensional spaces where co-ordinate methods are often not available.
- One of the reasons for doing the course P1 is to learn more abstract modes of thought. Sticking to concrete co-ordinate systems is hardly the way to go about it.

Returning to the ideas associated with Lemma 2.5 we make the following definitions.

Definition 2.17 *If U and V are finite dimensional vector spaces over \mathbb{F} and $\alpha : U \rightarrow V$ is linear then*

- (i) The rank $r(\alpha)$ of α is the dimension of the range space $\alpha(U)$.*
- (ii) The nullity $n(\alpha)$ of α is the dimension of the nul space $\alpha^{-1}(0)$.*

Theorem 2.18 *If U and V are finite dimensional vector spaces over \mathbb{F} and $\alpha : U \rightarrow V$ is linear then*

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

Further we can find bases for U and V such that the matrix A associated with α is given by $a_{ij} = 1$ if $1 \leq i = j \leq r(\alpha)$, $b_{ij} = 0$ otherwise.

Formula (*) is the famous ‘rank-nullity formula’.

Theorem 2.19 *If U is a vector space over \mathbb{F} and $\alpha : U \rightarrow V$ is linear then*

$$U \supseteq \alpha(U) \supseteq \alpha^2(U) \supseteq \dots \supseteq \alpha^k(U) \supseteq \alpha^{k+1}(U) \supseteq \dots .$$

If $\alpha^l(U) = \alpha^{l+1}(U)$ then $\alpha^k(U) = \alpha^l(U)$ for all $k \geq l$.

If U is finite dimensional then

$$\dim(U) \geq r(\alpha) \geq r(\alpha^2) \geq \dots ,$$

and there exists an l with $l \leq \dim(U)$ such that $r(\alpha^k) > r(\alpha^{k+1})$ for $k < l$ and $r(\alpha^k) = r(\alpha^l)$ for $k \geq l$. Further

$$r(\alpha^k) - r(\alpha^{k+1}) \geq r(\alpha^{k+1}) - r(\alpha^{k+2})$$

for all k .

Definition 2.20 *If A is a matrix with entries in \mathbb{F} then the column rank of A is the dimension of the space spanned by the column vectors of A and the row rank of A is the dimension of the space spanned by the row vectors of A .*

Lemma 2.21 *If A is matrix associated with a linear map α then the column rank of A is the rank of α .*

Later, in Theorem 3.17 (and in a ‘more natural manner’ in Theorem 6.10) we shall see that the row and column rank of a matrix are the same so that we can speak just of the ‘rank’ of a matrix. The following theorem used to be a high point of the first course in linear algebra but is now not even in the syllabus.

Theorem 2.22* *Let A be an $m \times n$ matrix and b an $m \times 1$ matrix (i.e. a column vector) over \mathbb{F} . Observe that the column rank of A is no greater than m . We define the ‘augmented matrix’ $(A|b)$ to be the $m \times (n+1)$ matrix whose*

first m columns are the columns of A and whose last column is b . Consider the system of m equations

$$\sum_{j=1}^n a_{ij}x_j = b_i \quad [1 \leq i \leq m], \quad (*)$$

which we write in matrix form as $Ax = b$ (where x is the $n \times 1$ matrix with entries x_j).

(i) The system $(*)$ of equations has a solution if and only if the column ranks of A and the augmented matrix $(A|b)$ are the same.

(ii) If A has column rank m then $(*)$ always has a solution. If A has column rank less than m (and so, in particular, if $n < m$) then there exist choices of b for which $(*)$ has no solution.

(iii) The column rank of A is no greater than n and there exists a subspace V of the space of $n \times 1$ matrices of dimension precisely $n - \text{column rank}(A)$ such that, if x' is a solution of $(*)$, then x is a solution of $(*)$ if and only if $x - x' \in V$. (In particular if $m < n$ $(*)$ can not have a unique solution.)

(iv) The system $(*)$ has 0, 1 or infinitely many solutions. It is a necessary (but not a sufficient) condition for $(*)$ to have a unique solution for each b that $n = m$.

As we have already noticed, the matrix associated with a linear map depends on the bases chosen.

Theorem 2.23 [The Change of Basis Theorem] Suppose that U is a finite dimensional vector space over \mathbb{F} with two bases u_1, u_2, \dots, u_m and u'_1, u'_2, \dots, u'_m and V is a finite dimensional vector space over \mathbb{F} with two bases v_1, v_2, \dots, v_n and v'_1, v'_2, \dots, v'_n . Then we can find an $m \times m$ matrix P and an $n \times n$ matrix Q such that

$$u_j = \sum_{i=1}^m p_{ij}u'_i \quad \text{for } 1 \leq j \leq m$$

$$v_s = \sum_{r=1}^n q_{rs}v'_r \quad \text{for } 1 \leq s \leq n.$$

The matrices P and Q are invertible and if $\alpha \in L(U, V)$ has matrix A with respect to the bases u_1, u_2, \dots, u_m of U and v_1, v_2, \dots, v_n of V and matrix B with respect to the bases u'_1, u'_2, \dots, u'_m of U and v'_1, v'_2, \dots, v'_n then

$$B = Q^{-1}AP.$$

Corollary 2.24 *Let U be an m dimensional vector space over \mathbb{F} and V an n dimensional vector space. Two $m \times n$ matrices A and B can represent the same linear map from U to V (with respect to appropriately chosen bases) if and only if there exist a non singular $m \times m$ matrix P and a non singular $n \times n$ matrix Q such that $B = QAP$.*

Using Theorem 2.18 we obtain the first and easiest result on ‘canonical forms’ of matrices.

Corollary 2.25 *An $m \times n$ matrix A over \mathbb{F} has column rank r if and only if there exist a non singular $m \times m$ matrix P and a non singular $n \times n$ matrix Q such that $B = QAP$ is an $m \times n$ matrix with $b_{ij} = 1$ if $1 \leq i = j \leq r$, $b_{ij} = 0$ otherwise.*

The reader will already have seen a computational proof of this based on Gaussian elimination.

We complete this section by proving an isomorphism theorem which directly parallels the isomorphism theorem for groups. It is starred but very easy.

Theorem 2.26 *If U and V are vector spaces over \mathbb{F} and $\alpha : U \rightarrow V$ is linear then there is a natural isomorphism $\tilde{\alpha} : U/\alpha^{-1}(0) \rightarrow \alpha(U)$ given by*

$$\tilde{\alpha}(u + \alpha^{-1}(0)) = \alpha(u).$$

As usual, a key point is to show that $\tilde{\alpha}$ is well defined.

3 Endomorphisms

If $\alpha \in L(U, V)$ we defined a matrix associated with α in terms of a basis of U and a basis of V . If $U = V$, i.e. if α is an endomorphism, it seems reasonable only to use one basis.

Definition 3.1 *If V is a vector space over \mathbb{F} with basis v_1, v_2, \dots, v_n . We say that the linear map $\alpha : V \rightarrow V$ has the $n \times n$ matrix A with respect to the given basis if*

$$\alpha(v_j) = \sum_{i=1}^n a_{ij}v_i.$$

In this context the change of basis theorem takes the following form.

Theorem 3.2 (*The Change of Basis Theorem For Endomorphisms*) Suppose that V is a finite dimensional vector space over \mathbb{F} with two bases v_1, v_2, \dots, v_n and v'_1, v'_2, \dots, v'_n . Then we can find an $n \times n$ matrix P such that

$$u_j = \sum_{i=1}^m p_{ij} u'_i \text{ for } 1 \leq i \leq n.$$

The matrix P is invertible and if the linear map $\alpha : V \rightarrow V$ has matrix A with respect to the base u_1, u_2, \dots, u_m and matrix B with respect to the base v'_1, v'_2, \dots, v'_n then

$$B = P^{-1}AP.$$

Corollary 3.3 Let V be an n dimensional vector space over \mathbb{F} . Two $n \times n$ matrices A and B can represent the same endomorphism of V (with respect to appropriately chosen bases) if and only if there exists a non singular $n \times n$ matrix P such that $B = P^{-1}AP$.

We are particularly interested in the case when the $n \times n$ matrix A can be diagonalised, that is we can find a $n \times n$ diagonal matrix D (a matrix D with $d_{ij} = 0$ for $i \neq j$) and an $n \times n$ invertible matrix P such that $D = P^{-1}AP$. In this case we say that A is ‘diagonalisable’. One of the many reasons for being interested in this phenomenon is indicated by the observation that

$$A^N = (PDP^{-1})^N = PD^N P^{-1}$$

and that D^N is a diagonal matrix whose diagonal entries are the N th powers of the corresponding diagonal entries of D . In the same way an endomorphism of a finite dimensional vector space V is called diagonalisable if we can find a basis of V with respect to which the matrix associated to α is diagonal. It is important to note that (even over \mathbb{C}) NOT ALL SQUARE MATRICES ARE DIAGONALISABLE.

Example 3.4 The matrix

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

is not diagonalisable over \mathbb{F} .

With very little exaggeration it may be said that all putative theorems about matrices should be tested on the matrix just given.

The problem of diagonalisation is closely linked to the existence of eigenvalues and eigenvectors.

Definition 3.5 Let V be a vector space over \mathbb{F} and let α be an endomorphism of V . If e is a non-zero vector, $\lambda \in \mathbb{F}$ and

$$\alpha e = \lambda e$$

then we say that λ is an eigenvalue and e an eigenvector of α .

Eigenobjects (though associated with infinite rather than finite dimensional spaces) occur throughout modern physics.

Lemma 3.6 An endomorphism of a finite dimensional vector space is diagonalisable if and only if the space has a basis of eigenvectors.

Lemma 3.7 Eigenvectors with distinct eigenvalues are linearly independent. In particular an endomorphism of an n dimensional vector space which has n distinct eigenvalues is diagonalisable.

Our treatment of eigenvectors will use determinants. Recall that the signature function ζ is the unique non trivial homomorphism from the group $S(n)$ of permutations on the set $\{1, 2, \dots, n\}$ to the multiplicative group $\{1, -1\}$.

Definition 3.8 If A is an $n \times n$ matrix over \mathbb{F} then

$$\det A = \sum_{\sigma \in S(n)} \zeta(\sigma) \prod_{i=1}^n a_{i, \sigma i}.$$

Lemma 3.9 If A is an $n \times n$ matrix with rows $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ let us write

$$f(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \det A.$$

Then

- (i) f is linear in each variable.
- (ii) f is alternating in the sense that

$$f(\mathbf{a}_{\sigma 1}, \mathbf{a}_{\sigma 2}, \dots, \mathbf{a}_{\sigma n}) = \zeta(\sigma) f(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$$

for all $\sigma \in S(n)$.

- (iii) If \mathbf{e}_i is the row vector of length n with j th component δ_{ij} then

$$f(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n) = 1.$$

Moreover conditions (i), (ii) and (iii) determine f completely.

Theorem 3.10 *If A and B are $n \times n$ matrices over \mathbb{F} then $\det A \det B = \det(AB)$.*

Corollary 3.11 *If A and B are matrices associated with an endomorphism of a finite dimensional vector space (with respect to appropriate bases) then $\det A = \det B$.*

The result just proved enables us to make the following definition

Definition 3.12 *If α is an endomorphism of a finite dimensional vector space V associated to a matrix A with respect to some basis we set $\det \alpha = \det A$.*

It is a blemish on our presentation that we define a co-ordinate free quantity $\det \alpha$ via co-ordinates but, since the quantity only exists (in general) for finite dimensional spaces some reference to bases (though not to matrices) is unavoidable. (If the structure of the course permitted it, I would prefer to define determinants via area.)

Corollary 3.13 *If α and β are endomorphisms of a finite dimensional vector space V and if ι is the identity map on V then $\det \alpha \det \beta = \det(\alpha\beta)$ and $\det \iota = 1$.*

Definition 3.14 *If A is an $n \times n$ matrix over \mathbb{F} we write A_{ij} for the determinant of the $(n-1) \times (n-1)$ matrix formed by removing the i th row and j th column from A . The $n \times n$ matrix whose (i, j) th entry is $(-1)^{i+j} A_{ji}$ is called the ‘adjugate matrix’ of A and written as $\text{adj}(A)$.*

So far as we are concerned the only purpose of introducing the adjugate is to prove the last two parts of the following theorem.

Theorem 3.15 (i) *If A is an $n \times n$ matrix then*

$$A \text{adj}(A) = \text{adj}(A)A = (\det A)I.$$

(ii) *If A is an $n \times n$ matrix then A is invertible if and only if $\det A \neq 0$.*

(iii) *An endomorphism α of a finite dimensional vector space V over \mathbb{F} is an automorphism (i.e. an isomorphism of V with itself) if and only if $\det \alpha \neq 0$.*

Before returning, in the next section, to the subject of eigenvalues we look down a couple of by-ways.

Lemma 3.16 *Suppose that in Theorem 2.22 we take $n = m$. Then $(*)$ has a unique solution for all b if and only if $\det A \neq 0$.*

If A is an $n \times m$ matrix with entries a_{ij} the transposed matrix A^T is the $m \times n$ matrix with entries a_{ji} .

Theorem 3.17 (i) If A is an $n \times m$ matrix and B an $m \times p$ matrix then $(AB)^T = B^T A^T$.

(ii) If A is an $n \times n$ matrix then $\det A^T = \det A$.

(iii) If A is an invertible $n \times n$ matrix so is A^T and then $(A^T)^{-1} = (A^{-1})^T$.

(iv) The row and column rank of a matrix are identical.

In Theorem 6.10 we give a ‘natural’ proof of part (iv) by using the idea of an adjoint map.

4 The Characteristic Polynomial

Definition 4.1 (i) If α is an endomorphism of an n dimensional space over \mathbb{F} we define its characteristic polynomial $P_\alpha : \mathbb{F} \rightarrow \mathbb{F}$ by

$$P_\alpha(t) = \det(tI - \alpha).$$

(ii) If A is an $n \times n$ matrix over \mathbb{F} we define its characteristic polynomial $P_A : \mathbb{F} \rightarrow \mathbb{F}$ by

$$P_A(t) = \det(tI - A).$$

Lemma 4.2 If α an endomorphism of a finite dimensional vector space V over \mathbb{F} is associated to a matrix A with respect to some basis then $P_A = p_\alpha$.

Theorem 4.3 Let α be an endomorphism of an n dimensional space V over \mathbb{F} .

(i) P_α is a polynomial of degree exactly n .

(ii) $\lambda \in \mathbb{F}$ is an eigenvalue if and only if $P_\alpha(\lambda) = 0$.

(iii) It is a necessary (but not, if $n \geq 2$, a sufficient) condition for α to be diagonalisable that P_α factorises completely into linear factors over \mathbb{F} .

(iv) It is a sufficient (but not, if $n \geq 2$, a necessary) condition for α to be diagonalisable that P_α has n distinct roots in \mathbb{F} .

(v) If $\mathbb{F} = \mathbb{C}$ then α has at least one eigenvalue (and so has an eigenvector).

The following definition may help illuminate parts (iii) and (iv) of Theorem 4.3.

Definition 4.4 Let α be an endomorphism of a finite dimensional vector space over \mathbb{F} .

(i) We say that an eigenvalue λ has algebraic multiplicity k if $(t - \lambda)^k$ is a factor of $P_\alpha(t)$ but $(t - \lambda)^{k+1}$ is not.

(ii) We say that an eigenvalue λ has geometric multiplicity k if $\alpha - \lambda I$ has nullity k , i.e.

$$\dim\{e \in V : \alpha e = \lambda e\} = k.$$

Lemma 4.5 The geometric multiplicity of an eigenvalue can not exceed its algebraic multiplicity.

Lemma 4.6 Let α be an endomorphism of an finite dimensional vector space over \mathbb{F} . The following statements about α are equivalent.

(i) α is diagonalisable.

(ii) The sum of the geometric multiplicities of its eigenvalues equals the dimension of V .

(iii) (a) the sum of the algebraic multiplicities of its eigenvalues is the dimension of V and (b) the algebraic multiplicity of each eigenvalue equals its geometric multiplicity.

Although we can not always diagonalise an endomorphism of a finite dimensional vector space over \mathbb{C} (see Example 3.4) we can find a basis with respect to which its matrix is triangular.

Theorem 4.7 (i) If α is an endomorphism of an n dimensional vector space V over \mathbb{C} then we can find a basis v_1, v_2, \dots, v_n of V such that

$$\alpha v_j \in \text{span}\{v_1, v_2, \dots, v_j\}.$$

(ii) If A is an $n \times n$ matrix over \mathbb{C} we can find an invertible $n \times n$ matrix P such that $B = P^{-1}AP$ is (upper) triangular (i.e. $b_{ij} = 0$ if $i > j$).

Our method of proof will be induction on the dimension of V starting from the observation that an endomorphism of a complex vector space always has an eigenvector. This method of proof will recur in the course P4. We note a profound difference between the real and complex cases.

Example 4.8 If we work over \mathbb{R} the the matrix

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

has no eigenvectors and so, in particular we can not find an invertible P such that $P^{-1}AP$ is triangular.

We now turn to the study of the characteristic polynomial for its own sake. By direct calculation, or by using Lemma 4.2 we know that $P_{Q^{-1}AQ} = P_A$ whenever Q is invertible. Thus if we write

$$P_A(t) = \sum_{r=0}^n c_r(A)t^r$$

we know that the coefficients $c_r(A)$ are ‘matrix conjugacy class invariants’ in the sense that $c_r(Q^{-1}AQ) = c_r(A)$. We can readily identify three of these invariants: $c_n(A) = 1$ (which is not very interesting), $c_0(A) = (-1)^n \det A$ (which we already knew to be invariant) and $c_{n-1}(A) = -\sum_{i=1}^n a_{ii}$.

Definition 4.9 *If A is an $n \times n$ matrix over \mathbb{F} then we define the trace $\text{tr} A$ of A by*

$$\text{tr} A = \sum_{i=1}^n a_{ii}.$$

The trace of an endomorphism on a finite dimensional space is defined to be the trace of any associated matrix.

In the sense made precise by the next lemma, trace is the only linear matrix conjugacy class invariant.

Lemma 4.10 * *Consider the collection $M_n(\mathbb{F})$ of $n \times n$ matrices over \mathbb{F} .*

(i) $\text{tr} : M_n(\mathbb{F}) \rightarrow \mathbb{F}$ is a linear map.

(ii) If $T : M_n(\mathbb{F}) \rightarrow \mathbb{F}$ is a linear map such that $T(P^{-1}AP) = T(A)$ for all $A \in M_n(\mathbb{F})$ and all invertible $P \in M_n(\mathbb{F})$ then $T = (T(I)n^{-1})\text{tr}$.

If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is smooth the Laplacian of f is the trace of its Hessian.

It should be noted that the characteristic polynomial, informative as it is, does not tell us everything about the associated matrix or endomorphism.

Example 4.11 *The matrices*

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

have the same characteristic polynomial, yet have different ranks. Further one is diagonalisable and the other is not.

If we write out the characteristic polynomial of an endomorphism of a finite dimensional vector space V as

$$P_\alpha(t) = \sum_{r=0}^n c_r(\alpha)t^r$$

we see that we may define $P_\alpha(\beta)$ for any endomorphism β of V by

$$P_\alpha(\beta) = \sum_{r=0}^n c_r(\alpha)\beta^r.$$

Observe that $P_\alpha(\beta)$ is itself an endomorphism. If $P_\alpha(\beta)$ is the zero endomorphism we say that β satisfies the characteristic equation of α . In exactly the same way if A and B are $n \times n$ matrices over \mathbb{F} we can define an $n \times n$ matrix $P_A(B)$. If $P_A(B)$ is the zero matrix we say that B satisfies the characteristic equation of A .

Example 4.12 Consider the two 2×2 matrices

$$A = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

The characteristic polynomial of A is $P_A(t) = t^2 - 2t + 1 = (t - 1)^2$ so $P_A(B)$ is the non-zero 2×2 matrix

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

so that B does not satisfy the characteristic equation of A although $\det(BI - A)$ is the scalar 0.

Lemma 4.13 (The Cayley-Hamilton Theorem for Triangular Matrices) If A is an $n \times n$ triangular matrix over \mathbb{F} with diagonal entries a_{ii} then

$$\det(tI - A) = (t - a_{11})(t - a_{22}) \dots (t - a_{nn})$$

and

$$(A - a_{11}I)(A - a_{22}I) \dots (A - a_{nn}I) = 0.$$

At first sight the result just proved seems very special but in Theorem 4.7 we showed that any endomorphism of a finite dimensional vector space over \mathbb{C} can be represented (with respect to an appropriate basis) by a triangular matrix. We can thus extend the Cayley-Hamilton Theorem to a wide range of cases.

Corollary 4.14 (i) If V is a finite dimensional vector space over \mathbb{C} then any linear map $\alpha : V \rightarrow V$ satisfies its own characteristic equation.

(ii) Any $n \times n$ matrix over \mathbb{C} satisfies its own characteristic equation.

(iii) Any $n \times n$ matrix over \mathbb{R} satisfies its own characteristic equation.

(iv) If V is a finite dimensional vector space over \mathbb{R} then any linear map $\alpha : V \rightarrow V$ satisfies its own characteristic equation.

We bring together the results of the corollary as a theorem.

Theorem 4.15 (*The Cayley-Hamilton Theorem*) (i) Any $n \times n$ matrix over \mathbb{F} satisfies its own characteristic equation.

(ii) If V is a finite dimensional vector space over \mathbb{F} then any linear map $\alpha : V \rightarrow V$ satisfies its own characteristic equation.

There are many different proofs of the Cayley-Hamilton Theorem but (so far as I know) no totally trivial ones. If you invent a new proof, first check it against Example 4.12 and then get your supervisor to check it.

5 Jordan Forms

It is unsatisfactory to leave non-diagonalisable endomorphisms (and matrices) over \mathbb{F} unexamined. In this section we show that they can be fully classified using the Jordan normal form. Although some of the material is more or less explicitly starred some is not and the development is pretty and instructive.

Our first steps are unstarred. Suppose V is a finite dimensional vector space over \mathbb{C} and α is an endomorphism of V . The Cayley Hamilton theorem tells us that there is a monic polynomial P_α with $P_\alpha(\alpha) = 0$. It follows that there must be a monic polynomial of least degree which annihilates α .

Definition 5.1 If V is a finite dimensional vector space over \mathbb{C} and α is an endomorphism of V the monic polynomial Q of least degree such that $Q(\alpha) = 0$ is called the minimal polynomial of α .

Lemma 5.2 Suppose that V is a finite dimensional vector space over \mathbb{C} and α is an endomorphism of V . The minimal polynomial divides (i.e. is a factor of) every polynomial P with $P(\alpha) = 0$. In particular the minimal polynomial divides the characteristic polynomial.

Example 5.3 Suppose that V is a finite dimensional vector space over \mathbb{C} and α is an endomorphism of V which may be diagonalised with associated diagonal matrix D . If the distinct diagonal entries of D are d_1, d_2, \dots, d_k then the minimal polynomial Q_α of α is given by

$$Q_\alpha(t) = \prod_{j=1}^k (t - d_j).$$

In particular the characteristic and minimal polynomials coincide if and only if $k = n$.

The next results may or may not be starred but are sufficiently important that everyone should know them.

Lemma 5.4 *Let $P(t) = \sum_{j=0}^n a_j t^j$ be a polynomial with coefficients in \mathbb{F} . If $P(t) = 0$ for $n + 1$ distinct values of $t \in \mathbb{F}$ then $a_0 = a_1 = \cdots = a_n = 0$. In particular $P(\alpha) = 0$ and $P(A) = 0$ whenever α is an appropriate endomorphism or A an appropriate matrix.*

Theorem 5.5 (Bezout's Theorem for polynomials) *If R_1, R_2, \dots, R_k are polynomials with highest common factor 1 (i.e. if no non-constant polynomial divides all of R_1, R_2, \dots, R_k) then we can find polynomials P_1, P_2, \dots, P_k such that*

$$\sum_{j=1}^k P_j R_j = 1.$$

We now apply the last two results to the minimal polynomial.

Theorem 5.6 *Suppose that V is a finite dimensional vector space over \mathbb{C} and α is an endomorphism of V with minimal polynomial Q . If we write $Q(t) = \prod_{j=1}^k Q_j$ where $Q_j(t) = (t - \lambda_j)^{n_j}$ with $n_j \geq 1$ and the λ_j distinct and set*

$$R_j(t) = \prod_{i \neq j} Q_i$$

we can find polynomials P_1, P_2, \dots, P_k such that

$$\sum_{j=1}^k P_j(\alpha) R_j(\alpha) = \iota.$$

Further, if we write $V_j = R_j(\alpha)V$ the following facts are true.

- (i) $V_j = Q_j(\alpha)^{-1}(0)$.*
- (ii) $\alpha(V_j) \subseteq V_j$.*
- (iii) $V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$,*

Lemma 5.7 *With the notation of the preceding theorem suppose \mathcal{E}_j is a basis for V_j . If $\alpha|_{V_j}$ considered as an endomorphism of V_j (see (ii) in the preceding theorem) has matrix A_j with respect to \mathcal{E}_j , then the matrix of α with respect to the basis $\mathcal{E} = \bigcup_{j=1}^k \mathcal{E}_j$ (see (iii) in the preceding theorem) consists of the square matrices A_j along the diagonal with all other entries 0.*

Thus we have reduced the problem

*Problem** Find a basis for which an endomorphism α has a nice matrix.
to a rather simpler problem

*Problem*** Suppose that the endomorphism α of V has the property that $(\alpha - \lambda\iota)^m = 0$. Find a basis for which an endomorphism α has a nice matrix.

If we observe that whenever α has matrix A with respect to a certain basis then $\alpha - \lambda\iota$ has matrix $A - \lambda I$ we can make the further reduction to *Problem**** Suppose that the endomorphism α on V has the property that $\alpha^m = 0$. Find a basis for which an endomorphism α has a nice matrix.

This suggests the following definition.

Definition 5.8 *An endomorphism α of a vector space is called nilpotent if we can find an $m \geq 0$ with $\alpha^m = 0$.*

To solve the problems just stated we need the notion of a Jordan matrix. We write $J(\lambda, n)$ for the $n \times n$ matrix with λ 's down the diagonal, 1's immediately below and zero every where else so that

$$J(\lambda, n) = \begin{pmatrix} \lambda & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & \lambda & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & \lambda & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & \lambda & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \lambda \end{pmatrix}.$$

We call $J(\lambda, n)$ a Jordan matrix.

We can now solve *Problem****.

Theorem 5.9 *Suppose that V is a finite dimensional vector space over \mathbb{C} and α is a nilpotent endomorphism of V . Then there is basis for V such that α has matrix A (relative to this basis) which consists of zeros except for Jordan matrices $J(0, n_i)$ [$1 \leq i \leq s$]. down the diagonal.*

The only proofs that I know of Theorem 5.9 are hard and in my view the only reason it is included is to show that your lecturers are cleverer than you are. The proof is starred but not the statement of the theorem. You should, however, convince yourselves that the result is plausible.

It is now easy to solve *Problem***.

Lemma 5.10 *Suppose that V is a finite dimensional vector space over \mathbb{C} and α is an endomorphism of V with the property that $(\alpha - \lambda\iota)^m = 0$. Then there is basis for V such that α has matrix A (relative to this basis) which consists of zeros except for Jordan matrices $J(\lambda, n_i)$ [$1 \leq i \leq s$]. down the diagonal.*

We can now solve our original problem, Problem*.

Theorem 5.11 (Jordan Normal Form) *Suppose that V is a finite dimensional vector space over \mathbb{C} and α is an endomorphism of V . Then there is basis for V such that α has matrix A (relative to this basis) which consists of zeros except for Jordan matrices $J(\lambda_i, m_{ij})$ [$1 \leq m_{ij} \leq m_i, 1 \leq i \leq s$]. We may demand that the λ_i be distinct and that $m_{i,1} \geq m_i, 2 \geq \dots$.*

A little thought shows that the Jordan form is unique up to shuffling the diagonal blocks.

The following observation is trivial but worth making.

Lemma 5.12 *With the notation of Theorem 5.9 α has characteristic polynomial*

$$P_\alpha(t) = \prod_{i=1}^s \prod_{j=1}^{m_i} (t - \lambda_i)^{m_{ij}}$$

and minimal polynomial

$$P_\alpha(t) = \prod_{i=1}^s (t - \lambda_i)^{m_{i,1}}.$$

The algebraic multiplicity of λ_i is $\sum_{j=1}^{m_i} m_{ij}$. and the geometric multiplicity is $\max_{1 \leq j \leq m_i} m_{ij}$.

In Course C1 you saw how the Jordan normal form was used to classify the solution of two simultaneous first order differential equations in two unknowns. The extension to n simultaneous first order differential equations in n unknowns is obvious.

6 Dual Spaces

We saw in Theorem 2.8 that the linear maps from one vector space U to another vector space V form a vector space $L(U, V)$. In the two previous sections we investigated the special case when $U = V$. Now we look at the case when V is one dimensional.

Definition 6.1 *If U is a vector space over \mathbb{F} the vector space of linear maps $u' : U \rightarrow \mathbb{F}$ is called the dual space of U and denoted by U' .*

Thus for example the trace map is in the dual space of $M_n(\mathbb{F})$.

Example 6.2 Let $C^\infty(\mathbb{R})$ be the space of infinitely differentiable functions $f : \mathbb{R} \rightarrow \mathbb{R}$. If we fix $x \in \mathbb{R}$ and set

$$J(f) = \int_{-1}^1 f(t)dt, \quad \delta_x(f) = f(x), \quad \delta'_x(f) = f'(x)$$

then J , δ_x and δ'_x are all in the dual space of $C^\infty(\mathbb{R})$.

As the example indicates analysts are very interested in objects which belong to dual spaces whilst algebraists are interested in dual objects in general. However, at this level, we can only say interesting things about the duals of finite dimensional spaces.

Lemma 6.3 Suppose that V is a vector space over \mathbb{F} with basis e_1, e_2, \dots, e_n . Then the dual space V' has the same dimension as V and may be given a basis (the so called dual basis) E_1, E_2, \dots, E_n with $E_i(e_j) = \delta_{ij}$.

Lemma 6.4 Let V be a finite dimensional vector space. If we write

$$\Phi(v)(v') = v'(v)$$

for all $v \in V$, $v' \in V'$ then $\Phi : V \rightarrow V''$ is an isomorphism.

We already knew from Lemma 6.3 that $\dim V = \dim V' = \dim V''$ but Φ gives us a ‘natural isomorphism’ defined without reference to some particularly chosen basis. The ‘standard convention’ is to write $v = \Phi(v)$, i.e. to identify V and V'' via Φ .

Definition 6.5 If U is a non-empty subset of a vector space V then we define the annihilator U° of U to be the subset of V' given by

$$U^\circ = \{v' \in V' : v'(u) = 0 \text{ for all } u \in U\}.$$

Lemma 6.6 If U is a non-empty subset of a vector space V then U is a subspace of V' . If, further, V is finite dimensional and U is a subspace of V then

- (i) $\dim U + \dim U^\circ = \dim V$.
- (ii) $\Phi(U) = U^{\circ\circ}$ so, using the standard convention, $U^{\circ\circ} = U$.

We conclude the section and the course by looking at the notion of the adjoint of a linear map.

Lemma 6.7 *If U and V are vector spaces over \mathbb{F} and $\alpha \in L(U, V)$ then the equation*

$$\alpha'(v')(u) = v'(\alpha u)$$

for all $v' \in V'$, $u \in U$ defines an $\alpha' \in L(V', U')$. (We call α' the dual, or the adjoint, of α).

Lemma 6.8 *If U and V are finite dimensional vector spaces and α has matrix A with respect to given bases of U and V then α' has matrix A^T with respect to the dual bases.*

Lemma 6.9 *If U , V and W are vector spaces over \mathbb{F} and $\alpha \in L(U, V)$, $\beta \in L(V, W)$ then $(\beta\alpha)' = \alpha'\beta'$.*

If we choose appropriate bases we recover the matricial formula $(BA)^T = A^T B^T$ for matrices of appropriate sizes.

Theorem 6.10 *If U and V are finite dimensional vector spaces. If we adopt the standard convention of identifying U'' with U and V'' with V then, if $\alpha \in L(U, V)$,*

(i) $\alpha'' = \alpha$.

(ii) $(\alpha U)^\circ = (\alpha')^{-1}(0)$.

(iii) $r(\alpha) = r(\alpha')$

(iv) The row rank and the column rank of any matrix are equal.

We have thus fulfilled our promise to prove the equivalence of row and column rank in a natural context.

Lemma 6.11 *If α is an endomorphism of a finite dimensional vector space V the α' is an endomorphism of V' and $\det \alpha' = \det \alpha$.*