LINEAR ALGEBRA

SIMON WADSLEY

Contents

Lecture 1

1. Vector spaces

Linear algebra can be summarised as the study of vector spaces and linear maps between them. This is a second 'first course' in Linear Algebra. That is to say, we will define everything we use but will assume some familiarity with the concepts (picked up from the IA course Vectors and Matrices for example).

1.1. **Definitions and examples.**

Examples.

(1) For each non-negative integer n, the set $\mathbf{R}^n$ of column vectors of length n with real entries is a vector space (over $\mathbf{R}$). An $(m \times n)$-matrix A with real entries can be viewed as a linear map $\mathbf{R}^n \to \mathbf{R}^m$ via $v \mapsto Av$. In fact, as we will see, every linear map from $\mathbf{R}^n \to \mathbf{R}^m$ is of this form.

(2) Let X be a set and $\mathbf{R}^X := \{f \colon X \to \mathbf{R}\}$ be equipped with an addition given by $(f+g)(x) := f(x) + g(x)$ and a multiplication by scalars (in $\mathbf{R}$) given by $(\lambda f)(x) = \lambda(f(x))$. Then $\mathbf{R}^X$ is a vector space (over $\mathbf{R}$).

(3) If $[a,b]$ is a closed interval in $\mathbf{R}$ then $C([a,b],\mathbf{R}) := \{f \in \mathbf{R}^{[a,b]} \mid f \text{ is continuous}\}$ is an $\mathbf{R}$-vector space by restricting the operations on $\mathbf{R}^{[a,b]}$. Similarly

$$C^\infty([a,b],\mathbf{R}) := \{f \in C([a,b],\mathbf{R}) \mid f \text{ is infinitely differentiable}\}$$

is an $\mathbf{R}$-vector space.

(4) The set of $(m \times n)$-matrices with real entries is a vector space over $\mathbf{R}$.

Convention. In this course we will use $\mathbf{F}$ to denote either $\mathbf{R}$ or $\mathbf{C}$. Most of the results will be true for any field $\mathbf{F}$; but since general fields are not officially defined until Groups, Rings and Modules next term we follow the schedules in not addressing that.

What do our examples of vector spaces above have in common? In each case we have a notion of addition of 'vectors' and scalar multiplication of 'vectors' by elements in $\mathbf{R}$.

Definition. An $\mathbf{F}$-*vector space* is an abelian group $(V,+)$ equipped with a function $\mathbf{F} \times V \to V$; $(\lambda, v) \mapsto \lambda v$ such that

(a) $\lambda(\mu v) = (\lambda\mu)v$ for all $\lambda, \mu \in \mathbf{F}$ and $v \in V$;
(b) $\lambda(u + v) = \lambda u + \lambda v$ for all $\lambda \in \mathbf{F}$ and $u, v \in V$;
(c) $(\lambda + \mu)v = \lambda v + \mu v$ for all $\lambda, \mu \in \mathbf{F}$ and $v \in V$;
(d) $1v = v$ for all $v \in V$.

Note that this means that we can add, subtract and rescale elements in a vector space and these operations behave in the ways that we are used to. Note also that in general a vector space does not come equipped with notions of length or of angle. We will discuss how to recover these at the end of the course. At that point particular properties of the field $\mathbf{F}$ will be important.

Convention. We will always write 0 to denote the additive identity of a vector space V. By slight abuse of notation we will also write 0 to denote the vector space $\{0\}$.

Exercise.

(1) Convince yourself that all the vector spaces mentioned thus far do indeed satisfy the axioms for a vector space.
(2) Show that for any v in any vector space V, $0v = 0$ and $(-1)v = -v$

Definition. Suppose that V is a vector space over $\mathbf{F}$. A subset $U \subset V$ is a *(linear) subspace* if

(a) for all $u_1, u_2 \in U$, $u_1 + u_2 \in U$;
(b) for all $\lambda \in \mathbf{F}$ and $u \in U$, $\lambda u \in U$;
(c) $0 \in U$.

Remarks.

(1) It is straightforward to see that $U \subset V$ is a subspace if and only if $U \neq \emptyset$ and $\lambda u_1 + \mu u_2 \in U$ for all $u_1, u_2 \in U$ and $\lambda, \mu \in F$.
(2) If U is a subspace of V then U is a vector space under the inherited operations.

Examples.

(1) $\left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbf{R}^3 \colon x_1 + x_2 + x_3 = t \right\}$ is a subspace of $\mathbf{R}^3$ if and only if $t = 0$.
(2) Let X be a set. We define the *support* of a function $f \colon X \to \mathbf{F}$ to be

$$\mathrm{supp} f := \{x \in X : f(x) \neq 0\}.$$

Then $\{f \in \mathbf{F}^X : |\mathrm{supp} f| < \infty\}$ is a subspace of $\mathbf{F}^X$.

Definition. Let V be a vector space over $\mathbf{F}$ and $S \subset V$ a subset of V. Then the *span of S in V*,

$$\langle S \rangle := \left\{ \sum_{i=1}^{n} \lambda_i s_i : \lambda_i \in \mathbf{F}, s_i \in S, n \geqslant 0 \right\}$$

Remark. For any subset $S \subset V$, $\langle S \rangle$ is the smallest subspace of V containing S.

Example. Suppose that V is $\mathbf{R}^3$.

$$\text{If } S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \right\} \text{ then } \langle S \rangle = \left\{ \begin{pmatrix} a \\ b \\ b \end{pmatrix} : a, b \in \mathbf{R} \right\}.$$

Note also that every subset of S of order 2 has the same span as S.

Example. Let X be a set and for each $x \in X$, define $\delta_x \colon X \to \mathbf{F}$ by

$$\delta_x(y) = \begin{cases} 1 \text{ if } y = x \\ 0 \text{ if } y \neq x. \end{cases}$$

Then $\langle \delta_x : x \in X \rangle = \{f \in \mathbf{F}^X : |\mathrm{supp} f| < \infty\}$.

Definition. Suppose that U and W are subspaces of a vector space V over $\mathbf{F}$. Then the *sum* of U and W is the set

$$U + W := \{u + w : u \in U, w \in W\}.$$

Proposition. *If U and W are subspaces of a vector space V over $\mathbf{F}$ then $U \cap W$ and $U + W$ are also subspaces of V.*

Proof. Certainly both $U \cap W$ and $U + W$ contain 0. Suppose that $v_1, v_2 \in U \cap W$, $u_1, u_2 \in U$, $w_1, w_2 \in W$,and $\lambda, \mu \in \mathbf{F}$. Then $\lambda v_1 + \mu v_2 \in U \cap W$ and

$$\lambda(u_1 + w_1) + \mu(u_2 + w_2) = (\lambda u_1 + \mu u_2) + (\lambda w_1 + \mu w_2) \in U + W.$$

So $U \cap W$ and $U + W$ are subspaces of V. $\square$

***Quotient spaces*.** Suppose that V is a vector space over $\mathbf{F}$ and U is a subspace of V. Then the quotient group V/U can be made into a vector space over $\mathbf{F}$ by definining

$$\lambda(v + U) = (\lambda v) + U$$

for $\lambda \in \mathbf{F}$ and $v \in V$.

Exercise. Justify the claim that this makes V/U into a vector space over $\mathbf{F}$.

Lecture 2

1.2. **Linear independence, bases and the Steinitz exchange lemma.**

Definition. Let V be a vector space over $\mathbf{F}$ and $S \subset V$.

(a) We say that S *spans* V if $V = \langle S \rangle$.
(b) We say that S is *linearly independent (LI)* if, whenever

$$\sum_{i=1}^{n} \lambda_i s_i = 0$$

with $\lambda_i \in \mathbf{F}$, and s_i distinct elements of S, it follows that $\lambda_i = 0$ for all i. If S is not linearly independent then we say that S is *linearly dependent (LD)*.
(c) We say that S is a *basis* for V if S spans and is linearly independent.

Example. Suppose that V is $\mathbf{R}^3$ and $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \right\}$. Then S is linearly dependent since $1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + (-1) \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = 0$. Moreover S does not span V since $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ is not in $\langle S \rangle$. However, every subset of S of order 2 is linearly independent and forms a basis for $\langle S \rangle$.

Remark. Note that no linearly independent set can contain the zero vector since $1 \cdot 0 = 0$.

Convention. The span of the empty set $\langle \emptyset \rangle$ is the zero subspace 0. Thus the empty set is a basis of 0. One may consider this to not be so much a convention as the only reasonable interpretation of the definitions of span, linearly independent and basis in this case.

Lemma. *A subset S of a vector space V over $\mathbf{F}$ is linearly dependent if and only if there exist $s_0, s_1, \ldots, s_n \in S$ distinct and $\lambda_1, \ldots, \lambda_n \in \mathbf{F}$ such that $s_0 = \sum_{i=1}^{n} \lambda_i s_i$.*

Proof. Suppose that S is linearly dependent so that $\sum \lambda_i s_i = 0$ for some $s_i \in S$ distinct and $\lambda_i \in \mathbf{F}$ with $\lambda_j \neq 0$ say. Then

$$s_j = \sum_{i \neq j} \frac{-\lambda_i}{\lambda_j} s_i.$$

Conversely, if $s_0 = \sum_{i=1}^n \lambda_i s_i$ then $(-1)s_0 + \sum_{i=1}^n \lambda_i s_i = 0$. □

Proposition. *Let V be a vector space over $\mathbf{F}$. Then $\{e_1, \ldots, e_n\}$ is a basis for V if and only if every element $v \in V$ can be written uniquely as $v = \sum_{i=1}^n \lambda_i e_i$ with $\lambda_i \in \mathbf{F}$.*

Proof. First we observe that by definition $\{e_1, \ldots, e_n\}$ spans V if and only if every element v of V can be written in at least one way as $v = \sum \lambda_i e_i$ with $\lambda_i \in \mathbf{F}$.

So it suffices to show that $\{e_1, \ldots, e_n\}$ is linearly independent if and only if there is at most one such expression for every $v \in V$.

Suppose that $\{e_1, \ldots, e_n\}$ is linearly independent and $v = \sum \lambda_i e_i = \sum \mu_i e_i$ with $\lambda_i, \mu_i \in \mathbf{F}$. Then, $\sum(\lambda_i - \mu_i)e_i = 0$. Thus by definition of linear independence, $\lambda_i - \mu_i = 0$ for $i = 1, \ldots, n$ and so $\lambda_i = \mu_i$ for all i.

Conversely if $\{e_1, \ldots, e_n\}$ is linearly dependent then we can write

$$\sum \lambda_i e_i = 0 = \sum 0 e_i$$

for some $\lambda_i \in \mathbf{F}$ not all zero. Thus there are two ways to write 0 as an $\mathbf{F}$-linear combination of the e_i. □

The following result is necessary for a good notion of dimension for vector spaces.

Theorem (Steinitz exchange lemma)**.** *Let V be a vector space over $\mathbf{F}$. Suppose that $S = \{e_1, \ldots, e_n\}$ is a linearly independent subset of V and $T \subset V$ spans V. Then there is a subset T' of T of order n such that $(T \backslash T') \cup S$ spans V. In particular $n \leqslant |T|$.*

This is sometimes stated as follows (with the assumption that T is finite).

Corollary. *If $\{e_1, \ldots, e_n\} \subset V$ is linearly independent and $\{f_1, \ldots, f_m\}$ spans V. Then $n \leqslant m$ and, possibly after reordering the f_i, $\{e_1, \ldots, e_n, f_{n+1}, \ldots, f_m\}$ spans V.*

We prove the theorem by replacing elements of T by elements of S one by one.

Proof of the Theorem. Suppose that we've already found a subset T'_r of T of order $0 \leqslant r < n$ such that $T_r := (T \backslash T'_r) \cup \{e_1, \ldots, e_r\}$ spans V. Then we can write

$$e_{r+1} = \sum_{i=1}^k \lambda_i t_i$$

with $\lambda_i \in \mathbf{F}$ and $t_i \in T_r$. Since $\{e_1, \ldots, e_{r+1}\}$ is linearly independent there must be some $1 \leqslant j \leqslant k$ such that $\lambda_j \neq 0$ and $t_j \notin \{e_1, \ldots, e_r\}$. Let $T'_{r+1} = T'_r \cup \{t_j\}$ and

$$T_{r+1} = (T \backslash T'_{r+1}) \cup \{e_1, \ldots, e_{r+1}\} = (T_r \backslash \{t_j\}) \cup \{e_{r+1}\}$$

.

Now

$$t_j = \frac{1}{\lambda_j} e_{r+1} - \sum_{i \neq j} \frac{\lambda_i}{\lambda_j} t_i,$$

so $t_j \in \langle T_{r+1}\rangle$ and $\langle T_{r+1}\rangle = \langle T_{r+1} \cup \{t_j\}\rangle \supset \langle T_r\rangle = V$.

Now we can inductively construct $T' = T'_n$ with the required properties. $\square$

Corollary. *Let V be a vector space with a basis of order n.*

(a) Every basis of V has order n.
(b) Every basis of a subspace U of V has order at most n.
(c) Any n LI vectors in V form a basis for V.
(d) Any n vectors in V that span V form a basis for V.
(e) Any set of linearly independent vectors in V can be extended to a basis.

Proof. Suppose that $S = \{e_1, \dots, e_n\}$ is a basis for V.

(a) Suppose that T is another basis of V. Since S spans V and any finite subset of T is linearly independent $|T| \leqslant n$. Since T spans and S is linearly independent $|T| \geqslant n$. Thus $|T| = n$ as required.

(b) Suppose T is a basis for U. Since T is a linearly independent subset of V, $|T| \leqslant n$.

(c) Suppose T is a LI subset of V of order n. If T did not span we could choose $v \in V \backslash \langle T\rangle$. Then $T \cup \{v\}$ is a LI subset of V of order $n+1$, a contradiction.

(d) Suppose T is spans V and has order n. If T were LD we could find $t_0, t_1, \dots, t_m$ in T distinct such that $t_0 = \sum_{i=1}^m \lambda_i t_i$ with $\lambda_i \in \mathbf{F}$. Thus $V = \langle T\rangle = \langle T\backslash\{t_0\}\rangle$ so $T\backslash\{t_0\}$ is a spanning set for V of order $n-1$, a contradiction.

(e) Let $T = \{t_1, \dots, t_m\}$ be a linearly independent subset of V. Since S spans V we can find $s_1, \dots, s_m$ in S such that $(S\backslash\{s_1, \dots, s_m\}) \cup T$ spans V. Since this set has order (at most) n it is a basis containing T. $\square$

Lecture 3

Definition. If a vector space V over $\mathbf{F}$ has a finite basis S then we say that V is *finite dimensional* (or *f. d.*). Moreover, we define the *dimension* of V by

$$\dim_{\mathbf{F}} V = \dim V = |S|.$$

If V does not have a finite basis then we will say that V is *infinite dimensional*.

Lemma. *If V is f.d. and $U \subsetneq V$ is a proper subspace then U is also f.d.. Moreover, $\dim U < \dim V$.*

Proof. Let $S \subset U$ be a LI subset of U of maximal possible size. Then $|S| \leqslant \dim V$ (by the last Corollary).

Suppose that $v \in V\backslash\langle S\rangle$ and $\lambda_0 v + \sum_{i=1}^m \lambda_i s_i{=}0$ with $\lambda_0 \dots, \lambda_m \in \mathbf{F}$, and $s_1, \dots, s_m$ is S distinct. Then $\lambda_0 = 0$ since $v \notin \langle S\rangle$. So $\lambda_1, \dots, \lambda_m = 0$ since S is LI. Thus $S \cup \{v\}$ is LI for every $v \in V\backslash\langle S\rangle$. In particular $U = \langle S\rangle$, else S does not have maximal size. Moreover since $U \neq V$, there is some $v \in V\backslash\langle S\rangle$ and $|S \cup \{v\}|$ is a LI subset of order $|S|+1$. So $|S| < \dim V$ as required. $\square$

Remarks.

(1) By the last corollary the dimension of a finite dimensional space V does not depend on the choice of basis S. However the dimension *does* depend on $\mathbf{F}$. For example $\mathbf{C}$ has dimension 1 viewed as a vector space over $\mathbf{C}$ (since $\{1\}$ is a basis) but dimension 2 viewed as a vector space over $\mathbf{R}$ (since $\{1, i\}$ is a basis).

(2) If we wanted to be more precise then we could define the dimension of an infinite dimensional space to be the cardinality of any basis for V. But we have not proven enough to see that this would be well-defined; in fact there are no problems.

Proposition. *Let U and W be subspaces of a finite dimensional vector space V over $\mathbf{F}$. Then*

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

Proof. Since dimension is defined in terms of bases and we have no way to compute it at present except by finding bases and counting the number of elements we must find suitable bases. The key idea is to be careful about how we choose our bases.

Slogan When choosing bases always choose the right basis for the job.

Let $R:=\{v_1,\ldots,v_r\}$ be a basis for $U\cap W$. Since $U\cap W$ is a subspace of U we can extend R to a basis $S:=\{v_1,\ldots,v_r,u_{r+1},\ldots,u_s\}$ for U. Similary we can extend R to a basis $T:=\{v_1,\ldots,v_r,w_{r+1},\ldots,w_t\}$ for W. We claim that $X:=S\cup T$ is a basis for $U+W$. This will suffice, since then

$$\dim(U+W)=|X|=s+t-r=\dim U+\dim W-\dim(U\cap W).$$

Suppose $u+w\in U+W$ with $u\in U$ and $w\in W$. Then $u\in\langle S\rangle$ and $w\in\langle T\rangle$. Thus $U+W$ is contained in the span of $X=S\cup T$. It is clear that $\langle X\rangle\subset U+W$ so X does span $U+W$ and it now suffices to show that X is linearly independent. Suppose that

$$\sum_{i=1}^{r}\lambda_i v_i+\sum_{j=r+1}^{s}\mu_j u_j+\sum_{k=r+1}^{t}\nu_k w_k=0.$$

Then we can write $\sum\mu_j u_j=-\sum\lambda_i v_i-\sum\nu_k w_k\in U\cap W$. Since the R spans $U\cap W$ and T is linearly independent it follows that all the ν_k are zero. Then $\sum\lambda_i v_i+\sum\mu_j u_j=0$ and so all the λ_i and μ_j are also zero since S is linearly independent. □

Lemma. *If $S\subset V$ is a finite spanning set then S contains a basis for V.*

Proof. By induction on —S—. If S is LI we're done. Otherwise there are $s_0,\ldots,s_n\in S$ and $\lambda_1,\ldots,\lambda_n\in\mathbf{F}$ such that $s_0=\sum_{i=1}^n\lambda_i s_i$. Thus $\langle S\rangle=\langle S\backslash\{s_0\}\rangle$. By the induction hypothesis $S\backslash\{s_0\}$ contains a basis. □

Exercise (non-examinable). Show that if V is a finite dimensional vector space over $\mathbf{F}$ and U is a subspace then

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

Hint. Show that if $\{u_1,\ldots,u_m\}$ is a basis for U and $\{v_1+U,\ldots,v_n+U\}$ is a basis for V/U then $\{u_1,\ldots,u_m,v_1,\ldots,v_n\}$ is a basis for V.

1.3. **Direct sum.** There are two related notions of direct sum of vector spaces and the distinction between them can often cause confusion to newcomers to the subject. The first is sometimes known as the *internal* direct sum and the latter as the *external* direct sum. However it is common to gloss over the difference between them.

Definition. Suppose that V is a vector space over $\mathbf{F}$ and U and W are subspaces of V. Recall that *sum* of U and W is defined to be

$$U + W = \{u + w : u \in U, w \in W\}.$$

We say that V is the *(internal) direct sum* of U and W, written $V = U \oplus W$, if $V = U + W$ and $U \cap W = 0$. Equivalently $V = U \oplus V$ if every element $v \in V$ can be written uniquely as $u + w$ with $u \in U$ and $w \in W$.

We also say that U and W are *complementary subspaces* in V.

Example. Suppose that $V = \mathbf{R}^3$ and

$$U = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : x_1 + x_2 + x_3 = 0 \right\}, W_1 = \left\langle \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\rangle \text{ and } W_2 = \left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\rangle$$

then $V = U \oplus W_1 = U \oplus W_2$.

Note in particular that U does not have only one complementary subspace in V.

Definition. Given any two vector spaces U and W over $\mathbf{F}$ the *(external) direct sum* $U \oplus W$ of U and W is defined to be the set of pairs

$$\{(u, w) : u \in U, w \in W\}$$

with addition given by

$$(u_1, w_1) + (u_2, w_2) = (u_1 + u_2, w_1 + w_2)$$

and scalar multiplication given by

$$\lambda(u, w) = (\lambda u, \lambda w).$$

Exercise. Show that $U \oplus W$ is a vector space over $\mathbf{F}$ with the given operations and that it is the internal direct sum of its subspaces

$$\{(u, 0) : u \in U\} \text{ and } \{(0, w) : w \in W\}.$$

More generally we can make the following definitions.

Definition. If $U_1, \ldots, U_n$ are subspaces of V then V is the *(internal) direct sum of* $U_1, \ldots, U_n$ written

$$V = U_1 \oplus \cdots \oplus U_n = \bigoplus_{i=1}^{n} U_i$$

if every element v of V can be written uniquely as $v = \sum_{i=1}^{n} u_i$ with $u_i \in U_i$.

Definition. If $U_1, \ldots, U_n$ are any vector spaces over $\mathbf{F}$ their *(external) direct sum* is the vector space

$$\bigoplus_{i=1}^{n} U_i := \{(u_1, \ldots, u_n) \mid u_i \in U_i\}$$

with natural coordinate-wise operations.

From now on we will drop the adjectives 'internal' and 'external' from 'direct sum'.

Lecture 4

2. Linear maps

2.1. Definitions and examples.

Definition. Suppose that U and V are vector spaces over a field $\mathbf{F}$. Then a function $\alpha\colon U \to V$ is a *linear map* if

(a) $\alpha(u_1 + u_2) = \alpha(u_1) + \alpha(u_2)$ for all $u_1, u_2 \in U$;
(b) $\alpha(\lambda u) = \lambda\alpha(u)$ for all $u \in U$ and $\lambda \in \mathbf{F}$.

Notation. We write $\mathcal{L}(U, V)$ for the set of linear maps $U \to V$.

Remarks.

(1) We can combine the two parts of the definition into one as: α is linear if and only if $\alpha(\lambda u_1 + \mu u_2) = \lambda\alpha(u_1) + \mu\alpha(u_2)$ for all $\lambda, \mu \in \mathbf{F}$ and $u_1, u_2 \in U$. Linear maps should be viewed as functions between vector spaces that respect their structure as vector spaces.
(2) If α is linear map then α is a homomorphism of the underlying abelian groups. In particular $\alpha(0) = 0$.
(3) If we want to stress the field $\mathbf{F}$ then we will say a map is $\mathbf{F}$-linear. For example, complex conjugation defines an $\mathbf{R}$-linear map from $\mathbf{C}$ to $\mathbf{C}$ but it is not $\mathbf{C}$-linear.

Examples.

(1) Let A be an $n \times m$ matrix with coefficients in $\mathbf{F}$. Then $\alpha\colon \mathbf{F}^m \to \mathbf{F}^n$; $\alpha(v) = Av$ is a linear map.

To see this let $\lambda, \mu \in \mathbf{F}$ and $u, v \in \mathbf{F}^m$. As usual, let A_{ij} denote the ijth entry of A and u_j, (resp. v_j) the jth coordinate of u (resp. v). Then for $1 \leqslant i \leqslant n$,

$$(\alpha(\lambda u + \mu v))_i = \sum_{j=1}^{m} A_{ij}(\lambda u_j + \mu v_j) = \lambda\alpha(u)_i + \mu\alpha(v)_i$$

so $\alpha(\lambda u + \mu v) = \lambda\alpha(u) + \mu\alpha(v)$ as required.
(2) If X is any set and $g \in \mathbf{F}^X$ then $g\colon \mathbf{F}^X \to \mathbf{F}^X$; $(gf)(x) := g(x)f(x)$ for $x \in X$ is linear.
(3) For all $x \in [a, b]$, $\delta_x\colon C([a, b], \mathbf{R}) \to \mathbf{R}$; $f \mapsto f(x)$ is linear.
(4) $I\colon C([a, b], \mathbf{R}) \to C([a, b], \mathbf{R})$; $I(f)(x) = \int_a^x f(t)\,dt$ is linear.
(5) $D\colon C^\infty([a, b], \mathbf{R}) \to C^\infty([a, b], \mathbf{R})$; $(Df)(t) = f'(t)$ is linear.
(6) If $\alpha, \beta\colon U \to V$ are linear and $\lambda \in \mathbf{F}$ then $\alpha + \beta\colon U \to V$ given by $(\alpha + \beta)(u) = \alpha(u) + \beta(u)$ and $\lambda\alpha\colon U \to V$ given by $(\lambda\alpha)(u) = \lambda(\alpha(u))$ are linear. In this way $\mathcal{L}(U, V)$ is a vector space over $\mathbf{F}$.

Definition. We say that a linear map $\alpha\colon U \to V$ is an *isomorphism* if there is a linear map $\beta\colon V \to U$ such that $\beta\alpha = \mathrm{id}_U$ and $\alpha\beta = \mathrm{id}_V$.

Lemma. *Suppose that U and V are vector spaces over $\mathbf{F}$. A linear map $\alpha\colon U \to V$ is an isomorphism if and only if α is a bijection.*

Proof. Certainly an isomorphism $\alpha\colon U \to V$ is a bijection since it has an inverse as a function between the underlying sets U and V. Suppose that $\alpha\colon U \to V$ is a linear bijection and let $\beta\colon V \to U$ be its inverse as a function. We must show that β is also linear. Let $\lambda, \mu \in \mathbf{F}$ and $v_1, v_2 \in V$. Then

$$\alpha\beta\left(\lambda v_1 + \mu v_2\right) = \lambda\alpha\beta(v_1) + \mu\alpha\beta(v_2) = \alpha\left(\lambda\beta(v_1) + \mu\beta(v_2)\right).$$

Since α is injective it follows that β is linear as required. $\square$

Definition. Suppose that $\alpha\colon U \to V$ is a linear map.

- The *image* of α, $\operatorname{Im}\alpha := \{\alpha(u) : u \in U\}$.
- The *kernel* of α, $\ker\alpha := \{u \in U : \alpha(u) = 0\}$.

Examples.

(1) Let A be an $n \times m$-matrix with coefficients in $\mathbf{F}$ and let $\alpha\colon \mathbf{F}^m \to \mathbf{F}^n$ be the linear map defined by $x \mapsto Ax$. Then the system of equations

$$\sum_{j=1}^{m} A_{ij}x_j = b_i; \qquad 1 \leqslant i \leqslant n$$

has a solution if and only if $\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in \operatorname{Im}\alpha$. The kernel of α consists of the solutions $\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$ to the homogeneous equations

$$\sum_{j=1}^{m} A_{ij}x_j = 0; \qquad 1 \leqslant i \leqslant n$$

(2) Let $\beta\colon C^\infty(\mathbf{R},\mathbf{R}) \to C^\infty(\mathbf{R},\mathbf{R})$ be given by

$$\beta(f)(t) = f''(t) + p(t)f'(t) + q(t)f(t)$$

for some $p, q \in C^\infty(\mathbf{R},\mathbf{R})$. A function $g \in C^\infty(\mathbf{R},\mathbf{R})$ is in the image of β precisely if

$$f''(t) + p(t)f'(t) + q(t) = g(t)$$

has a solution in $C^\infty(\mathbf{R},\mathbf{R})$. Moreover, $\ker\beta$ consists of the solutions to the differential equation

$$f''(t) + p(t)f'(t) + q(t)f(t) = 0$$

in $C^\infty(\mathbf{R},\mathbf{R})$.

Note that α is injective if and only if $\ker\alpha = 0$ and that α is surjective if and only if $\operatorname{Im}\alpha = V$.

Proposition. *Suppose that $\alpha\colon U \to V$ is an $\mathbf{F}$-linear map.*

(a) If α is injective and $S \subset U$ is linearly independent then $\alpha(S) \subset V$ is linearly independent.
(b) If α is surjective and $S \subset U$ spans U then $\alpha(S)$ spans V.
(c) If α is an isomorphism and S is a basis then $\alpha(S)$ is a basis.

Proof. (a) Suppose α is injective, $S \subset U$ and $\alpha(S)$ is linearly dependent. Then there are $s_0, \ldots, s_n \in S$ distinct and $\lambda_1, \ldots, \lambda_n \in \mathbf{F}$ such that

$$\alpha(s_0) = \sum \lambda_i \alpha(s_i) = \alpha\left(\sum_{i=1}^{n} \lambda_i s_i\right).$$

Since α is injective it follows that $s_0 = \sum_1^n \lambda_i s_i$ and S is LD.

(b) Now suppose that α is surjective, $S \subset U$ spans U and let v in V. There is $u \in U$ such that $\alpha(u) = v$ and there are $s_1, \ldots, s_n \in S$ and $\lambda_1, \ldots, \lambda_n \in \mathbf{F}$ such that $\sum \lambda_i s_i = u$. Then $\sum \lambda_i \alpha(s_i) = v$. Thus $\alpha(S)$ spans V.

(c) Follows immediately from (a) and (b). $\square$

Corollary. *If two finite dimensional vector spaces are isomorphic then they have the same dimension.*

Proof. If $\alpha\colon U \to V$ is an isomorphism and S is a finite basis for U then $\alpha(S)$ is a basis of V by the proposition. Since α is an injection $|S| = |\alpha(S)|$. $\square$

Proposition. *Suppose that V is a vector space over $\mathbf{F}$ of dimension $n < \infty$. Writing $e_1, \ldots, e_n$ for the standard basis for $\mathbf{F}^n$, there is a bijection Φ between the set of isomorphisms $\mathbf{F}^n \to V$ and the set of (ordered) bases for V that sends the isomorphism $\alpha\colon \mathbf{F}^n \to V$ to the (ordered) basis $\langle \alpha(e_1), \ldots, \alpha(e_n) \rangle$.*

Proof. That the map Φ is well-defined follows immediately from part (c) of the last Proposition.

If $\Phi(\alpha) = \Phi(\beta)$ then

$$\alpha\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) = \sum_{i=1}^n x_i \alpha(e_i) = \sum_{i=1}^n x_i \beta(e_i) = \beta\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right)$$

for all $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{F}^n$ so $\alpha = \beta$ and Φ is injective.

Suppose now that $\langle v_1, \ldots, v_n \rangle$ is an ordered basis for V and define $\alpha\colon \mathbf{F}^n \to V$ by

$$\alpha\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) = \sum_{i=1}^n x_i v_i.$$

Then α is injective since $v_1, \ldots, v_n$ are LI and α is surjective since $v_1, \ldots, v_n$ span V and α is easily seen to be linear. Thus α is an isomorphism such that $\Phi(\alpha) = \langle v_1, \ldots, v_n \rangle$ and Φ is surjective as required. $\square$

Thus choosing a basis for an n-dimensional vector space V corresponds to choosing an identification of V with $\mathbf{F}^n$.

Lecture 5

2.2. Linear maps and matrices.

Proposition. *Suppose that U and V are vector spaces over $\mathbf{F}$ and $S := \{e_1, \ldots, e_n\}$ is a basis for U. Then every function $f\colon S \to V$ extends uniquely to a linear map $\alpha\colon U \to V$.*

Slogan To define a linear map it suffices to specify its values on a basis.

Proof. First we prove uniqueness: suppose that $f\colon S \to V$ and α andd β are two linear maps $U \to V$ extending f. Let $u \in U$ so that $u = \sum u_i e_i$ for some $u_i \in \mathbf{F}$. Then

$$\alpha(u) = \alpha\left(\sum_{i=1}^n u_i e_i\right) = \sum_{i=1}^n u_i \alpha(e_i).$$

Similarly, $\beta(u) = \sum_1^n u_i \beta(e_i)$. Since $\alpha(e_i) = f(e_i) = \beta(e_i)$ for each $1 \leqslant i \leqslant n$ we see that $\alpha(u) = \beta(u)$ for all $u \in U$ and so $\alpha = \beta$.

That argument also shows us how to construct a linear map α that extends f. Every $u \in U$ can be written uniquely as $u = \sum_{i=1}^n u_i e_i$ with $u_i \in \mathbf{F}$. Thus we can define $\alpha(u) = \sum u_i f(e_i)$ without ambiguity. Certainly α extends f so it remains to show that α is linear. So we compute for $u = \sum u_i e_i$ and $v = \sum v_i e_i$,

$$\begin{aligned} \alpha(\lambda u + \mu v) &= \alpha\left(\sum_{i=1}^n (\lambda u_i + \mu v_i) e_i\right) \\ &= \sum_{i=1}^n (\lambda u_i + \mu v_i) f(e_i) \\ &= \lambda \sum_{i=1}^n u_i f(e_i) + \mu \sum_{i=1}^n v_i f(e_i) \\ &= \lambda \alpha(u) + \mu \alpha(v) \end{aligned}$$

as required. □

Remarks.

(1) With a little care the proof of the proposition can be extended to the case U is not assumed finite dimensional.

(2) It is not hard to see that the only subsets S of U that satisfy the conclusions of the proposition are bases: spanning is necessary for the uniqueness part and linear independence is necessary for the existence part. The proposition should be considered a key justification for the definition of a basis.

Corollary. *If U and V are finite dimensional vector spaces over $\mathbf{F}$ with (ordered) bases $\langle e_1, \ldots, e_m\rangle$ and $\langle f_1, \ldots, f_n\rangle$ respectively then there is a bijection*

$$\mathrm{Mat}_{n,m}(\mathbf{F}) \leftrightarrow \mathcal{L}(U, V)$$

that sends a matrix A to the unique linear map α such that $\alpha(e_i) = \sum a_{ji} f_j$.

Interpretation The ith column of the matrix A tells where the ith basis vector of U goes (as a linear combination of the basis vectors of V).

Proof. If $\alpha\colon U \to V$ is a linear map then for each $1 \leqslant i \leqslant m$ we can write $\alpha(e_i)$ uniquely as $\alpha(e_i) = \sum a_{ji} f_j$ with $a_{ji} \in \mathbf{F}$. The proposition tells us that every matrix $A = (a_{ij})$ arises in this way from some linear map and that α is determined by A. □

Definition. We call the matrix corresponding to a linear map $\alpha \in \mathcal{L}(U, V)$ under this corollary the *matrix representing α with respect to $\langle e_1, \ldots, e_m\rangle$ and $\langle f_1, \ldots, f_n\rangle$*.

Exercise. The bijection given by the corollary is even an isomorphism of vector spaces. Thus $\dim \mathcal{L}(U, V) = \dim U \dim V$.

Proposition. *Suppose that U, V and W are finite dimensional vector spaces over $\mathbf{F}$ with bases $R := \langle u_1, \ldots, u_r\rangle$, $S := \langle v_1, \ldots, v_s\rangle$ and $T := \langle w_1, \ldots, w_t\rangle$ respectively. If $\alpha\colon U \to V$ is a linear map represented by the matrix A with respect to R and S and $\beta\colon V \to W$ is a linear map represented by the matrix B with respect to S and T then $\beta\alpha$ is the linear map $U \to W$ represented by BA with respect to R and T.*

Proof. Verifying that $\beta\alpha$ is linear is straightforward: suppose $x, y \in U$ and $\lambda, \mu \in \mathbf{F}$ then

$$\beta\alpha(\lambda x + \mu y) = \beta(\lambda\alpha(x) + \mu\alpha(y)) = \lambda\beta\alpha(x) + \mu\beta\alpha(y).$$

Next we compute $\beta\alpha(u_i)$ as a linear combination of w_j.

$$\beta\alpha(u_i) = \beta\left(\sum_k A_{ki}v_k\right) = \sum_k A_{ki}\beta(v_k) = \sum_{k,j} A_{ki}B_{jk}w_j = \sum_j (BA)_{ji}w_j$$

as required.

□

2.3. **The first isomorphism theorem and the rank-nullity theorem.** The following analogue of the first isomorphism theorem for groups holds for vector spaces.

Lemma (The first isomorphism theorem)**.** *Let $\alpha\colon U \to V$ be a linear map between vector spaces over $\mathbf{F}$. Then $\ker\alpha$ is a subspace of U and $\operatorname{Im}\alpha$ is a subspace of V. Moreover α induces an isomorphism $U/\ker\alpha \to \operatorname{Im}\alpha$ given by*

$$\overline{\alpha}(u + \ker\alpha) = \alpha(u).$$

Proof. Certainly $0 \in \ker\alpha$. Suppose that $u_1, u_2 \in \ker\alpha$ and $\lambda, \mu \in \mathbf{F}$. Then

$$\alpha(\lambda u_1 + \mu u_2) = \lambda\alpha(u_1) + \mu\alpha(u_2) = 0 + 0 = 0.$$

Thus $\ker\alpha$ is a subspace of U. Similarly $0 \in \operatorname{Im}\alpha$ and for $u_1, u_2 \in U$,

$$\lambda\alpha(u_1) + \mu\alpha(u_2) = \alpha(\lambda u_1 + \mu u_2) \in \operatorname{Im}(\alpha).$$

The remainder is left as a (straightforward yet non-examinable) exercise. [Hint: the first isomorphism theorem for groups gives that $\overline{\alpha}$ is a bijective homomorphism of the underlying abelian groups so it remains to verify that $\overline{\alpha}$ respects multiplication by scalars.] □

Definition. Suppose that $\alpha\colon U \to V$ is a linear map between finite dimensional vector spaces.

- The number $n(\alpha) := \dim\ker\alpha$ is called the *nullity* of α.
- The number $r(\alpha) := \dim\operatorname{Im}\alpha$ is called the *rank* of α.

Corollary (The rank-nullity theorem)**.** *If $\alpha\colon U \to V$ is a linear map between f.d. vector spaces over $\mathbf{F}$ then*

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

Proof. Since $U/\ker\alpha \cong \operatorname{Im}\alpha$ this follows immediately from an earlier exercise. □

We are about to give another proof of the rank-nullity theorem not using quotient spaces or the first isomorphism theorem. However, the proof above is illustrative of the power of considering quotients.

Proposition. *Suppose that $\alpha\colon U \to V$ is a linear map between finite dimensional vector spaces then there are bases $\{e_1, \ldots, e_n\}$ for U and $\{f_1, \ldots, f_m\}$ for V such that the matrix representing α is*

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

where $r = \mathrm{r}(\alpha)$.

Proof. Let $e_{k+1}, \ldots, e_n$ be a basis for $\ker\alpha$ (here $\mathrm{n}(\alpha) = n - k$) and extend it to a basis $e_1, \ldots, e_n$ for U (we're being careful about ordering now so that we don't have to change it later). Let $f_i = \alpha(e_i)$ for $1 \leqslant i \leqslant k$.

We claim that $\{f_1, \ldots, f_k\}$ form a basis for $\operatorname{Im}\alpha$ (so that $k = \mathrm{r}(\alpha)$). Suppose first that $\sum_{i=1}^k \lambda_i f_i = 0$ for some $\lambda_i \in \mathbf{F}$. Then $\alpha\left(\sum_{i=1}^k \lambda_i e_i\right) = 0$ and so $\sum_{i=1}^k \lambda_i e_i \in \ker\alpha$. But $\ker\alpha \cap \langle e_1, \ldots, e_k \rangle = 0$ by construction and so $\sum_{i=1}^k \lambda_i e_i = 0$. Since $e_1, \ldots, e_k$ are LI, each $\lambda_i = 0$. Thus we have shown that $\{f_1, \ldots, f_k\}$ is LI.

Now suppose that $v \in \operatorname{Im}\alpha$, so that $v = \alpha(\sum_{i=1}^n \mu_i e_i)$ for some $\mu_i \in \mathbf{F}$. Since $\alpha(e_i) = 0$ for $i > k$ and $\alpha_i(e_i) = f_i$ for $i \leqslant k$, $v = \sum_{i=1}^k \mu_i f_i \in \langle f_1, \ldots, f_k \rangle$. So $\{f_1, \ldots, f_k\}$ is a basis for $\operatorname{Im}\alpha$ as claimed (and $k = r$).

We can extend $\{f_1, \ldots, f_r\}$ to a basis $\{f_1, \ldots, f_m\}$ for V.

Now

$$\alpha(e_i) = \begin{cases} f_i & 1 \leqslant i \leqslant r \\ 0 & r+1 \leqslant i \leqslant m \end{cases}$$

so the matrix representing α with respect to our choice of basis is as in the statement. $\square$

The proposition says that the rank of a linear map between two finite dimensional vector spaces is its only basis-independent invariant (or more precisely any other invariant can be deduced from it).

Corollary (The rank-nullity theorem)**.** *If $\alpha\colon U \to V$ is a linear map between finite dimensional vector spaces then*

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

Proof. This can easily be read off from either the statement or the proof of the Proposition. $\square$

Lecture 6

Recall the statement of the rank-nullity theorem.

Theorem (The rank-nullity theorem)**.** *If $\alpha\colon U \to V$ is a linear map between finite dimensional vector spaces then*

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

This result is very useful for computing dimensions of vector spaces in terms of known dimensions of other spaces.

Example. Let $W = \{\mathbf{x} \in \mathbf{R}^5 \mid x_1 + x_2 + x_5 = 0 \text{ and } x_3 - x_4 - x_5 = 0\}$. What is $\dim W$?

Consider $\alpha\colon \mathbf{R}^5 \to \mathbf{R}^2$ given by $\alpha(\mathbf{x}) = \begin{pmatrix} x_1 + x_2 + x_5 \\ x_3 - x_4 - x_5 \end{pmatrix}$. Then α is a linear map with image $\mathbf{R}^2$ (since

$$\alpha\left(\begin{pmatrix}1\\0\\0\\0\\0\end{pmatrix}\right) = \begin{pmatrix}1\\0\end{pmatrix} \text{ and } \alpha\left(\begin{pmatrix}0\\0\\1\\0\\0\end{pmatrix}\right) = \begin{pmatrix}0\\1\end{pmatrix}.)$$

and $\ker\alpha = W$. Thus $\dim W = n(\alpha) = 5 - \mathrm{r}(\alpha) = 5 - 2 = 3$.

More generally, one can use the rank-nullity theorem to see that m linear equations in n unknowns have a space of solutions of dimension at least $n - m$.

Example. Suppose that U and W are subspaces of a finite dimensional vector space V then let $\alpha\colon U \oplus W \to V$ be the linear map given by $\alpha((u, w)) = u + w$. Then $\ker\alpha = \{(u, -u) \mid u \in U \cap W\} \cong U \cap W$, and $\operatorname{Im}\alpha = U + W$. Thus

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

We can then recover $\dim U + \dim W = \dim (U + W) + \dim (U \cap W)$.

Corollary (of the rank-nullity theorem)**.** *Suppose that $\alpha\colon U \to V$ is a linear map between two vector spaces of dimension $n < \infty$. Then the following are equivalent:*
(a) α is injective;
(b) α is surjective;
(c) α is an isomorphism.

Proof. It suffices to see that (a) is equivalent to (b) since these two together are already known to be equivalent to (c). Now α is injective if and only if $n(\alpha) = 0$. By the rank-nullity theorem $n(\alpha) = 0$ if and only if $r(\alpha) = n$ and the latter is equivalent to α being surjective. □

This enables us to prove the following fact about matrices.

Lemma. *Let A be an $n \times n$ matrix over $\mathbf{F}$. The following are equivalent*
(i) there is a matrix B such that $BA = I_n$;
(ii) there is a matrix C such that $AC = I_n$.
Moreover, if (i) and (ii) hold then $B = C$ and we write $A^{-1} = B = C$; we say A is invertible.

Proof. Let $\alpha, \beta, \gamma, \iota\colon \mathbf{F}^n \to \mathbf{F}^n$ be the linear maps represented by A, B, C and I_n respectively (with respect to the standard basis for $\mathbf{F}^n$). Then (i) implies that $\beta\alpha = \iota$ thus α is injective and so an isomorphism. Thus $\alpha^{-1} = \beta$ is represented by B with respect to the standard basis and $AB = I_n$. Similarly (ii) implies that $\alpha\gamma = \iota$ thus α is surjective and so an isomorphism. Thus $\alpha^{-1} = \gamma$ is represented by C with respect to the standard basis and $CA = I_n$.

Finally $C = I_n C = BAC = BI_n = B$. □

2.4. Change of basis.

Theorem. *Suppose that $\langle e_1, \ldots, e_m\rangle$ and $\langle u_1, \ldots, u_m\rangle$ are two bases for a vector space U over $\mathbf{F}$ and $\langle f_1, \ldots, f_n\rangle$ and $\langle v_1, \ldots, v_n\rangle$ are two bases of another veector space V. Let $\alpha\colon U \to V$ be a linear map, A be the matrix representing α with* *respect to* $\langle e_1, \dots, e_m\rangle$ *and* $\langle f_1, \dots, f_m\rangle$ *and* B *be the matrix representing* α *with respect to* $\langle u_1, \dots, u_m\rangle$ *and* $\langle v_1, \dots, v_m\rangle$ *then*

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

where $u_i = \sum P_{ki}e_k$ *for* $i = 1, \dots, m$ *and* $v_j = \sum Q_{lj}f_l$ *for* $j = 1, \dots, n$.

Note that one can view P as the matrix representing the identity map from U with basis $\langle u_1, \dots, u_m\rangle$ to U with basis $\langle e_1, \dots, e_m\rangle$ and Q as the matrix representing the identity map from V with basis $\langle v_1, \dots, v_n\rangle$ to V with basis $\langle f_1, \dots, f_n\rangle$. Thus both are invertible with inverses represented by the identity maps going in the opposite directions.

Proof. On the one hand, by definition

$$\alpha(u_i) = \sum_j B_{ji}v_j = \sum_{j,l} B_{ji}Q_{lj}f_l = \sum_l (QB)_{li}f_l.$$

On the other hand, also by definition

$$\alpha(u_i) = \alpha\left(\sum_k P_{ki}e_k\right) = \sum_{k,l} P_{ki}A_{lk}f_l = \sum_l (AP)_{li}f_l.$$

Thus $QB = AP$ as the f_l are LI. Since Q is invertible the result follows. □

Definition. We say two matrices $A, B \in \mathrm{Mat}_{n,m}(\mathbf{F})$ are *equivalent* if there are invertible matrices $P \in \mathrm{Mat}_m(\mathbf{F})$ and $Q \in \mathrm{Mat}_n(\mathbf{F})$ such that $Q^{-1}AP = B$.

Note that equivalence is an equivalence relation. It can be reinterpreted as follows: two matrices are equivalent precisely if they respresent the same linear map with respect to different bases.

We saw earlier that for every linear map α between f.d. vector spaces there are bases for the domain and codomain such that α is represented by a matrix of the form

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

Moreover $r = r(\alpha)$ is independent of the choice of bases. We can now rephrase this as follows.

Corollary. *If* $A \in \mathrm{Mat}_{n,m}(\mathbf{F})$ *there are invertible matrices* $P \in \mathrm{Mat}_m(\mathbf{F})$ *and* $Q \in \mathrm{Mat}_n(\mathbf{F})$ *such that* $Q^{-1}AP$ *is of the form*

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

Moreover r *is uniquely determined by* A. *i.e. every equivalence class contains precisely one matrix of this form.* □

Definition. If $A \in \mathrm{Mat}_{n,m}(\mathbf{F})$ then

- *column rank* of A, written $r(A)$ is the dimension of the subspace of $\mathbf{F}^n$ spanned by the columns of A;
- the *row rank* of A is the column rank of A^T.

Note that if we take α to be a linear map represented by A with respect to the standard bases of $\mathbf{F}^m$ and $\mathbf{F}^n$ then $r(A) = \mathrm{r}(\alpha)$. i.e. 'column rank=rank'. Moreover, since $r(\alpha)$ is defined in a basis-invariant way, the column rank of A is constant on equivalence classes.

Corollary (Row rank equals column rank)**.** *If $A \in \mathrm{Mat}_{m,n}(\mathbf{F})$ then $r(A) = r(A^T)$.*

Proof. Let $r = r(A)$. There exist P, Q such that

$$Q^{-1}AP = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Thus

$$P^T A^T (Q^{-1})^T = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

and so $r = r(A^T)$. Thus A and A^T have the same rank. □

Lecture 7

2.5. Elementary matrix operations.

Definition. We call the following three types of invertible $n \times n$ matrices *elementary matrices*

$$S^n_{ij} := \begin{array}{c} \begin{matrix} \quad & \quad & i & \quad & j & \quad & \quad \end{matrix} \\ \begin{pmatrix} 1 & & & & & & 0 \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{pmatrix} \end{array} \text{ for } i \neq j,$$

$$E^n_{ij}(\lambda) := \begin{array}{c} \begin{matrix} \quad & \quad & i & \quad & j & \quad & \quad \end{matrix} \\ \begin{pmatrix} 1 & & & & & & 0 \\ & \ddots & & & & & \\ & & 1 & & \lambda & & \\ & & & \ddots & & & \\ & & 0 & & 1 & & \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{pmatrix} \end{array} \text{ for } i \neq j,\ \lambda \in \mathbf{F} \text{ and}$$

$$T^n_i(\lambda) := \begin{array}{c} \begin{matrix} \quad & \quad & \quad & i & \quad & \quad & \quad \end{matrix} \\ \begin{pmatrix} 1 & & & & & & 0 \\ & \ddots & & & & & \\ & & 1 & & 0 & & \\ & & & \lambda & & & \\ & & 0 & & 1 & & \\ & & & & & \ddots & \\ 0 & & & & & & 1 \end{pmatrix} \end{array} \text{ for } \lambda \in \mathbf{F} \backslash \{0\}.$$

We make the following observations: if A is an $m \times n$ matrix then AS^n_{ij} (resp. $S^m_{ij}A$) is obtained from A by swapping the ith and jth columns (resp. rows), $AE^n_{ij}(\lambda)$ (resp. $E^m_{ij}(\lambda)A$) is obtained from A by adding $\lambda \cdot$ (column i) to column j (resp. adding $\lambda \cdot$ (row j) to row i) and $AT^n_i(\lambda)$ (resp. $T^m_i(\lambda)A$) is obtained from A by multiplying column (resp. row) i by λ.

Recall the following result.

Proposition. *If $A \in \mathrm{Mat}_{n,m}(\mathbf{F})$ there are invertible matrices $P \in \mathrm{Mat}_m(\mathbf{F})$ and $Q \in \mathrm{Mat}_n(\mathbf{F})$ such that $Q^{-1}AP$ is of the form*

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

Pure matrix proof of the Proposition. We claim that there are elementary matrices $E_1^n, \ldots, E_a^n$ and $F_1^m, \ldots, F_b^m$ such that $E_a^n \cdots E_1^n A F_1^m \cdots F_b^m$ is of the required form. This suffices since all the elementary matrices are invertible and products of invertible matrices are invertible.

Moreover, to prove the claim it suffices to show that there is a sequence of elementary row and column operations that reduces A to the required form.

If $A = 0$ there is nothing to do. Otherwise, we can find a pair i, j such that $A_{ij} \neq 0$. By swapping rows 1 and i and then swapping rows 1 and j we can reduce to the case that $A_{11} \neq 0$. By multiplying row 1 by $\frac{1}{A_{11}}$ we can further assume that $A_{11} = 1$.

Now, given $A_{11} = 1$ we can add $-A_{1j}$ times column 1 to column j for each $1 < j \leqslant m$ and then add $-A_{i1}$ times row 1 to row i for each $1 < i \leqslant n$ to reduce further to the case that A is of the form

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

Now by induction on the size of A we can find elementary row and column operations that reduces B to the required form. Applying these 'same' operations to A we complete the proof. □

Note that the algorithm described in the proof can easily be implemented on a computer in order to actually compute the matrices P and Q.

Exercise. Show that elementary row and column operations do not alter $r(A)$ or $r(A^T)$. Conclude that the r in the statement of the proposition is thus equal to $r(A)$ and to $r(A^T)$.

3. Determinants of matrices

Recall that S_n is the group of permutations of the set $\{1, \ldots, n\}$. Moreover we can define a group homomorphism $\epsilon\colon S_n \to \{\pm 1\}$ such that $\epsilon(\sigma) = 1$ whenever σ is a product of an even number of transpositions and $\epsilon(\sigma) = -1$ whenever σ is a product of an odd number of transpositions.

Definition. If $A \in \mathrm{Mat}_n(\mathbf{F})$ then the *determinant of* A

$$\det A := \sum_{\sigma \in S_n} \epsilon(\sigma) \left(\prod_{i=1}^n A_{i\sigma(i)} \right).$$

Example. If $n = 2$ then $\det A = A_{11}A_{22} - A_{12}A_{21}$.

Lemma. $\det A = \det A^T$

Proof.

$$\begin{aligned} \det A^T &= \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^n A_{\sigma(i)i} \\ &= \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^n A_{i\sigma^{-1}(i)} \\ &= \sum_{\tau \in S_n} \epsilon(\tau^{-1}) \prod_{i=1}^n A_{i\tau(i)} \\ &= \det A \end{aligned}$$

□

Lemma. *Let $A \in \mathrm{Mat}_n(\mathbf{F})$ be upper triangular ie*

$$A = \begin{pmatrix} a_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & a_n \end{pmatrix}$$

then $\det A = \prod_{i=1}^n a_i$.

Proof.

$$\det A = \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^n A_{i\sigma(i)}.$$

Now $A_{i\sigma(i)} = 0$ if $i > \sigma(i)$. So $\prod_{i=1}^n A_{i\sigma(i)} = 0$ unless $i \leqslant \sigma(i)$ for all $i = 1, \ldots, n$. Since σ is a permutation $\prod_{i=1}^n A_{i\sigma(i)}$ is only non-zero when $\sigma = \mathrm{id}$. The result follows immediately. □

Definition. A *volume form* d on $\mathbf{F}^n$ is a function $\mathbf{F}^n \times \mathbf{F}^n \times \cdots \times \mathbf{F}^n \to \mathbf{F}$; $(v_1, \ldots, v_n) \mapsto d(v_1, \ldots, v_n)$ such that

(i) d is *multi-linear* i.e. for each $1 \leqslant i \leqslant n$

$$d(v_1, \ldots, \lambda v_i + \mu v_i', \ldots, v_n) = \lambda d(v_1, \ldots, v_i, \ldots, v_n) + \mu d(v_1, \ldots, v_i', \ldots, v_n);$$

(ii) d is *alternating* i.e. whenever $v_i = v_j$ for some $i \neq j$ then $d(v_1, \ldots, v_n) = 0$.

Note that one may view a matrix $A \in \mathrm{Mat}_n(\mathbf{F})$ as an n-tuple of elements of $\mathbf{F}^n$ given by its columns $A = (A^{(1)} \cdots A^{(n)})$ with $A^{(1)}, \ldots, A^{(n)} \in \mathbf{F}^n$.

Lemma. $\det\colon \mathbf{F}^n \times \cdots \mathbf{F}^n \to \mathbf{F}$*;* $(A^{(1)}, \ldots, A^{(n)}) \mapsto \det A$ *is a volume form.*

Proof. To see that det is multilinear it suffices to see that $\prod_{i=1}^n A_{i\sigma(i)}$ is multilinear for each $\sigma \in S_n$ since a sum of (multi)-linear functions is (multi)-linear. Since one term from each column appears in each such product this is easy to see.

Suppose now that $A^{(k)} = A^{(l)}$ for some $k \neq l$. Let τ be the transposition (kl). Then $a_{ij} = a_{i\tau(j)}$ for every i, j in $\{1, \ldots, n\}$. We can write S_n is a disjoint union of cosets $A_n \coprod \tau A_n$.

Then

$$\sum_{\sigma \in A_n} \prod a_{i\sigma(i)} = \sum_{\sigma \in A_n} \prod a_{i\tau\sigma(i)} = \sum_{\sigma \in \tau A_n} \prod a_{i\sigma(i)}$$

Thus $\det A = \mathrm{LHS} - \mathrm{RHS} = 0$. □

Lecture 8

We continue thinking about volume forms.

Lemma. *Let d be a volume form. Swapping two entries changes the sign. i.e.*

$$d(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -d(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n).$$

Proof. Consider $d(v_1, \ldots, v_i + v_j, \ldots, v_i + v_j, \ldots, v_n) = 0$. Expanding the left-hand-side using linearity of the ith and jth coordinates we obtain

$$\begin{aligned} d(v_1, \ldots, v_i, \ldots, v_i, \ldots, v_n) + d(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) + & \\ d(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) + d(v_1, \ldots, v_j, \ldots, v_j, \ldots, v_n) &= 0. \end{aligned}$$

Since the first and last terms on the left are zero, the statement follows immediately. □

Corollary. *If $\sigma \in S_n$ then $d(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) = \epsilon(\sigma) d(v_1, \ldots, v_n)$.* □

Theorem. *Let d be a volume form on $\mathbf{F}^n$. Let A be a matrix with ith column $A^{(i)} \in \mathbf{F}^n$. Then*

$$d(A^{(1)}, \ldots, A^{(n)}) = \det A \cdot d(e_1, \ldots, e_n).$$

In order words det *is the unique volume form d such that $d(e_1, \ldots, e_n) = 1$.*

Proof. We compute

$$\begin{aligned} d(A^{(1)}, \ldots, A^{(n)}) &= d(\sum_{i=1}^n A_{i1} e_i, A^{(2)}, \ldots, A^{(n)}) \\ &= \sum_i A_{i1} d(e_i, A^{(2)}, \ldots, A^{(n)}) \\ &= \sum_{i,j} A_{i1} A_{j2} d(e_i, e_j, \ldots, A^{(n)}) \\ &= \sum_{i_1, \ldots, i_n} \left(\prod_{j=1}^n A_{i_j j} \right) d(e_{i_1}, \ldots, e_{i_n}) \end{aligned}$$

But $d(e_{i_1}, \ldots, e_{i_n}) = 0$ unless $i_1, \ldots, i_n$ are distinct. That is unless there is some $\sigma \in S_n$ such that $i_j = \sigma(j)$. Thus

$$d(A^{(1)}, \ldots, A^{(n)}) = \sum_{\sigma \in S_n} \left(\prod_{j=1}^n A_{\sigma(j)j} \right) d(e_{\sigma(1)}, \ldots, e_{\sigma(n)}).$$

But $d(e_{\sigma(1)}, \ldots, e_{\sigma(n)} = \epsilon(\sigma) d(e_1, \ldots, e_n)$ so we're done. □

Remark. We can interpret this as saying that for every matrix A,

$$d(Ae_1, \ldots, Ae_n) = \det A \cdot d(e_1, \ldots, e_n).$$

The same proof gives $d(Av_1, \ldots, Av_n) = \det A \cdot d(v_1, \ldots, v_n)$ for all $v_1, \ldots, v_n \in \mathbf{F}^n$. We can view this result as the motivation for the formula defining the determinant; $\det A$ is the unique way to define the 'volume scaling factor' of the linear map given by A.

Theorem. *Let $A, B \in \mathrm{Mat}_n(\mathbf{F})$. Then $\det(AB) = \det A \det B$.*

Proof. Let d be a non-zero volume form on $\mathbf{F}^n$, for example det. Then we can compute

$$d(ABe_1, \ldots, ABe_n) = \det(AB) \cdot d(e_1, \ldots, e_n)$$

by the last theorem. But we can also compute

$$d(ABe_1, \ldots, ABe_n) = \det A \cdot d(Be_1 \ldots, Be_n) = \det A \det B \cdot d(e_1, \ldots, e_n)$$

by the remark extending the last theorem. Thus as $d(e_1, \ldots, e_n) \neq 0$ we can see that $\det(AB) = \det A \det B$ □

Corollary. *If A is invertible then $\det A \neq 0$ and $\det(A^{-1}) = \frac{1}{\det A}$.*

Proof. We can compute

$$1 = \det I_n = \det(AA^{-1}) = \det A \det A^{-1}.$$

Thus $\det A^{-1} = \frac{1}{\det A}$ as required. $\square$

Theorem. *Let $A \in \mathrm{Mat}_n(\mathbf{F})$. The following statements are equivalent:*
(a) A is invertible;
(b) $\det A \neq 0$;
(c) $r(A) = n$.

Proof. We've seen that (a) implies (b) above.

Suppose that $r(A) < n$. Then by the rank-nullity theorem $n(A) >_0$ and so there is some $\lambda \in \mathbf{F}^n \backslash 0$ such that $A\lambda = 0$ i.e. there is a linear relation between the columns of A; $\sum_{i=1}^n \lambda_i A^{(i)} = 0$ for some $\lambda_i \in \mathbf{F}$ not all zero.

Suppose that $\lambda_k \neq 0$ and let B be the matrix with ith column e_i for $i \neq k$ and kth column λ. Then AB has kth column 0. Thus $\det AB = 0$. But we can compute $\det AB = \det A \det B = \lambda_k \det A$. Since $\lambda_k \neq 0$, $\det A = 0$. Thus (b) implies (c).

Finally (c) implies (a) by the rank-nullity theorem: $r(A) = n$ implies $n(A) = 0$ and the linear map corresponding to A is bijective as required. $\square$

Notation. Let $\widehat{A_{ij}}$ denote the submatrix of A obtained by deleting the ith row and the jth column.

Lemma. *Let $A \in \mathrm{Mat}_n(\mathbf{F})$. Then*
(a) expanding determinant along the jth column $\det A = \sum_{i=1}^n (-1)^{i+j} A_{ij} \det \widehat{A_{ij}}$;
(b) expanding determinant along the ith row $\det A = \sum_{j=1}^n (-1)^{i+j} A_{ij} \det \widehat{A_{ij}}$.

Proof. Since $\det A = \det A^T$ it suffices to verify (a).

Now

$$\begin{aligned}
\det A &= \det(A^{(1)}, \ldots, A^{(n)}) \\
&= \det(A^{(1)}, \ldots, \sum_i A_{ij} e_i, \ldots, A^{(n)}) \\
&= \sum_i A_{ij} \det(A^{(1)}, \ldots, e_i, \ldots, A^{(n)}) \\
&= \sum_i A_{ij} (-1)^{i+j} \det B
\end{aligned}$$

where

$$B = \begin{pmatrix} \widehat{A_{ij}} & 0 \\ * & 1 \end{pmatrix}.$$

Finally for $\sigma \in S_n$, $\prod_{i=1}^n B_{i\sigma(i)} = 0$ unless $\sigma(n) = n$ and we see easily that $\det B = \det \widehat{A_{ij}}$ as required. $\square$

Definition. Let $A \in \mathrm{Mat}_n(\mathbf{F})$. The *adjugate matrix* $\operatorname{adj} A$ is the element of $\mathrm{Mat}_n(\mathbf{F})$ such that

$$(\operatorname{adj} A)_{ij} = (-1)^{i+j} \det \widehat{A_{ji}}.$$

Theorem. *Let $A \in \mathrm{Mat}_n(\mathbf{F})$. Then*

$$(\operatorname{adj} A)A = A(\operatorname{adj} A) = (\det A) I_n.$$

Thus if $\det A \neq 0$ then $A^{-1} = \frac{1}{\det A} \operatorname{adj} A$

Proof. We compute

$$\begin{aligned}((\operatorname{adj} A)A)_{jk} &= \sum_{i=1}^{n} (\operatorname{adj} A)_{ji} A_{ik} \\ &= \sum_{i=1}^{n} (-1)^{j+i} \det \widehat{A_{ij}} A_{ik}\end{aligned}$$

The right-hand-side is $\det A$ if $k = j$. If $k \neq j$ then the right-hand-side is the determinant of the matrix obtained by replacing the jth column of A by the kth column. Since the resulting matrix has two identical columns $((\operatorname{adj} A)A)_{jk} = 0$ in this case. Therefoe $(\operatorname{adj} A)A = (\det A)I_n$ as required.

We can now obtain $A \operatorname{adj} A = (\det A)I_n$ either by using a similar argument using the rows or by considering the transpose of $A \operatorname{adj} A$. The final part follows immediately. □

Remark. Note that the entries of the adjugate matrix are all given polynomials in the entries of A. Since the determinant is also a polynomial, it follows that the entries of the inverse of an invertible square matrix are given by a rational function (i.e. a ratio of two polynomial functions) in the entries of A. Whilst this is a very useful fact from a theoretical point of view, computationally there are better ways of computing the determinant and inverse of a matrix than using these formulae.

Lecture 9

We'll complete this section on determinants of matrices with a couple of results about block triangular matrices.

Lemma. *Let A and B be square matrices. Then*

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

Proof. Suppose $A \in \mathrm{Mat}_k(\mathbf{F})$ and $B \in \mathrm{Mat}_l(\mathbf{F})$ and $k + l = n$ so $C \in \mathrm{Mat}_{k,l}(\mathbf{F})$. Define

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

then

$$\det X = \sum_{\sigma \in S_n} \epsilon(\sigma) \left(\prod_{i=1}^{n} X_{i\sigma(i)} \right).$$

Since $X_{ij} = 0$ whenever $i > k$ and $j \leqslant k$ the terms with σ such that $\sigma(i) \leqslant k$ for some $i > k$ are all zero. So we may restrict the sum to those σ such that $\sigma(i) > k$ for $i > k$ i.e. those σ that restrict to a permutation of $\{1, \ldots, k\}$. We may factorise these σ as $\sigma = \sigma_1\sigma_2$ with $\sigma_1 \in S_k$ and σ_2 a permuation of $\{k+1, \ldots, n\}$. Thus

$$\begin{aligned}\det X &= \sum_{\sigma_1}\sum_{\sigma_2} \epsilon(\sigma_1\sigma_2) \left(\prod_{i=1}^{k} X_{i\sigma_1(i)} \right) \left(\prod_{j=1}^{l} X_{j+k,\sigma_2(j+k)} \right) \\ &= \left(\sum_{\sigma_1 \in S_k} \epsilon(\sigma_1) \left(\prod_{i=1}^{k} A_{i\sigma_1(i)} \right) \right) \left(\sum_{\sigma_2' \in S_l} \epsilon(\sigma_2') \left(\prod_{j=1}^{l} B_{j\sigma_2'(j)} \right) \right) \\ &= \det A \det B\end{aligned}$$

□

Corollary.
$$\det \begin{pmatrix} A_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & A_k \end{pmatrix} = \prod_{i=1}^{k} \det A_i \quad \square$$

Warning: it is *not* true in general that if $A, B, C, D \in \mathrm{Mat}_n(\mathbf{F})$ and M is the element of $\mathrm{Mat}_{2n}(\mathbf{F})$ given by
$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$
then $\det M = \det A \det D - \det B \det C$.

4. Endomorphisms

4.1. Invariants.

Definition. Suppose that V is a finite dimensional vector space over $\mathbf{F}$. An *endomorphism* of V is a linear map $\alpha \colon V \to V$. Let $\mathrm{End}(V)$ denote the vector space of endomorphisms of V. We'll write ι to denote the identity endomorphism of V.

When considering endomorphisms as matrices it is usual to choose the same basis for V for both the domain and the range.

Lemma. *Suppose that $\langle e_1, \ldots, e_n \rangle$ and $\langle f_1, \ldots, f_n \rangle$ are bases for V such that $f_i = \sum P_{ki} e_k$. Let $\alpha \in \mathrm{End}(V)$, A be the matrix representing α with respect to $\langle e_1, \ldots, e_n \rangle$ and B the matrix representing α with respect to $\langle f_1, \ldots, f_n \rangle$. Then $B = P^{-1}AP$.*

Proof. This is a special case of the change of basis formula for all linear maps between f.d. vector spaces. □

Definition. We say matrices A and B are *similar* (or *conjugate*) if $B = P^{-1}AP$ for some invertible matrix P.

Recall $GL_n(\mathbf{F})$ denotes all the invertible matrices in $\mathrm{Mat}_n(\mathbf{F})$. Then $GL_n(\mathbf{F})$ acts on $\mathrm{Mat}_n(\mathbf{F})$ by conjugation and two such matrices are similar precisely if they lie in the same orbit. Thus similarity is an equivalence relation.

An important problem is to classify elements of $\mathrm{Mat}_n(\mathbf{F})$ up to similarity (ie classify $GL_n(\mathbf{F})$-orbits). It will help us to find basis independent invariants of the corresponding endomorphisms. For example we'll see that given $\alpha \in \mathrm{End}(V)$ the rank, trace, determinant, characteristic polynomial and eigenvalues of α are all basis-independent.

Definition. The *trace* of $A \in \mathrm{Mat}_n(\mathbf{F})$ is defined by $\operatorname{tr} A = \sum A_{ii} \in \mathbf{F}$.

Note that trace is a linear map from $\mathrm{Mat}_n(\mathbf{F}) \to \mathbf{F}$.

Lemma.

(a) If $A \in \mathrm{Mat}_{n,m}(\mathbf{F})$ and $B \in \mathrm{Mat}_{m,n}(\mathbf{F})$ then $\operatorname{tr} AB = \operatorname{tr} BA$.
(b) If A and B are similar then $\operatorname{tr} A = \operatorname{tr} B$.
(c) If A and B are similar then $\det A = \det B$.

Proof. (a)

$$\begin{aligned} \operatorname{tr} AB &= \sum_{i=1}^{n}\left(\sum_{j=1}^{m} A_{ij}B_{ji}\right) \\ &= \sum_{j=1}^{m}\left(\sum_{i=1}^{n} B_{ji}A_{ij}\right) \\ &= \operatorname{tr} BA \end{aligned}$$

If $B = P^{-1}AP$ then,
(b) $\operatorname{tr} B = \operatorname{tr}(P^{-1}A)P = \operatorname{tr} P(P^{-1}A) = \operatorname{tr} A$.
(c) $\det B = \det P^{-1} \det A \det P = \frac{1}{\det P} \det A \det P = \det A$. □

Definition. Let $\alpha \in \operatorname{End}(V)$, $\langle e_1, \dots, e_n\rangle$ be a basis for V and A the matrix representing α with respect to $\langle e_1, \dots, e_n\rangle$. Then the *trace* of α written $\operatorname{tr}\alpha$ is defined to be the trace of A and the *determinant* of α written $\det\alpha$ is defined to be the determinant of A.

We've proven that the trace and determinant of α do not depend on the choice of basis $\langle e_1, \dots, e_n\rangle$.

Definition. Let $\alpha \in \operatorname{End}(V)$.

(a) $\lambda \in \mathbf{F}$ is an *eigenvalue* of α if there is $v \in V\backslash 0$ such that $\alpha v = \lambda v$.
(b) $v \in V$ is an *eigenvector* for α if $\alpha(v) = \lambda v$ for some $\lambda \in \mathbf{F}$.
(c) When $\lambda \in \mathbf{F}$, the λ-*eigenspace* of α, written $E_\alpha(\lambda)$ or simply $E(\lambda)$ is the set of λ-eigenvectors of α; i.e. $E(\lambda) = \ker(\alpha - \lambda\iota)$.
(d) The *characteristic polynomial* of α is defined by

$$\chi_\alpha(t) = \det(t\iota - \alpha).$$

Remarks.

(1) $\chi_\alpha(t)$ is a monic polynomial in t of degree n.
(2) $\lambda \in \mathbf{F}$ is an eigenvalue of α if and only if $\ker(\alpha - \lambda\iota) \neq 0$ if and only if λ is a root of $\chi_\alpha(t)$.
(3) If $A \in \operatorname{Mat}_n(F)$ we can define $\chi_A(t) = \det(tI_n - A)$. Then similar matrices have the same characteristic polynomials.

Lemma. *Let $\alpha \in \operatorname{End}(V)$ and $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of α. Then $E(\lambda_1) + \cdots + E(\lambda_k)$ is a direct sum of the $E(\lambda_i)$.*

Proof. Suppose that $\sum_{i=1}^k x_i = \sum_{i=1}^k y_i$ with $x_i, y_i \in E(\lambda_i)$. Consider the linear maps

$$\beta_j := \prod_{i\neq j}(\alpha - \lambda_i\iota).$$

Then

$$\begin{aligned}\beta_j(\sum_{i=1}^k x_i) &= \sum_{i=1}^k \beta_j(x_i) \\ &= \sum_{i=1}^k \left(\prod_{r\neq j}(\alpha - \lambda_r\iota)(x_i)\right) \\ &= \sum_{i=1}^k \left(\prod_{r\neq j}(\lambda_i - \lambda_r)x_i\right) \\ &= \prod_{r\neq j}(\lambda_j - \lambda_r)x_i\end{aligned}$$

Similarly, $\beta_j(\sum_{i=1}^k y_i) = \prod_{r\neq j}(\lambda_j - \lambda_r)y_i$. Thus since $\prod_{r\neq j}(\lambda_j - \lambda_r) \neq 0$, $x_j = y_j$ and the expression is unique. □

Note that the proof of this lemma show that any set of non-zero eigenvectors with distinct eigenvalues is LI.

Lecture 10

Definition. $\alpha \in \mathrm{End}(V)$ is *diagonalisable* if there is a basis for V such that the corresponding matrix is diagonal.

Theorem. *Let $\alpha \in \mathrm{End}(V)$. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of α. Write $E_i = E(\lambda_i)$. Then the following are equivalent*
(a) α is diagonalisable;
(b) V has a basis consisting of eigenvectors of α;
(c) $V = \oplus_{i=1}^k E_i$;
(d) $\sum \dim E_i = \dim V$.

Proof. Suppose that $\langle e_1, \ldots, e_n\rangle$ is a basis for V and A is the matrix representing α with respect to this basis. Then $\alpha(e_i) = \sum A_{ji}e_j$. Thus A is diagonal if and only if each e_i is an eigenvector for α. i.e. (a) and (b) are equivalent.

Now (b) is equivalent to $V = \sum E_i$ and we've proven that $\sum E_i = \oplus_{i=1}^k E_i$ so (b) and (c) are equivalent.

The equivalence of (c) and (d) is a basic fact about direct sums that follows from Example Sheet 1 Q10. □

4.1.1. *An aside on polynomials.*

Definition. A polynomial function $f\colon \mathbf{F} \to \mathbf{F}$ is one of the form

$$f(t) = a_m t^m + \cdots + a_1 t + a_0$$

for some $m \geqslant 0$ and $a_0, \ldots, a_m \in \mathbf{F}$. The largest n such that $a_n \neq 0$ is the *degree* of f written $\deg f$. Thus $\deg 0 = -\infty$.

It is straightforward to show that

$$\deg(f+g) \leqslant \max(\deg f, \deg g)$$

and

$$\deg fg = \deg f + \deg g.$$

Notation. We write $\mathbf{F}[t] := \{\text{polynomials with coefficients in } \mathbf{F}\}$.

Lemma (Polynomial division)**.** *Given $f, g \in \mathbf{F}[t]$, $g \neq 0$ there exist $q, r \in \mathbf{F}[t]$ such that $f(t) = q(t)g(t) + r(t)$ and $\deg r < \deg g$.*

Lemma. *If $\lambda \in \mathbf{F}$ is a root of a polynomial $f(t)$, i.e. $f(\lambda) = 0$, then $f(t) = (t - \lambda)g(t)$ for some $g(t) \in \mathbf{F}[t]$.*

Proof. There are $q, r \in \mathbf{F}[t]$ such that $f(t) = (t - \lambda)q(t) + r(t)$ with $\deg r < 1$. But $\deg r < 1$ means $r(t) = r_0$ some $r_0 \in \mathbf{F}$. But then $0 = f(\lambda) = (\lambda - \lambda)q(\lambda) + r_0 = r_0$. So $r_0 = 0$ and we're done. $\square$

Definition. If $f \in \mathbf{F}[t]$ and $\lambda \in \mathbf{F}$ is a root of f we say that λ is a *root of multiplicity k* if $(t - \lambda)^k$ is a factor of $f(t)$ but $(t - \lambda)^{k+1}$ is not a factor of f. i.e. if $f(t) = (t - \lambda)^k g(t)$ for some $g(t) \in \mathbf{F}[t]$ with $g(\lambda) \neq 0$.

We can use the last lemma and induction to show that every $f(t)$ can be written as

$$f(t) = \prod_{i=1}^{r} (t - \lambda_i)^{a_i} g(t)$$

with $r \geqslant 0$, $a_1, \ldots, a_r \geqslant 1$, $\lambda_1, \ldots, \lambda_r \in \mathbf{F}$ and $g(t) \in \mathbf{F}(t)$ with no roots in $\mathbf{F}$.

Lemma. *A polynomial $f \in \mathbf{F}[t]$ of degree $n \geqslant 0$ has at most n roots counted with multiplicity.*

Corollary. *Suppose $f, g \in \mathbf{F}[t]$ have degrees $< n$ and $f(t_i) = g(t_i)$ for $t_1, \ldots, t_n \in \mathbf{F}$ distinct. Then $f = g$.*

Proof. Consider $f - g$ which has degree $< n$ but at least n roots, namely $t_1, \ldots, t_n$. Thus $\deg(f - g) = -\infty$ and so $f = g$. $\square$

Theorem (Fundamental Theorem of Algebra)**.** *Every polynomial $f \in \mathbf{C}[t]$ of degree at least 1 has a root in $\mathbf{C}$.*

It follows that $f \in \mathbf{C}[t]$ has precisely n roots in $\mathbf{C}$ counted with multiplicity.

4.1.2. *Minimal polynomials.*

Notation. Given $f(t) = \sum_{i=0}^{m} a_i t^i \in \mathbf{F}[t]$, $A \in \mathrm{Mat}_n(\mathbf{F})$ and $\alpha \in \mathrm{End}(V)$ we write

$$f(A) := \sum_{i=0}^{m} a_i A^i$$

and

$$f(\alpha) := \sum_{i=0}^{m} a_i \alpha^i.$$

Here $A^0 = I_n$ and $\alpha^0 = \iota$.

Theorem. *Suppose that $\alpha \in \mathrm{End}(V)$. Then α is diagonalisable if and only if there is a non-zero polynomial $p(t) \in \mathbf{F}[t]$ that can be expressed as a product of distinct linear factors such that $p(\alpha) = 0$.*

Proof. Suppose that α is diagonalisable and $\lambda_1, \dots, \lambda_k \in \mathbf{F}$ are the distinct eigenvalues of α. Thus if v is an eigenvector for α then $\alpha(v) = \lambda_i v$ for some $i = 1, \dots, k$. Let $p(t) = \prod_{i=1}^k (t - \lambda_i)$

Since α is diagonalisable there is a basis $e_1, \dots, e_n$ for V such that each e_r is an eigenvector of α with eigenvalue λ_{j_r}, say. Then $p(\alpha)(e_r) = \prod_{i=1}^k (\lambda_{j_r} - \lambda_i) e_r = 0$ since $\lambda_{j_r} = \lambda_i$ for some $i \in \{1, \dots, k\}$. Thus $p(\alpha)(v) = 0$ for all v in a basis for V and so $p(\alpha) = 0 \in \operatorname{End}(V)$.

Conversely, if $p(\alpha) = 0$ for $p(t) = \prod_{i=1}^k (t - \lambda_i)$ for $\lambda_1, \dots, \lambda_k \in \mathbf{F}$ distinct — note that without loss of generality we may assume p has leading coefficient equal to 1. We will show that $V = \bigoplus_{i=1}^k E(\lambda_i)$. Since the sum of eigenspaces is always direct it suffices to show that every element $v \in V$ can be written as a sum of eigenvectors.

Let

$$p_j(t) := \prod_{\substack{i=1 \\ i \neq j}}^{k} \frac{(t - \lambda_i)}{(\lambda_j - \lambda_i)}$$

for $j = 1, \dots, k$. Thus $p_j(\lambda_i) = \delta_{ij}$.

Thus $\sum_{j=1}^k p_j(t)$ is a polynomial of degree at most $k - 1$ such that $p_j(\lambda_i) = 1$ for each $i = 1, \dots, k$. It follows that $\sum_{j=1}^k p_j(t) = 1$.

Let $\Pi_j \colon V \to V$ be defined by $\Pi_j = p_j(\alpha)$. Then $\sum \Pi_j = \sum p_j(\alpha) = \iota$.

Let $v \in V$. Then $v = \iota(v) = \sum \Pi_j(v)$. But $(\alpha - \lambda_j \iota) p_j(\alpha) = \frac{p(\alpha)}{\prod_{i \neq j}(\lambda_j - \lambda_i)} v = 0$. Thus $\Pi_j(v) \in \ker(\alpha - \lambda_j \iota) = E(\lambda_j)$ and we're done. $\square$

Remark. In the above proof, if $v \in E(\lambda_i)$ then $\Pi_j(v) = p_j(\lambda_i) v = \delta_{ij} v$. So Π_j is a projection onto $E(\lambda_j)$ along $\oplus_{i \neq j} E(\lambda_i)$.

Definition. The *minimal polynomial* of $\alpha \in \operatorname{End}(V)$ is the non-zero monic polynomial (i.e. leading coefficient is 1) $m_\alpha(t)$ of least degree such that $m_\alpha(\alpha) = 0$.

Of course we can define the minimal polynomial of a square matrix in a similar fashion.

Note that if $\dim V = n < \infty$ then $\dim \operatorname{End}(V) = n^2$, so $\iota, \alpha, \alpha^2, \dots, \alpha^{n^2}$ are linearly dependent since there are $n^2 + 1$ of them. Thus there is some non-trivial linear equation $\sum_{i=0}^{n^2} a_i \alpha^i = 0$. i.e. there is a non-zero polynomial $p(t)$ of degree at most n^2 such that $p(\alpha) = 0$.

Lemma. *Let $\alpha \in \operatorname{End}(V)$, $p \in \mathbf{F}[t]$ then $p(\alpha) = 0$ if and only if $m_\alpha(t)$ is a factor of $p(t)$. In particular $m_\alpha(t)$ is well-defined.*

Proof. We can find $q, r \in \mathbf{F}[t]$ such that $p(t) = q(t) m_\alpha(t) + r(t)$ with $\deg r < \deg m_\alpha$. Then $p(\alpha) = q(\alpha) m_\alpha(\alpha) + r(\alpha) = 0 + r(\alpha)$. Thus $p(\alpha) = 0$ if and only if $r(\alpha) = 0$. But the minimality of the degree of m_α means that $r(\alpha) = 0$ if and only if $r = 0$ ie if and only if m_α is a factor of p.

Now if m_1, m_2 are both minimal polynomials for α then m_1 divides m_2 and m_2 divides m_1 so as both are monic $m_2 = m_1$. $\square$

Lecture 11

Note that if A and B are similar matrices; so $B = P^{-1} A P$ say, then for any polynomial $p(t) \in \mathbf{F}[t]$ we can compute $p(B) = p(P^{-1} A P) = P^{-1} p(A) P$. So as 0 is only similar to itself we see that $p(B) = 0$ if and only if $p(A) = 0$. Thus similar matrices have the same minimal polynomial.

Another way to see that is to observe that $m_A = m_\alpha$ for any endomorphism α such that A represents α with respect to some basis. But then if B is similar to A, B must also represent α with respect to some (other) basis. Thus $m_B = m_\alpha = m_A$.

Example. If $V = \mathbf{F}^2$ then

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

both have characteristic polynomial $(t-1)^2$ but only the first one is diagonalisable so they cannot be similar. One can see that $m_A(t) = t - 1$ but $m_B(t) = (t-1)^2$ so minimal polynomials distinguish these two similarity classes.

Theorem (Diagonalisability Theorem)**.** *Let $\alpha \in \mathrm{End}(V)$ then α is diagonalisable if and only if $m_\alpha(t)$ has distinct linear factors.*

Proof. If α is diagonalisable there is some polynomial $p(t)$ with distinct linear factors such that $p(\alpha) = 0$ then m_α divides $p(t)$ so must also have distinct linear factors. The converse is already proven. □

Theorem. *Let $\alpha, \beta \in \mathrm{End}(V)$ be diagonalisable. Then α, β are simultaneously diagonalisable (i.e. there is a single basis with respect to which the matrices representing α and β are both diagonal) if and only if α and β commute.*

Proof. Certainly if there is a basis $\langle e_1, \ldots, e_n \rangle$ such that α and β are represented by diagonal matrices, A and B respectively, then α and β commute since A and B commute and $\alpha\beta$ is represented by AB and $\beta\alpha$ by BA.

For the converse, suppose that α and β commute. Let $\lambda_1, \ldots, \lambda_k$ denote the distinct eigenvalues of α and let $E_i = E_\alpha(\lambda_i)$ for $i = 1, \ldots, k$. Then as α is diagonalisable we know that $V = \bigoplus_{i=1}^k E_i$.

We claim that $\beta(E_i) \subset E_i$ for each $i = 1, \ldots, k$. To see this, suppose that $v \in E_i$ for some such i. Then

$$\alpha\beta(v) = \beta\alpha(v) = \beta(\lambda_i v) = \lambda_i \beta(v)$$

and so $\beta(v) \in E_i$ as claimed. Thus we can view $\beta|_{E_i}$ as an endomorphism of E_i.

Now since β is diagonalisable, the minimal polynomial m_β of β has distinct linear factors. But $m_\beta(\beta|_{E_i}) = m_\beta(\beta)|_{E_i} = 0$. Thus $\beta|_{E_i}$ is diagonalisable for each E_i and we can find B_i a basis of E_i consisting of eigenvectors of β. Then $B = \bigcup_{i=1}^k B_i$ is a basis for V. Moreover α and β are both diagonal with respect to this basis. □

Remark. By slightly adapting the proof we can extend this to show that any set of diagonalisable endomorphisms of V is simultaneously diagonalisable precisely if they commute pairwise.

4.2. **The Cayley-Hamilton Theorem.**

Definition. $\alpha \in \mathrm{End}(V)$ is *triangulable* if there is a basis for V such that the corresponding matrix is upper triangular.

Recall that the characteristic polynomial of an endomorphism $\alpha \in \mathrm{End}(V)$ is defined by $\chi_\alpha(t) = \det(t\iota - \alpha)$.

Lemma. *A linear map α is triangulable if and only if $\chi_\alpha(t)$ can be written as a product of linear factors.*

Proof. Suppose that α is triangulable and is represented by

$$\begin{pmatrix} a_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & a_n \end{pmatrix}$$

with respect to some basis. Then

$$\begin{aligned} \chi_\alpha(t) &= \det\left(tI_n - \begin{pmatrix} a_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & a_n \end{pmatrix} \right) \\ &= \prod (t - a_i). \end{aligned}$$

Thus χ_α is a product of linear factors.

We'll prove the converse by induction on $n = \dim V$. If $n = 1$ every matrix is triangulable. Suppose that $n > 1$ and the result holds for all endomorphisms of spaces of smaller dimension. By hypothesis $\chi_\alpha(t)$ has a root $\lambda \in \mathbf{F}$. Let $U = E(\lambda) \neq 0$. Let W be a vector space complement for U in V. Let $u_1, \ldots, u_r$ be a basis for U and $w_{r+1}, \ldots, w_n$ a basis for W so that $u_1, \ldots, u_r, w_{r+1}, \ldots, w_n$ is a basis for V. Then α is represented by a matrix of the form

$$\begin{pmatrix} \lambda I_r & * \\ 0 & B \end{pmatrix}.$$

Moreover because this matrix is block triangular we know that

$$\chi_\alpha(t) = \chi_{\lambda I_r}(t)\chi_B(t).$$

Thus as χ_α is a product of linear factors χ_B must be also. Let β be the linear map $W \to W$ defined by B. (**Warning:** β is not just $\alpha|_W$ in general. However it is true that $(\beta - \alpha)(w) \in U$ for all $w \in W$.) Since $\dim W < \dim V$ there is another basis $v_{r+1}, \ldots, v_n$ for W such that the matrix C representing β is upper-triangular. Since for each $j = 1, \ldots, n-r$, $\alpha(v_{j+r}) = u'_j + \sum_{k=1}^{n-r} C_{kj} v_k$ for some $u'_j \in U$, the matrix representing α with respect to the basis $u_1, \ldots, u_r, v_{r+1}, \ldots, v_n$ is of the form

$$\begin{pmatrix} \lambda I_r & * \\ 0 & C \end{pmatrix}$$

which is upper triangular. □

Thus by the Fundamental Theorem of Algebra every endomorphism of a f.d. complex vector space is triangulable.

Corollary. *Every $A \in \mathrm{Mat}_n(\mathbf{C})$ is similar to an upper triangular matrix.*

Example. The real matrix

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

is not similar to an upper triangular matrix over $\mathbf{R}$ for $\theta \notin \pi\mathbf{Z}$ since its eigenvalues are $e^{\pm i\theta} \notin \mathbf{R}$. Of course it is similar to a diagonal matrix over $\mathbf{C}$.

Lecture 12

Theorem (Cayley–Hamilton Theorem)**.** *Suppose that V is a f.d. vector space over $\mathbf{F}$ and $\alpha \in \operatorname{End}(V)$. Then $\chi_\alpha(\alpha) = 0$. In particular m_α divides χ_α (and so $\deg m_\alpha \leqslant \dim V$).*

Remarks.

(1) It is tempting to substitute '$t = A$' into $\chi_A(t) = \det(tI_n - A)$ but it is not possible to make sense of this.

(2) If $p(t) \in \mathbf{F}[t]$ and

$$A = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix}$$

is diagonal then

$$p(A) = \begin{pmatrix} p(\lambda_1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & p(\lambda_n) \end{pmatrix}.$$

So as $\chi_A(t) = \prod_{i=1}^n (t - \lambda_i)$ we see $\chi_A(A) = 0$. So Cayley–Hamilton is obvious when α is diagonalisable.

Proof of Cayley–Hamilton when $\mathbf{F} = \mathbf{C}$. Since $\mathbf{F} = \mathbf{C}$ we've seen that there is a basis $\langle e_1, \ldots, e_n \rangle$ such that α is represented by an upper triangular matrix

$$A = \begin{pmatrix} \lambda_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_n \end{pmatrix}.$$

Then we can compute $\chi_\alpha(t) = \prod_{i=1}^n (t - \lambda_i)$. Let V_j be the span of $e_1, \ldots, e_j$ for $j = 0, \ldots, n$ so we have

$$0 = V_0 \subset V_1 \subset \cdots \subset V_{n-1} \subset V_n = V$$

with $\dim V_j = j$. Since $\alpha(e_i) = \sum_{k=1}^n A_{ki} e_k = \sum_{k=1}^i A_{ki} e_k$, we see that

$$\alpha(V_j) \subset V_j \text{ for each } j = 0, \ldots, n.$$

Moreover $(\alpha - \lambda_j \iota)(e_j) = \sum_{k=1}^{j-1} A_{kj} e_k$ so

$$(\alpha - \lambda_j \iota)(V_j) \subset V_{j-1} \text{ for each } j = 1, \ldots, n.$$

Thus we see inductively that $\prod_{i=j}^n (\alpha - \lambda_i \iota)(V_n) \subset V_{j-1}$. In particular

$$\prod_{i=1}^n (\alpha - \lambda_i \iota)(V) \subset V_0 = 0.$$

Thus $\chi_\alpha(\alpha) = 0$ as claimed. □

Remark. It is straightforward to extend this to the case $\mathbf{F} = \mathbf{R}$: since $\mathbf{R} \subset \mathbf{C}$, if $A \in \operatorname{Mat}_n(\mathbf{R})$ then we can view A as an element of $\operatorname{Mat}_n(\mathbf{C})$ to see that $\chi_A(A) = 0$. But then if $\alpha \in \operatorname{End}(V)$ for any vector space V over $\mathbf{R}$ we can take A to be the matrix representing α over some basis. Then $\chi_\alpha(\alpha) = \chi_A(\alpha)$ is represented by $\chi_A(A)$ and so it zero.

Second proof of Cayley–Hamilton. Let $A \in \mathrm{Mat}_n(\mathbf{F})$ and let $B = tI_n - A$. We can compute that $\operatorname{adj} B$ is an $n \times n$-matrix with entries elements of $\mathbf{F}[t]$ of degree at most $n-1$. So we can write

$$\operatorname{adj} B = B_{n-1}t^{n-1} + B_{n-2}t^{n-2} + \cdots + B_1 t + B_0$$

with each $B_i \in \mathrm{Mat}_n(\mathbf{F})$. Now we know that $B \operatorname{adj} B = \det B I_n = \chi_A(t) I_n$. ie

$$(tI_n - A)(B_{n-1}t^{n-1} + B_{n-2}t^{n-2} + \cdots + B_1 t + B_0) = (t^n + a_{n-1}t^{n-1} + \cdots + a_0)I_n$$

where $\chi_A(t) = t^n + a_{n-1}t^{n_1} + \cdots a_0$. Comparing coefficients in t^k for $k = n, \ldots, 0$ we see

$$\begin{aligned} B_{n-1} - 0 &= I_n \\ B_{n-2} - AB_{n-1} &= a_{n-1}I_n \\ B_{n-3} - AB_{n-2} &= a_{n-2}I_n \\ \cdots &= \cdots \\ 0 - AB_0 &= a_0 I_n \end{aligned}$$

Thus

$$\begin{aligned} A^n B_{n-1} - 0 &= A^n \\ A^{n-1}B_{n-2} - A^n B_{n-1} &= a_{n-1}A^{n-1} \\ A^{n-2}B_{n-3} - A^{n-1}B_{n-2} &= a_{n-2}A^{n-2} \\ \cdots &= \cdots \\ 0 - AB_0 &= a_0 I_n \end{aligned}$$

Summing we get $0 = \chi_A(A)$ as required. $\square$

Lemma. *Let $\alpha \in End(V)$, $\lambda \in \mathbf{F}$. Then the following are equivalent*

(a) λ is an eigenvalue of α;
(b) λ is a root of $\chi_\alpha(t)$;
(c) λ is a root of $m_\alpha(t)$.

Proof. λ is an eigenvalue of α if and only if $\ker(\alpha - \lambda\iota) \neq 0$ if and only if $\det(\alpha - \lambda\iota) = 0$ if and only if $\chi_\alpha(\lambda) = 0$. Thus (a) is equivalent to (b).

Suppose that λ is an eigenvalue of α. There is some $v \in V$ non-zero such that $\alpha v = \lambda v$. Then for any polynomial $p \in \mathbf{F}[t]$, $p(\alpha)v = p(\lambda)v$ so

$$0 = m_\alpha(\alpha)v = m_\alpha(\lambda)(v).$$

Since $v \neq 0$ it follows that $m_\alpha(\lambda) = 0$. Thus (a) implies (c).

Finally suppose that $m_\alpha(\lambda) = 0$. Then $m_\alpha(t) = (t - \lambda)g(t)$ for some $g \in \mathbf{F}[t]$. Since $\deg g < \deg m$, $g(\alpha) \neq 0$. Thus there is some $v \in V$ such that $g(\alpha)(v) \neq 0$. But then $(\alpha - \lambda\iota)\left(g(\alpha)(v)\right) = m_\alpha(\alpha)(v) = 0$. So $g(\alpha)(v) \neq 0$ is a λ-eigenvector of α. Thus (c) implies (a). $\square$

Note we could've used the Cayley–Hamilton Theorem to see that (c) implies (b).

Example. What is the minimal polynomial of

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

We can compute $\chi_A(t) = (t-1)^2(t-2)$. So by Cayley–Hamilton $m_\alpha(t)$ is a factor of $(t-1)^2(t-2)$. Moreover by the lemma it must be a multiple of $(t-1)(t-2)$. So m_A is one of $(t-1)(t-2)$ and $(t-1)^2(t-2)$.

We can compute

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

Thus $m_A(t) = (t-1)(t-2)$. Since this has distict roots, A is diagonalisable.

4.3. **Multiplicities of eigenvalues and Jordan Normal Form.**

Definition (Multiplicity of eigenvalues)**.** Suppose that $\alpha \in \mathrm{End}(V)$ and λ is an eigenvalue of α:

(a) the *algebraic multiplicity of* λ is

$$a_\lambda := \text{ the multiplicity of } \lambda \text{ as a root of } \chi_\alpha(t);$$

(b) the *geometric multiplicity of* λ is

$$g_\lambda := \dim E_\alpha(\lambda);$$

(c) another useful number is

$$c_\lambda := \text{ the multiplicity of } \lambda \text{ as a root of } m_\alpha(t).$$

Lemma. *Let $\alpha \in \mathrm{End}(V)$ and $\lambda \in \mathbf{F}$ an eigenvalue of α. Then*

(a) $1 \leqslant g_\lambda \leqslant a_\lambda$ and
(b) $1 \leqslant c_\lambda \leqslant a_\lambda$.

Proof. (a) By definition if λ is an eigenvalue of α then $E_\alpha(\lambda) \neq 0$ so $g_\lambda \geqslant 1$. Suppose that $v_1 \ldots, v_g$ is a basis for $E(\lambda)$ and extend it to a basis $v_1, \ldots, v_n$ for V. Then α is represented by a matrix of the form

$$\begin{pmatrix} \lambda I_g & * \\ 0 & B \end{pmatrix}.$$

Thus $\chi_\alpha(t) = \chi_{\lambda I_g}(t)\chi_B(t) = (t-\lambda)^g \chi_B(t)$. So $a_\lambda \geqslant g = g_\lambda$.

(b) We've seen that if λ is an eigenvalue of α then α is a root of $m_\alpha(t)$ so $c_\lambda \geqslant 1$. Cayley–Hamilton says $m_\alpha(t)$ divides $\chi_\alpha(t)$ so $c_\lambda \leqslant a_\lambda$. □

Examples.

(1) If $A = \begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix} \in \mathrm{Mat}_n(\mathbf{F})$ then $g_\lambda = 1$ and $a_\lambda = c_\lambda = n$.

(2) If $A = \lambda I$ then $g_\lambda = a_\lambda = n$ and $c_\lambda = 1$.

Lecture 13

Lemma. *Suppose that* $\mathbf{F} = \mathbf{C}$ *and* $\alpha \in \mathrm{End}(V)$. *Then the following are equivalent:*
(a) α *is diagonalisable;*
(b) $a_\lambda = g_\lambda$ *for all eigenvalues* λ *of* α*;*
(c) $c_\lambda = 1$ *for all eigenvalues* λ *of* α*.*

Proof. To see that (a) is equivalent to (b) suppose that the distict eigenvalues of α are $\lambda_1, \ldots, \lambda_k$. Then α is diagonalisable if and only if $\dim V = \sum_{i=1}^k \dim E(\lambda_i) = \sum_{i=1}^n g_{\lambda_i}$. But $g_\lambda \leqslant a_\lambda$ for each eigenvalue λ and $\sum_{i=1}^k a_{\lambda_i} = \deg \chi_\alpha = \dim V$ by the Fundamental Theorem of Algebra. Thus α is diagonalisable if and only if $g_{\lambda_i} = a_{\lambda_i}$ for each $i = 1, \ldots, k$.

Since by the Fundamental Theorem of Algebra for any such α, $m_\alpha(t)$ may be written as a product of linear factors, α is diagonalisable if and only if these factors are distinct. This is equivalent to $c_\lambda = 1$ for every eigenvalue λ of α. $\square$

Remark. Let A be a block diagonal square matrix; ie

$$A = \begin{pmatrix} A_1 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & A_k \end{pmatrix}$$

then $\chi_A(t) = \prod_{i=1}^k \chi_{A_i}(t)$. Moreover, if $p \in \mathbf{F}[t]$ then

$$p(A) = \begin{pmatrix} p(A_1) & 0 & 0 & 0 \\ 0 & p(A_2) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & p(A_k) \end{pmatrix}$$

so $m_A(t)$ is the lowest common multiple of $m_{A_1}(t), \ldots, m_{A_k}(t)$.

We even have $n(p(A)) = \sum_{i=1}^k n(p(A_i))$ for any $p \in \mathbf{F}[t]$.

Definition. We say that a matrix $A \in \mathrm{Mat}_n(\mathbf{C})$ is in *Jordan Normal Form (JNF)* if it is a block diagonal matrix

$$A = \begin{pmatrix} J_{n_1}(\lambda_1) & 0 & 0 & 0 \\ 0 & J_{n_2}(\lambda_2) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & J_{n_k}(\lambda_k) \end{pmatrix}$$

where $k \geqslant 1$, $n_1, \ldots, n_k \in \mathbb{N}$ such that $\sum_{i=1}^k n_i = n$ and $\lambda_1, \ldots, \lambda_k \in \mathbf{C}$ (not necessarily distinct) and $J_m(\lambda) \in \mathrm{Mat}_m(\mathbf{C})$ has the form

$$J_m(\lambda) := \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & \ddots & 0 \\ 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}.$$

We call the $J_m(\lambda)$ *Jordan blocks*

Note $J_m(\lambda) = \lambda I_m + J_m(0)$.

Examples.

$$J_1(\lambda) = \begin{pmatrix} \lambda \end{pmatrix},$$

$$J_2(\lambda) = \begin{pmatrix} \lambda & 1 \\ 0 & 1 \end{pmatrix},$$

$$J_3(\lambda) = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}.$$

Theorem (Jordan Normal Form)**.** *Every matrix $A \in \mathrm{Mat}_n(\mathbf{C})$ is similar to a matrix in JNF. Moreover this matrix in JNF is uniquely determined by A up to reordering the Jordan blocks.*

Remarks.

(1) Of course, we can rephrase this as whenever α is an endomorphism of a f.d. $\mathbf{C}$-vector space V, there is a basis of V such that α is represented by a matrix in JNF. Moreover, this matrix is uniquely determined by α up to reordering the Jordan blocks.

(2) Two matrices in JNF that differ only in the ordering of the blocks are similar. A corresponding basis change arises as a reordering of the basis vectors.

(3) $A \in \mathrm{Mat}_n(\mathbf{C})$ is diagonalisable if and only if all Jordan blocks have size one if and only if $a_\lambda = g_\lambda$ for all eigenvalues λ of A if and only if $c_\lambda = 1$ for all eigenvalues λ of A.

Examples.

(1) Every 2×2 matrix in JNF is of the form $\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$ with $\lambda \neq \mu$ or $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$ or $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$. The minimal polynomials are $(t-\lambda)(t-\mu)$, $(t-\lambda)$ and $(t-\lambda)^2$ respectively. The characteristic polynomials are $(t-\lambda)(t-\mu)$, $(t-\lambda)^2$ and $(t-\lambda)^2$ respectively. Thus we see that the JNF is determined by the minimal polynomial of the matrix in this case (but not by just the characteristic polynomial).

(2) Suppose now that λ_1, λ_2 and λ_3 are distinct complex numbers. Then every 3×3 matrix in JNF is one of six forms

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}, \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}, \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_2 \end{pmatrix}$$

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix}, \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{pmatrix} \text{ and } \begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{pmatrix}.$$

The minimal polynomials are $(t-\lambda_1)(t-\lambda_2)(t-\lambda_3)$, $(t-\lambda_1)(t-\lambda_2)$, $(t-\lambda_1)(t-\lambda_2)^2$, $(t-\lambda_1)$, $(t-\lambda_1)^2$ and $(t-\lambda_1)^3$ respectively. The characteristic polynomials are $(t-\lambda_1)(t-\lambda_2)(t-\lambda_3)$, $(t-\lambda_1)(t-\lambda_2)^2$, $(t-\lambda_1)(t-\lambda_2)^2$, $(t-\lambda_1)^3$, $(t-\lambda_1)^3$ and $(t-\lambda_1)^3$ respectively. So in this case the minimal polynomial does not determine the JNF by itself but the minimal and characteristic polynomials together do determine the JNF. In general even these two bits of data together don't suffice to determine everything.

We recall that

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & \ddots & 0 \\ 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}.$$

Thus if $\langle e_1, \dots, e_n \rangle$ is the standard basis for $\mathbb{C}^n$ we can compute $(J_n(\lambda) - \lambda I_n)e_1 = 0$ and $(J_n(\lambda) - \lambda I_n)e_i = e_{i-1}$ for $1 < i \leqslant n$. Thus $(J_n(\lambda) - \lambda I_n)^k$ maps $e_1, \dots, e_k$ to 0 and e_{k+j} to e_j for $1 \leqslant j \leqslant n-k$. That is

$$(J_n(\lambda) - \lambda I_n)^k = \begin{pmatrix} 0 & I_{n-k} \\ 0 & 0 \end{pmatrix} \text{ for } k < n$$

and $(J_n(\lambda) - \lambda I_n)^k = 0$ for $k \geqslant n$.

Thus if $A = J_n(\lambda)$ is a single Jordan block, then $\chi_A(t) = m_A(t) = (t-\lambda)^n$, so λ is the only eigenvalue of A. Moreover $\dim E(\lambda) = 1$. Thus $a_\lambda = c_\lambda = n$ and $g_\lambda = 1$.

In general a_λ is the sum of the sizes of the blocks with eigenvalue λ which is the same as the number of λs on the diagonal. g_λ is the number of blocks with eigenvalue λ and c_λ is the size of the largest block with eigenvalue λ.

Theorem. *If $\alpha \in \mathrm{End}(V)$ and A in JNF represents α with respect to some basis then the number of Jordan blocks $J_n(\lambda)$ of A with eigenvalue λ and size $n \geqslant k \geqslant 1$ is given by*

$$|\{\textit{Jordan blocks } J_n(\lambda) \textit{ in } A \textit{ with } n \geqslant k\}| = n\left((\alpha - \lambda\iota)^k\right) - n\left((\alpha - \lambda\iota)^{k-1}\right)$$

Proof. We work blockwise. We can compute that

$$n\left((J_m(\lambda) - \lambda I_n)^k\right) = \min(m,k)$$

and

$$n\left((J_m(\mu) - \lambda I_n)^k\right) = 0$$

when $\mu \neq \lambda$.

Adding up for each block we get for $k \geqslant 1$

$$\begin{aligned} n\left((\alpha - \lambda\iota)^k\right) - n\left((\alpha - \lambda\iota)^{k-1}\right) &= n\left((A - \lambda I)^k\right) - n\left((A - \lambda I)^{k-1}\right) \\ &= \sum_{\substack{i=1 \\ \lambda_i = \lambda}}^{k} (\min(k, n_i) - \min(k-1, n_i) \\ &= |\{1 \leqslant i \leqslant k \mid \lambda_i = \lambda, n_i \geqslant k\} \\ &= |\{\text{Jordan blocks } J_n(\lambda) \text{ in } A \text{ with } n \geqslant k\}| \end{aligned}$$

as required. □

Because these nullities are basis-invariant, it follows that if it exists then the Jordan normal form representing α is unique up to reordering the blocks as claimed.

Lecture 14

Theorem (Jordan Normal Form)**.** *Let V be a f.d. $\mathbf{C}$-vector space, and $\alpha \in \mathrm{End}(V)$. Then there is a basis for V such that α is represented by a block diagonal matrix of* *the form*

$$A = \begin{pmatrix} J_{n_1}(\lambda_1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & J_{n_k}(\lambda_k) \end{pmatrix}$$

where $k \geqslant 1$, $n_1, \ldots, n_k \in \mathbb{N}$ *such that* $\sum_{i=1}^k n_i = \dim V$ *and* $\lambda_1, \ldots, \lambda_k \in \mathbf{C}$ *(not necessarily distinct). Moreover, this matrix is uniquely determined by* α *up to reordering the blocks.*

Theorem (Generalised eigenspace decompostion)**.** *Let* V *be a f.d.* $\mathbb{C}$*-vector space and* $\alpha \in \operatorname{End}(V)$*. Suppose that*

$$m_\alpha(t) = (t - \lambda_1)^{c_1} \cdots (t - \lambda_k)^{c_k}$$

with $\lambda_1, \ldots, \lambda_k$ *distinct. Then*

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

where $V_j = \ker((\alpha - \lambda_j)^{c_j})$ *is an* α*-invariant subspace (called a generalised eigenspace).*

Note that in the case $c_1 = c_2 = \cdots = c_k = 1$ we recover the diagonalisability theorem.

Sketch of proof. Let $p_j(t) = \prod_{\substack{i \neq j \\ i=1}}^k (t - \lambda_i)^{c_i}$. Then $p_1, \ldots, p_k$ have no common factor i.e. they are coprime. Thus by Euclid's algorithm we can find $q_1, \ldots, q_k \in \mathbf{C}[t]$ such that $\sum_{i=1}^k q_i p_i = 1$.

Let $\Pi_j = q_j(\alpha) p_j(\alpha)$ for $j = 1, \ldots, k$. Then $\sum_{j=1}^k \Pi_j = \iota$. Since $m_\alpha(\alpha) = 0$, $(\alpha - \lambda_j)^{c_j} \Pi_j = 0$, thus $\operatorname{Im} \Pi_j \subset V_j$.

Suppose that $v \in V$ then

$$v = \iota(v) = \sum_{j=1}^k \Pi_j(v) \in \sum V_j.$$

Thus $V = \sum V_j$.

But $\Pi_i \Pi_j = 0$ for $i \neq j$ and so $\Pi_i = \Pi_i(\sum_{j=1}^k \Pi_j) = \Pi_i^2$ for $1 \leqslant i \leqslant k$. Thus $\Pi_j|_{V_j} = \iota_{V_j}$ and if $v = \sum v_j$ with $v_j \in V_j$ then $v_j = \Pi_j(v)$. So $V = \bigoplus V_j$ as claimed. $\square$

Using this theorem we can, by restricting to its generalised eigenspaces, reduce the proof of the existence of Jordan normal form for α to the case that it has only one eigenvalue λ. By considering $(\alpha - \lambda\iota)$ we can even reduce to the case that 0 is the only eigenvalue.

Definition. We say that $\alpha \in \operatorname{End}(V)$ is *nilpotent* if there is some $k \geqslant 0$ such that $\alpha^k = 0$.

Note that α is nilpotent if and only if $m_\alpha(t) = t^k$ for some $1 \leqslant k \leqslant n$. When $\mathbf{F} = \mathbf{C}$ this is equivalent to 0 being the only eigenvalue for α.

Example. Let

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

Find an invertible matrix P such that $P^{-1}AP$ is in JNF.

First we compute the eigenvalues of A:

$$\chi_A(t) = \det \begin{pmatrix} t-3 & 2 & 0 \\ -1 & t & 0 \\ -1 & 0 & t-1 \end{pmatrix} = (t-1)(t(t-3)+2) = (t-1)^2(t-2).$$

Next we compute the eigenspaces

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

which has rank 2 and kernel spanned by $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Thus $E_A(1) = \left\langle \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle$. Similarly

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

also has rank 1 and kernel spanned by $\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$ thus $E_A(2) = \left\langle \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} \right\rangle$. Since $\dim E_A(1) + \dim E_A(2) = 2 < 3$, A is not diagonalisable. Thus

$$m_A(t) = \chi_A(t) = (t-1)^2(t-2)$$

and the JNF of A is

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

So we want to find a basis $\langle v_1, v_2, v_3 \rangle$ such that $Av_1 = v_1$, $Av_2 = v_1 + v_2$ and $Av_3 = 2v_3$ or equivalently $(A-I)v_2 = v_1$, $(A-I)v_1 = 0$ and $(A-2I)v_3 = 0$. Note that under these conditions $(A-I)^2 v_2 = 0$ but $(A-I)v_2 \neq 0$.

We compute

$$(A-I)^2 = \begin{pmatrix} 2 & -2 & 0 \\ 1 & -1 & 0 \\ 2 & -2 & 0 \end{pmatrix}$$

Thus

$$\ker(A-I)^2 = \left\langle \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle$$

Take $v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $v_1 = (A-I)v_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ and $v_3 = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$. Then these are LI so form a basis for $\mathbf{C}^3$ and if we take P to have columns v_1, v_2, v_3 we see that $P^{-1}AP = J$ as required.

5. Duality

5.1. **Dual spaces.** To specify a subspace of $\mathbf{F}^n$ we can write down a set of linear equations that every vector in the space satisfies. For example if $U = \left\langle \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right\rangle \subset \mathbf{F}^3$ we can see that

$$U = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} : 2x_1 - x_2 = 0, x_1 - x_3 = 0 \right\}.$$

These equations are determined by linear maps $\mathbf{F}^n \to \mathbf{F}$. Moreover if $\theta_1, \theta_2 \colon \mathbf{F}^n \to \mathbf{F}$ are linear maps that vanish on U and $\lambda, \mu \in \mathbf{F}$ then $\lambda\theta_1 + \mu\theta_2$ vanishes on U. Since the 0 map vanishes on evey subspace, one may study the subspace of linear maps $\mathbf{F}^n \to \mathbf{F}$ that vanish on U.

Definition. Let V be a vector space over $\mathbf{F}$. The *dual space* of V is the vector space

$$V^* := \mathcal{L}(V, \mathbf{F}) = \{\alpha \colon V \to \mathbf{F} \text{ linear}\}$$

with pointwise addition and scalar mulitplication. The elements of V^* can called *linear forms* or *linear functionals* on V.

Examples.

(a) $V = \mathbf{R}^3$, $\theta \colon V \to \mathbf{R}$; $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto x_3 - x_1$.

(b) $V = C([0,1], \mathbf{R})$, then $V \to \mathbf{R}$; $f \mapsto \int_0^1 f(t)dt \in V^*$.

Lemma. *Suppose that V is a f.d. vector space over $\mathbf{F}$ with basis $\langle e_1, \ldots, e_n \rangle$. Then V^* has a basis $\langle \epsilon_1, \ldots, \epsilon_n \rangle$ such that $\epsilon_i(e_j) = \delta_{ij}$.*

Definition. We call the basis $\langle \epsilon_1, \ldots, \epsilon_n \rangle$ the *dual basis* of V^* with respect to $\langle e_1, \ldots, e_n \rangle$.

Proof of Lemma. We know that to define a linear map it suffices to define it on a basis so there are unique elements $\epsilon_1, \ldots, \epsilon_n$ such that $\epsilon_i(e_j) = \delta_{ij}$. We must show that they span and are LI.

Suppose that $\theta \in V^*$ is any linear map. Then let $\lambda_i = \theta(e_i) \in \mathbf{F}$. We claim that $\theta = \sum_{i=1}^n \lambda_i \epsilon_i$. It suffices to show that the two elements agree on the basis $e_1, \ldots, e_n$ of V. But $\sum_{i=1}^n \lambda_i \epsilon_i(e_j) = \lambda_j = \theta(e_j)$. So the claim is true that $\epsilon_1, \ldots, \epsilon_n$ do span V^*.

Next, suppose that $\sum \mu_i \epsilon_i = 0 \in V^*$ for some $\mu_1, \ldots, \mu_n \in \mathbf{F}$. Then $0 = \sum \mu_i \epsilon_i(e_j) = \mu_j$ for each $j = 1, \ldots, n$. Thus $\epsilon_1, \ldots, \epsilon_n$ are LI as claimed. □

Lecture 15

Recall the following lemma/definition.

Lemma. *Suppose that V is a f.d. vector space over $\mathbf{F}$ with basis $\langle e_1, \ldots, e_n \rangle$. Then V^* has a basis $\langle \epsilon_1, \ldots, \epsilon_n \rangle$ called the dual basis with respect to $\langle e_1, \ldots, e_n \rangle$ such that $\epsilon_i(e_j) = \delta_{ij}$.* □

Corollary. *If V is f.d. then* $\dim V^* = \dim V$. □

Proposition. *Suppose that V is a f.d. vector space over $\mathbf{F}$ with bases $\langle e_1, \ldots, e_n\rangle$ and $\langle f_1, \ldots, f_n\rangle$ such that P is the change of basis matrix from $\langle e_1, \ldots, e_n\rangle$ to $\langle f_1, \ldots, f_n\rangle$ i.e. $f_i = \sum_{k=1}^n P_{ki} e_k$ for $1 \leqslant i \leqslant n$.*

Let $\langle \epsilon_1, \ldots, \epsilon_n\rangle$ and $\langle \eta_1, \ldots, \eta_n\rangle$ be the corresponding dual bases so that

$$\epsilon_i(e_j) = \delta_{ij} = \eta_i(f_j) \text{ for } 1 \leqslant i, j \leqslant n.$$

Then the change of basis matrix from $\langle \epsilon_1, \ldots, \epsilon_n\rangle$ to $\langle \eta_1, \ldots, \eta_n\rangle$ is given by $(P^{-1})^T$ ie $\epsilon_i = \sum P^T_{li}\eta_l$. .

Proof. Let $Q = P^{-1}$. Then $e_j = \sum Q_{kj} f_k$, so we can compute

$$(\sum_l P_{il}\eta_l)(e_j) = \sum_{k,l}(P_{il}\eta_l)(Q_{kj} f_k) = \sum_{k,l} P_{il}\delta_{kl}Q_{kj} = \delta_{ij}.$$

Thus $\epsilon_i = \sum_l P_{il}\eta_l$ as claimed. □

Remark. If we think of elements of V as column vectors with respect to some basis

$$\sum x_i e_i = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

then we can view elements of V^* as row vectors with respect to the dual basis

$$\sum a_i \epsilon_i = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}.$$

Then

$$\left(\sum a_i\epsilon_i\right)\left(\sum x_j e_j\right) = \sum a_i x_i = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

Definition.

(a) If $U \subset V$ then the *annihilator* of U, $U^\circ := \{\theta \in V^* \mid \theta(u) = 0 \quad \forall u \in U\} \subset V^*$.

(b) If $W \subset V^*$, then the *annihilator* of $W^\circ := \{v \in V \mid \theta(v) = 0 \quad \forall \theta \in W\} \subset V$.

Example. Consider $\mathbf{R}^3$ with standard basis $\langle e_1, e_2, e_3\rangle$ and $(\mathbf{R}^3)^*$ with dual basis $\langle \epsilon_1, \epsilon_2, \epsilon_3\rangle$, $U = \langle e_1 + 2e_2 + e_3\rangle \subset \mathbf{R}^3$ and $W = \langle \epsilon_1 - \epsilon_3, \epsilon_1 - 2\epsilon_2\rangle \subset (\mathbf{R}^3)^*$. Then $U^\circ = W$ and $W^\circ = U$.

Proposition. *Suppose that V is f.d. over $\mathbf{F}$ and $U \subset V$ is a subspace. Then*

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

Proof 1. Let $\langle e_1, \ldots, e_k\rangle$ be a basis for U and extend to a basis $\langle e_1, \ldots, e_n\rangle$ for V and consider the dual basis $\langle \epsilon_1, \ldots, \epsilon_n\rangle$ for V^*.

We claim that U° is spanned by $\epsilon_{k+1}, \ldots, \epsilon_n$.

Certainly if $j > k$, then $\epsilon_j(e_i) = 0$ for each $1 \leqslant i \leqslant k$ and so $\epsilon_j \in U^\circ$. Suppose now that $\theta \in U^\circ$. We can write $\theta = \sum_{i=1}^n \lambda_i\epsilon_i$ with $\lambda_i \in \mathbf{F}$. Now,

$$0 = \theta(e_j) = \lambda_j \text{ for each } 1 \leqslant j \leqslant k.$$

So $\theta = \sum_{j=k+1}^n \lambda_i\epsilon_i$. Thus U° is the span of $\epsilon_{k+1}, \ldots, \epsilon_n$ and

$$\dim U^\circ = n - k = \dim V - \dim U$$

as claimed. □

Proof 2. Consider the restriction map $V^* \to U^*$ given by $\theta \mapsto \theta|_U$. Since every linear map $U \to \mathbf{F}$ can be extended to a linear map $V \to F$ this map is a linear surjection. Moreover its kernel is U°. Thus $\dim V^* = \dim U^* + \dim U^\circ$ by the rank-nullity theorem. The proposition follows from the statements $\dim U = \dim U^*$ and $\dim V = \dim V^*$. □

5.2. **Dual maps.**

Definition. Let V and W be vector spaces over $\mathbf{F}$ and suppose that $\alpha\colon V \to W$ is a linear map. The *dual map* to α is the map $\alpha^*\colon W^* \to V^*$ is given by $\theta \mapsto \theta\alpha$.

Note that $\theta\alpha$ is the composite of two linear maps and so is linear. Moreover, if $\lambda, \mu \in \mathbf{F}$ and $\theta_1, \theta_2 \in W^*$ and $v \in V$ then

$$\begin{aligned} \alpha^*(\lambda\theta_1 + \mu\theta_2)(v) &= (\lambda\theta_1 + \mu\theta_2)\alpha(v) \\ &= \lambda\theta_1\alpha(v) + \mu\theta_2\alpha(v) \\ &= \left(\lambda\alpha^*(\theta_1) + \mu\alpha^*(\theta_2)\right)(v). \end{aligned}$$

Therefore $\alpha^*(\lambda\theta_1 + \mu\theta_2) = \lambda\alpha^*(\theta_1) + \mu\alpha^*(\theta_2)$ and α^* is linear ie $\alpha^* \in \mathcal{L}(W^*, V^*)$.

Lemma. *Suppose that V and W are f.d. with bases $\langle e_1, \ldots, e_n\rangle$ and $\langle f_1, \ldots, f_m\rangle$ respectively. Let $\langle \epsilon_1, \ldots, \epsilon_n\rangle$ and $\langle \eta_1, \ldots, \eta_m\rangle$ be the corresponding dual bases. Then if $\alpha\colon V \to W$ is represented by A with respect to $\langle e_1, \ldots, e_n\rangle$ and $\langle f_1, \ldots, f_m\rangle$ then α^* is represented by A^T with respect to $\langle \epsilon_1, \ldots, \epsilon_n\rangle$ and $\langle \eta_1, \ldots, \eta_m\rangle$.*

Proof. We're given that $\alpha(e_i) = \sum A_{ji} f_j$ and must compute $\alpha^*(\eta_i)$ in terms of $\epsilon_1, \ldots, \epsilon_n$.

$$\begin{aligned} \alpha^*(\eta_i)(e_k) &= \eta_i(\alpha(e_k)) \\ &= \eta_i(\sum_j A_{jk} f_j) \\ &= \sum_j A_{jk}\delta_{ij} = A_{ik} \end{aligned}$$

Thus $\alpha(\eta_i)(e_j) = \sum_k A_{ik}\epsilon_k(e_j) = \sum_k A^T_{ki}\epsilon_k(e_j)$ so $\alpha(\eta_i) = \sum_K A^T_{ki}\epsilon_k$ as required. □

Remarks.

(1) If $\alpha\colon U \to V$ and $\beta\colon V \to W$ are linear maps then $(\beta\alpha)^* = \alpha^*\beta^*$.
(2) If $\alpha, \beta\colon U \to V$ then $(\alpha + \beta)^* = \alpha^* + \beta^*$.
(3) If $\alpha \in \mathrm{End}(V)$ then $\det \alpha^* = \det \alpha$ since $\det(A^T) = \det A$.
(4) If $B = Q^{-1}AP$ is an equality of matrices with P and Q invertible, then

$$B^T = P^T A^T \left(Q^{-1}\right)^T = \left(\left(P^{-1}\right)^T\right)^{-1} A^T \left(Q^{-1}\right)^T$$

as we should expect at this point.

Lecture 16

Lemma. *Suppose that $\alpha \in \mathcal{L}(V, W)$ with V, W f.d. over $\mathbf{F}$. Then*

(a) $\ker \alpha^* = (\operatorname{Im} \alpha)^\circ$;
(b) $r(\alpha^*) = r(\alpha)$ *and*
(c) $\operatorname{Im} \alpha^* = (\ker \alpha)^\circ$

Proof. (a) Suppose $\theta \in W^*$. Then $\theta \in \ker \alpha^*$ if and only if $\alpha^*(\theta) = 0$ if and only if $\theta\alpha(v) = 0$ for all $v \in V$ if and only if $\theta \in (\operatorname{Im}\alpha)^\circ$.

(b) As $\operatorname{Im}\alpha$ is a subspace of W, we've seen that $\dim \operatorname{Im}\alpha + \dim(\operatorname{Im}\alpha)^\circ = \dim W$. Using part (a) we can deduce that $r(\alpha) + n(\alpha^*) = \dim W = \dim W^*$. But the rank-nullity theorem gives $r(\alpha^*) + n(\alpha^*) = \dim W^*$.

(c) Suppose that $\phi \in \operatorname{Im}\alpha^*$. Then there is some $\theta \in W^*$ such that $\phi = \alpha^*(\theta) = \theta\alpha$. Therefore for all $v \in \ker\alpha$, $\phi(v) = \theta\alpha(v) = \theta(0) = 0$. Thus $\operatorname{Im}\alpha^* \subset (\ker\alpha)^\circ$.

But $\dim\ker\alpha + \dim(\ker\alpha)^\circ = \dim V$. So

$$\dim(\ker\alpha)^\circ = \dim V - n(\alpha) = r(\alpha) = r(\alpha^*) = \dim \operatorname{Im}\alpha^*.$$

and so the inclusion must be an equality. □

Notice that we have reproven that row-rank=column rank in a more conceptually satisfying way.

Lemma. *Let V be a vector space over $\mathbf{F}$ there is a canonical linear map* $\mathrm{ev}\colon V \to V^{**}$ *given by* $\mathrm{ev}(v)(\theta) = \theta(v)$.

Proof. First we must show that $\mathrm{ev}(v) \in V^{**}$ whenever $v \in V$. Suppose that $\theta_1, \theta_2 \in V^*$ and $\lambda, \mu \in \mathbf{F}$. Then

$$\mathrm{ev}(v)(\lambda\theta_1 + \mu\theta_2) = \lambda\theta_1(v) + \mu\theta_2(v) = \lambda\mathrm{ev}(v)(\theta_1) + \mu\mathrm{ev}(v)(\theta_2).$$

Next, we must show ev is linear, ie $\mathrm{ev}(\lambda v_1 + \mu v_2) = \lambda\mathrm{ev}(v_1) + \mathrm{ev}(v_2)$ whenever $v_1, v_2 \in V$, $\lambda, \mu \in \mathbf{F}$. We can show this by evaluating both sides at each $\theta \in V^*$. Then

$$\mathrm{ev}(\lambda v_1 + \mu v_2)(\theta) = \theta(\lambda v_1 + \mu v_2) = (\lambda\mathrm{ev}(v_1) + \mu\mathrm{ev}(v_2))(\theta)$$

so ev is linear. □

Lemma. *Suppose that V is f.d. then the canonical linear map* $\mathrm{ev}\colon V \to V^{**}$ *is an isomorphism.*

Proof. Suppose that $\mathrm{ev}(v) = 0$. Then $\theta(v) = \mathrm{ev}(v)(\theta) = 0$ for all $\theta \in V^*$. Thus the annilhilator of the span of V has dimension $\dim V$. It follows that the span of v is a space of dimension 0 so $v = 0$. In particular we've proven that ev is injective.

To complete the proof it suffices to observe that $\dim V = \dim V^* = \dim V^{**}$ so any injective linear map $V \to V^{**}$ is an isomorphism. □

Remarks.

(1) The lemma tells us more than that there is an isomorphism between V and V^{**}. It tells us that there is a way to define such an isomorphism canonically, that is to say without choosing bases. This means that we can, and from now on we will identify V and V^{**} whenever V is f.d. In particular for $v \in V$ and $\theta \in V^*$ we can write $v(\theta) = \theta(v)$.

(2) Although the canonical linear map is $\mathrm{ev}\colon V \to V^{**}$ always exists it is not an isomorphism in general if V is not f.d.

Lemma. *Suppose V and W are f.d. over $\mathbf{F}$. After identifying V with V^{**} and W with W^{**} via* ev *we have*

(a) If U is a subspace of V then $U^{\circ\circ} = U$.
(b) If $\alpha \in \mathcal{L}(V, W)$ then $\alpha^{**} = \alpha$.

Proof. (a) Let $u \in U$. Then $u(\theta) = \theta(u) = 0$ for all $\theta \in U^\circ$. Thus $u \in U^{\circ\circ}$. ie $U \subset U^{\circ\circ}$. But

$$\dim U = \dim V - \dim U^\circ = \dim V^* - \dim U^\circ = \dim U^{\circ\circ}.$$

(b) Suppose that $\langle e_1, \ldots, e_n \rangle$ is a basis for V and $\langle f_1, \ldots, f_m \rangle$ is a basis for W and $\langle \epsilon_1, \ldots, \epsilon_n \rangle$ and $\langle \eta_1, \ldots, \eta_m \rangle$ are the corresponding dual bases. Then if α is represented by A with respect to $\langle e_1, \ldots, e_n \rangle$ and $\langle f_1, \ldots, f_m \rangle$, α^* is represented by A^T with respect to $\langle \epsilon_1, \ldots, \epsilon_n \rangle$ and $\langle \eta_1, \ldots, \eta_n \rangle$.

Since we can view $\langle e_1, \ldots, e_n \rangle$ as the dual basis to $\langle \epsilon_1, \ldots, \epsilon_n \rangle$ as

$$e_i(\epsilon_j) = \epsilon_j(e_i) = \delta_{ij},$$

and $\langle f_1, \ldots, f_m \rangle$ as the dual basis of $\langle \eta_1, \ldots, \eta_m \rangle$ (by a similar computation), α^{**} is represented by $(A^T)^T = A$. □

Proposition. *Suppose V is f.d. over $\mathbf{F}$ and U_1, U_2 are subspaces of V then*
(a) $(U_1 + U_2)^\circ = U_1^\circ \cap U_2^\circ$ *and*
(b) $(U_1 \cap U_2)^\circ = U_1^\circ + U_2^\circ$.

Proof. (a) Suppose that $\theta \in V^*$. Then $\theta \in (U_1 + U_2)^\circ$ if and only if $\theta(u_1 + u_2) = 0$ for all $u_1 \in U_1$ and $u_2 \in U_2$ if and only if $\theta(u) = 0$ for all $u \in U_1 \cup U_2$ if and only if $\theta \in U_1^\circ \cap U_2^\circ$.

(b) by part (a), $U_1 \cap U_2 = U_1^{\circ\circ} \cap U_2^{\circ\circ} = (U_1^\circ + U_2^\circ)^\circ$. Thus

$$(U_1 \cap U_2)^\circ = (U_1^\circ + U_2^\circ)^{\circ\circ} = U_1^\circ + U_2^\circ$$

as required □

6. Bilinear Forms

6.1. **Definitions and Examples.** Let V and W be vector spaces over $\mathbf{F}$.

Definition. $\psi \colon V \times W \to \mathbf{F}$ is a *bilinear form* if it is linear in both arguments; i.e. if $\psi(v, -) \colon W \to \mathbf{F} \in W^*$ for all $v \in V$ and $\psi(-, w) \colon V \to \mathbf{F} \in V^*$ for all $w \in W$.

Examples.
(1) $V = \mathbf{R}^n$; $\psi(x, y) = \sum_{i=1}^n x_i y_i$ is a bilinear form.
(2) Suppose that $A \in \mathrm{Mat}_{m,n}(\mathbf{F})$ then $\psi \colon \mathbf{F}^m \times \mathbf{F}^n \to \mathbf{F}$; $\psi(v, w) = v^T A w$ is a bilinear form.
(3) The map $V \times V^* \to \mathbf{F}$; $(v, \theta) \mapsto \theta(v)$ is a bilinear form.
(4) If $V = W = C([0,1], \mathbf{R})$ then $\psi(f, g) = \int_0^1 f(t)g(t)\,dt$ is a bilinear form.

Lecture 17

We recall the following definition from last time.

Definition. $\psi \colon V \times W \to \mathbf{F}$ is a *bilinear form* if it is linear in both arguments; i.e. if $\psi(v, -) \colon W \to \mathbf{F} \in W^*$ for all $v \in V$ and $\psi(-, w) \colon V \to \mathbf{F} \in V^*$ for all $w \in W$.

We can see that a bilinear form ψ gives linear maps $\psi_L \colon V \to W^*$ and $\psi_R \colon W \to V^*$ by the formulae

$$\psi_L(v)(w) = \psi(v, w) = \psi_R(w)(v)$$

for $v \in V$ and $w \in W$.

Definition. Let $\langle e_1, \dots, e_n \rangle$ be a basis for V and $\langle f_1, \dots, f_m \rangle$ be a basis of W and $\psi\colon V \times W \to \mathbf{F}$ a bilinear form. Then the *matrix A representing ψ with respect to $\langle e_1, \dots, e_n \rangle$ and $\langle f_1, \dots, f_m \rangle$* is given by $A_{ij} = \psi(e_i, f_j)$.

Remark. If $v = \sum \lambda_i e_i$ and $w = \sum \mu_j f_j$ then

$$\psi\left(\sum \lambda_i e_i, \mu_j f_j\right) = \sum_{i=1}^{n} \lambda_i \psi\left(e_i, \sum \mu_j f_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \lambda_i \mu_j \psi(e_i, f_j).$$

Therefore if A is the matrix representing ψ with respect to $\langle e_1, \dots, e_n \rangle$ and $\langle f_1, \dots, f_m \rangle$ we have

$$\psi(v, w) = \begin{pmatrix} \lambda_1 & \cdots & \lambda_n \end{pmatrix} A \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}$$

and ψ is determined by the matrix representing it.

Lemma. *Let $\langle \epsilon_1, \dots, \epsilon_n \rangle$ be the dual basis to $\langle e_1, \dots, e_n \rangle$ and $\langle \eta_1, \dots, \eta_m \rangle$ be the dual basis to $\langle f_1, \dots, f_m \rangle$. Then A represents ψ_R with respect to $\langle f_1, \dots, f_m \rangle$ and $\langle \epsilon_1, \dots, \epsilon_m \rangle$ and A^T represents ψ_L with respect to $\langle e_1, \dots, e_n \rangle$ and $\langle \eta_1, \dots, \eta_m \rangle$.*

Proof. We can compute $\psi_L(e_i)(f_j) = \psi(e_i, f_j) = A_{ij}$ and so $\psi_L(e_i) = \sum_{j=1}^m A^T_{ji} \eta_j$ and $\psi_R(f_j)(e_i) = \psi(e_i, f_j) = A_{ij}$ and so $\psi_L(f_j) = \sum_{i=1}^n A_{ij} \epsilon_i$. $\square$

Proposition. *Suppose that $\langle e_1, \dots, e_n \rangle$ and $\langle v_1, \dots, v_n \rangle$ are two bases of V such that $v_i = \sum_{j=1}^n P_{ji} e_j$ for $i = 1, \dots, n$ and $\langle f_1, \dots, f_m \rangle$ and $\langle w_1, \dots, w_m \rangle$ are two bases of W such that $w_i = \sum_{j=1}^m Q_{ji} f_j$ for $i = 1, \dots, m$. Let $\psi\colon V \times W \to \mathbf{F}$ be a bilinear form represented by A with respect to $\langle e_1, \dots, e_n \rangle$ and $\langle f_1, \dots, f_m \rangle$ and by B with respect to $\langle v_1, \dots, v_n \rangle$ and $\langle w_1, \dots, w_m$ then*

$$B = P^T A Q.$$

Proof.

$$\begin{aligned} B_{ij} &= \psi(v_i, w_j) \\ &= \psi\left(\sum_{k=1}^n P_{ki} e_k, \sum_{l=1}^m Q_{lj} f_l\right) \\ &= \sum_{k,l} P_{ki} Q_{lj} \psi(e_k, f_l) \\ &= (P^T A Q)_{ij} \end{aligned}$$

$\square$

Corollary. *Let V and W be f.d. vector spaces over $\mathbf{F}$ and $\psi\colon V \times W \to \mathbf{F}$ a bilinear form. There are bases $\langle v_1, \dots, v_m \rangle$ of V and $\langle w_1, \dots, w_n \rangle$ of W and an $r \leqslant \min(m, n)$ such that $\psi\left(\sum_{i=1}^m \lambda_i v_i, \sum_{j=1}^n \mu_j w_j\right) = \sum_{i=1}^r \lambda_i \mu_i$.*

Proof. Let $\langle e_1, \dots, e_m \rangle$ and $\langle f_1, \dots, f_n \rangle$ be any bases of V and W respectively and let A be the matrix representing ψ with respect to this pair of bases. We know that there are $R \in GL_m(\mathbf{F})$ and $Q \in GL_n(\mathbf{F})$ such that

$$R^{-1} A Q = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

for $r = r(A)$. Taking $P = (R^{-1})^T$, $v_i = \sum_{k=1}^{m} P_{ki}e_k$ for $1 \leqslant i \leqslant m$ and $w_j = \sum_{l=1}^{n} Q_{lj}e_l$ for $1 \leqslant j \leqslant n$ gives the result. □

We note that r only depends on ψ since $r(P^TAQ) = r(A)$ whenever P, Q are invertible. We call r the *rank* of ψ written $r(\psi)$.

Definition. We say a bilinear form $\psi\colon V \times W \to \mathbf{F}$ is *degenerate* if there is either some $v \in V\backslash 0$ such that $\psi(v, -) = 0 \in W^*$ or there is some $w \in W\backslash 0$ such that $\psi(-, w) = 0 \in V^*$. Otherwise we say that ψ is *non-degenerate*.

Lemma. *Let V and W be f.d. vector spaces over $\mathbf{F}$ with bases $\langle e_1, \ldots, e_n\rangle$ and $\langle f_1, \ldots, f_m\rangle$ and let $\psi\colon W \times V \to \mathbf{F}$ be a bilinear form represented by the matrix A with respect to those bases. Then ψ is non-degenerate if and only if the matrix A is invertible. In particular, if ψ non-degenerate then $\dim V = \dim W$.*

Proof. The condition that ψ is non-degenerate is equivalent to $\ker \psi_L = 0$ and $\ker \psi_R = 0$ which is in turn equivalent to $n(A) = 0 = n(A^T)$. This last is equivalent to $r(A) = \dim V$ and $r(A^T) = \dim W$. Since row-rank and column-rank agree we can see that this final statement is equivalent to A being invertible as required. □

It follows that, when V and W are f.d., defining a non-degenerate bilinear form $\psi\colon V \times W \to \mathbf{F}$ is equivalent to defining an isomorphism $\psi_L\colon V \to W^*$ (or equivalently an isomorphism $\psi_R\colon W \to V^*$).

6.2. Symmetric bilinear forms and quadratic forms.

Definition. Let V be a vector space over $\mathbf{F}$. A bilinear form $\phi\colon V \times V \to \mathbf{F}$ is *symmetric* if $\phi(v_1, v_2) = \phi(v_2, v_1)$ for all $v \in V$.

Example. Suppose $S \in \mathrm{Mat}_n(\mathbf{F})$ is a symmetric matrix (ie $S^T = S$), then we can define a symmetric bilinear form $\phi\colon \mathbf{F}^n \times \mathbf{F}^n \to \mathbf{F}$ by

$$\phi(x, y) = x^TSy = \sum_{i,j=1}^{n} x_iS_{ij}y_j$$

In fact that example is completely typical.

Lemma. *Suppose that V is a f.d. vector space over $\mathbf{F}$ and $\phi\colon V \times V \to \mathbf{F}$ is a bilinear form. Let $\langle e_1, \ldots, e_n\rangle$ be a basis for V and M be the matrix representing ϕ with respect to this basis, i.e. $M_{ij} = \phi(e_i, e_j)$. Then ϕ is symmetric if and only if M is symmetric.*

Proof. If ϕ is symmetric then $M_{ij} = \phi(e_i, e_j) = \phi(e_j, e_i) = M_{ji}$ so M is symmetric. Conversely if M is symmetric, then

$$\phi(x, y) = \sum_{i,j=1}^{n} x_iM_{ij}y_j = \sum_{i,j=1}^{n} y_jM_{ji}x_i = \phi(y, x).$$

Thus ϕ is symmetric. □

It follows that if ϕ is represented by a symmetric matrix with respect to one basis then it is represented by a symmetric matrix with respect to every basis.

Lemma. *Suppose that V is a f.d. vector space over $\mathbf{F}$, $\phi\colon V \times V \to \mathbf{F}$ is a bilinear form and $\langle e_1, \dots, e_n \rangle$ and $\langle f_1, \dots, f_n \rangle$ are two bases of V such that $f_i = \sum P_{ki} e_k$ for $i = 1, \dots n$. If A represents ϕ with respect to $\langle e_1, \dots, e_n \rangle$ and B represents ϕ with respect to $\langle f_1, \dots, f_n \rangle$ then*

$$B = P^T A P$$

Proof. We compute

$$B_{ij} = \phi(f_i, f_j) = \phi(\sum_k P_{ki} e_k, \sum_l P_{lj} e_l) = \sum_{k,l} P_{ki} P_{lj} \phi(e_k, e_l).$$

Thus $B_{ij} = \sum_{k,l} P^T_{ik} A_{kl} P_{lj} = [P^T A P]_{ij}$. $\square$

Definition. We say that square matrices A and B are *congruent* if there is an invertible matrix P such that $B = P^T A P$.

Congruence is an equivalence relation. Two matrices are congruent precisely if they represent the same bilinear form $\phi\colon V \times V \to \mathbf{F}$ with respect to different bases for V. Thus to classify (symmetric) bilinear forms on a f.d. vector space is to classify (symmetric) matrices up to congruence.

Lecture 18

Definition. If $\phi\colon V \times V \to \mathbf{F}$ is a bilinear form then we call the map $V \to \mathbf{F}$; $v \mapsto \phi(v, v)$ a *quadratic form* on V.

Example. If $V = \mathbf{R}^2$ and ϕ is represented by the matrix A with respect to the standard basis then the corresponding quadratic form is

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x & y \end{pmatrix} A \begin{pmatrix} x \\ y \end{pmatrix} = A_{11}x^2 + (A_{12} + A_{21})xy + A_{22}y^2$$

Note that if we replace A by the symmetric matrix $\frac{1}{2}\left(A + A^T\right)$ we get the same quadratic form.

Proposition (Polarisation identity)**.** *If $q\colon V \to \mathbf{F}$ is a quadratic form then there exists a unique symmetric bilinear form $\phi\colon V \times V \to \mathbf{F}$ such that $q(v) = \phi(v, v)$ for all $v \in V$.*

Proof. Let ψ be a bilinear form on $V \times V$ such that $\psi(v, v) = q(v)$ for all $v \in V$. Then

$$\phi(v, w) := \frac{1}{2}\left(\psi(v, w) + \psi(w, v)\right)$$

is a symmetric bilinear form such that $\phi(v, v) = q(v)$ for all $v \in V$.

It remains to prove uniqueness. Suppose that ϕ is such a symmetric bilinear form. Then for $v, w \in V$,

$$\begin{aligned} q(x + y) &= \phi(v + w, v + w) \\ &= \phi(v, v) + \phi(v, w) + \phi(w, v) + \phi(w, w) \\ &= q(v) + 2\phi(v, w) + q(w). \end{aligned}$$

Thus $\phi(v, w) = \frac{1}{2}\left(q(v + w) - q(v) - q(w)\right)$. $\square$

Theorem (Canonical form for symmetric bilinear forms)**.** *If $\phi\colon V \times V \to \mathbf{F}$ is a symmetric bilinear form on a f.d. vector space V over $\mathbf{F}$, then there is a basis $\langle e_1, \dots, e_n \rangle$ for V such that ϕ is represented by a diagonal matrix.*

Proof. By induction on $n = \dim V$. If $n = 0, 1$ the result is clear. Suppose that we have proven the result for all spaces of dimension strictly smaller than n.

If $\phi(v, v) = 0$ for all $v \in V$, then by the polarisation identity ϕ is identically zero and is represented by the zero matrix with respect to every basis. Otherwise, we can choose $e_1 \in V$ such that $\phi(e_1, e_1) \neq 0$. Let

$$U = \{u \in V \mid \phi(e_1, u) = 0\} = \ker \phi(e_1, -)\colon V \to \mathbf{F}.$$

By the rank-nullity theorem, U has dimension $n-1$ and $e_1 \notin U$ so U is a complement to the span of e_1 in V.

Consider $\phi|_{U\times U}\colon U \times U \to \mathbf{F}$, a symmetric bilinear form on U. By the induction hypothesis, there is a basis $\langle e_2, \ldots, e_n\rangle$ for U such that $\phi|_{U\times U}$ is represented by a diagonal matrix. The basis $\langle e_1, \ldots, e_n\rangle$ satisfies $\phi(e_i, e_j) = 0$ for $i \neq j$ and we're done. $\square$

Example. Let q be the quadratic form on $\mathbf{R}^3$ given by

$$q\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = x^2 + y^2 + z^2 + 2xy + 4yz + 6xz.$$

Find a basis $\langle f_1, f_2, f_3\rangle$ for $\mathbf{R}^3$ such that q is of the form

$$q(af_1 + bf_2 + cf_3) = \lambda a^2 + \mu b^2 + \nu c^2.$$

Method 1 Let ϕ be the bilinear form represented by the matrix

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

so that $q(v) = \phi(v, v)$ for all $v \in \mathbf{R}^3$.

Now $q(e_1) = 1 \neq 0$ so let $f_1 = e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Then $\phi(f_1, v) = f_1^T Av = v_1 + v_2 + 3v_3$.

So we choose f_2 such that $\phi(f_1, f_2) = 0$ but $\phi(f_2, f_2) \neq 0$. For example

$$q\left(\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}\right) = 0 \text{ but } q\left(\begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}\right) = -8 \neq 0.$$

So we can take $f_2 = \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}$. Then $\phi(f_2, v) = f_2^T Av = \begin{pmatrix} 0 & 1 & 8 \end{pmatrix} v = v_2 + 8v_3$.

Now we want $\phi(f_1, f_3) = \phi(f_2, f_3) = 0$, $f_3 = (5, -8, 1)^T$ will work. Then

$$q(af_1 + bf_2 + cf_3) = a^2 + (-8)b^2 + 8c^2.$$

Method 2 Complete the square

$$\begin{aligned} x^2 + y^2 + z^2 + 2xy + 4yz + 6xz &= (x + y + 3z)^2 + (-2yz) - 8z^2 \\ &= (x + y + 3z)^2 - 8\left(z + \frac{y}{8}\right)^2 + \frac{y^2}{8} \end{aligned}$$

Now solve $x + y + 3z = 1$, $z + \frac{y}{8} = 0$ and $y = 0$ to obtain $f_1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T$, solve $x + y + 3z = 0$, $z + \frac{y}{8} = 1$ and $y = 0$ to obtain $f_2 = \begin{pmatrix} -3 & 0 & 1 \end{pmatrix}^T$ and solve $x + y + 3z = 0$, $z + \frac{y}{8} = 0$ and $y = 1$ to obtain $f_3 = \begin{pmatrix} -\frac{5}{8} & 1 & -\frac{1}{8} \end{pmatrix}^T$.

Lecture 19

Corollary. *Let ϕ be a symmetric bilinear form on a f.d $\mathbf{C}$-vector space V. Then there is a basis $\langle v_1, \ldots, v_n \rangle$ for V such that ϕ is represented by a matrix of the form*

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

with $r = r(\phi)$ or equivalently such that the corresponding quadratic form q is given by $q(\sum_{i=1}^n a_i v_i) = \sum_{i=1}^r a_i^2$.

Proof. We have already shown that there is a basis $\langle e_1, \ldots, e_n \rangle$ such that $\phi(e_i, e_j) = \delta_{ij}\lambda_j$ for some $\lambda_1, \ldots, \lambda_n \in \mathbf{C}$. By reordering the e_i we can assume that $\lambda_i \neq 0$ for $1 \leqslant i \leqslant r$ and $\lambda_i = 0$ for $i > r$. Since we're working over $\mathbf{C}$ for each $1 \leqslant i \leqslant r$, λ_i has a non-zero square root μ_i, say. Defining $v_i = \frac{1}{\mu_i} e_i$ for $1 \leqslant i \leqslant r$ and $v_i = e_i$ for $r + 1 \leqslant i \leqslant n$, we see that $\phi(v_i, v_j) = 0$ if $i \neq j$ or $i = j > r$ and $\phi(v_i, v_i) = 1$ if $1 \leqslant i \leqslant r$ as required. □

Corollary. *Every symmetric matrix in* $\mathrm{Mat}_n(\mathbf{C})$ *is congruent to a matrix of the form*

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

□

Corollary. *Let ϕ be a symmetric bilinear form on a f.d $\mathbf{R}$-vector space V. Then there is a basis $\langle v_1, \ldots, v_n \rangle$ for V such that ϕ is represented by a matrix of the form*

$$\begin{pmatrix} I_s & 0 & 0 \\ 0 & -I_{r-s} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with $r = r(\phi)$ and $0 \leqslant s \leqslant r$ or equivalently such that the corresponding quadratic form q is given by $q(\sum_{i=1}^n a_i v_i) = \sum_{i=1}^s a_i^2 - \sum_{i=s+1}^r a_i^2$.

Proof. We have already shown that there is a basis $\langle e_1, \ldots, e_n \rangle$ such that $\phi(e_i, e_j) = \delta_{ij}\lambda_j$ for some $\lambda_1, \ldots, \lambda_n \in \mathbf{R}$. By reordering the e_i we can assume that there is an s such that $\lambda_i > 0$ for $1 \leqslant i \leqslant s$ and $\lambda_i < 0$ for $s + 1 \leqslant i \leqslant r$ and $\lambda_i = 0$ for $i > r$. Since we're working over $\mathbf{R}$ we can define $\mu_i = \sqrt{\lambda_i}$ for $1 \leqslant i \leqslant s$, $\mu_i = \sqrt{-\lambda_i}$ for $s + 1 \leqslant i \leqslant r$ and $\mu_i = 1$ for $i = 1$. Defining $v_i = \frac{1}{\mu_i} e_i$ we see that ϕ is represented by the given matrix with respect to $v_1, \ldots, v_n$. □

Corollary. *Every real symmetric matrix is congruent to a matrix of the form*

$$\begin{pmatrix} I_s & 0 & 0 \\ 0 & -I_{r-s} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

□

Definition. A symmetric bilinear form ϕ on a real vector space V is

(a) *positive definite* if $\phi(v, v) > 0$ for all $v \in V \backslash 0$;
(b) *positive semi-definite* if $\phi(v, v) \geqslant 0$ for all $v \in V$;
(c) *negative definite* if $\phi(v, v) < 0$ for all $v \in V \backslash 0$;
(d) *negative semi-definite* if $\phi(v, v) \leqslant 0$ for all $v \in V$.

We say a quadratic form is *...-definite* if the corresponding bilinear form is so.

Example. $\phi(x,y) := \sum_{i=1}^n x_i y_i$ is positive definite on $\mathbf{R}^n$.

Theorem (Sylvester's Law of Inertia)**.** *Let V be an n-dimensional real vector space and let ϕ be a symmetric bilinear form on V. Then there are* unique *integers s, r such that V has a basis $v_1, \ldots, v_n$ with respect to which ϕ is represented by the matrix*

$$\begin{pmatrix} I_s & 0 & 0 \\ 0 & -I_{r-s} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Proof. We've already done the existence part. We also already know that $r = r(\phi)$ is unique. To see s is unique we'll prove that s is the largest dimension of a subspace P of V such that $\phi|_{P\times P}$ is positive definite.

Let $v_1, \ldots, v_n$ be some basis with respect to which ϕ is represented by

$$\begin{pmatrix} I_s & 0 & 0 \\ 0 & -I_{r-s} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

for some choice of s. Then ϕ is positive definite on the space spanned by $v_1, \ldots, v_s$. Thus it remains to prove that there is no larger such subspace.

Let P be any subspace of V such that $\phi|_{P\times P}$ is positive definite and let Q be the space spanned by $v_{s+1}, \ldots, v_n$. The restriction of ϕ to $Q \times Q$ is negative semi-definite so $P \cap Q = 0$. So $\dim P + \dim Q = \dim P + Q \leqslant n$. Thus $\dim P \leqslant s$ as required. □

Definition. The *signature* of the symmetric bilinear form ϕ given in the Theorem is defined to be $s - (r - s) = 2s - r$.

6.3. **Hermitian forms.** Let V be a vector space over $\mathbf{C}$ and let ϕ be a symmetric bilinear form on V. Then ϕ can never be positive definite since $\phi(iv, iv) = -\phi(v,v)$ for all $v \in V$. We'd like to fix this.

Definition. Let V and W be vector spaces over $\mathbf{C}$. Then a *sesquilinear form* is a function $\phi\colon V \times W \to \mathbf{C}$ such that

$$\begin{aligned} \phi(\lambda_1 v_1 + \lambda_2 v_2, w) &= \overline{\lambda_1}\phi(v_1, w) + \overline{\lambda_2}\phi(v_2, w) \text{ and} \\ \phi(v, \mu_1 w_1 + \mu_2 w_2) &= \mu_1\phi(v, w_1) + \mu_2\phi(v, w_2) \end{aligned}$$

for all $\lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbf{C}$, $v, v_1, v_2 \in V$ and $w, w_1, w_2 \in W$.

Definition. Let ϕ be a sesquilinear form on $V \times W$ and let V have basis $\langle v_1, \ldots, v_m\rangle$ and W have basis $\langle w_1, \ldots, w_m\rangle$. The *matrix A representing* ϕ with respect to these bases is defined by $A_{ij} = \phi(v_i, w_j)$.

Suppose that $\sum \lambda_i v_i \in V$ and $\sum \mu_j w_j \in W$ then

$$\phi(\sum \lambda_i v_i, \sum \mu_j w_j) = \sum_{i=1}^m \sum_{j=1}^n \overline{\lambda_i} A_{ij} \mu_j = \overline{\lambda}^T A\mu.$$

Definition. A sesquilinear form $\phi\colon V \times V \to \mathbf{C}$ is said to be *Hermitian* if $\phi(x,y) = \overline{\phi(y,x)}$ for all $x, y \in V$.

Lemma. *Let $\phi\colon V \times V \to \mathbf{C}$ be a sesquilinear form on a complex vector space V with basis $\langle v_1, \ldots, v_n\rangle$. Then ϕ is Hermitian if and only if the matrix A representing ϕ with respect to this basis satisfies $A = \overline{A}^T$ (we also say the matrix A is* Hermitian*).*

Proof. If ϕ is Hermitian then

$$A_{ij} = \phi(v_i, v_j) = \overline{\phi(v_j, v_i)} = \overline{A_{ji}}.$$

Conversely if $A = \overline{A}^T$ then

$$\phi\left(\sum \lambda_i v_i, \sum \mu_j v_j\right) = \overline{\lambda}^T A\mu = \mu^T A^T \overline{\lambda} = \overline{\overline{\mu}^T A \lambda} = \overline{\phi\left(\sum \mu_j v_j, \sum \lambda_i v_i\right)}$$

as required $\square$

Lecture 20

Notice that if ϕ is a Hermitian form on V then $\phi(x,x) \in \mathbf{R}$ for all $x \in V$ and $\phi(\lambda x, \lambda x) = |\lambda|^2 \phi(x,x)$ for all $\lambda \in \mathbf{C}$.

Proposition (Change of basis)**.** *Suppose that ϕ is a Hermitian form on a f.d. complex vector space V and that $\langle e_1, \ldots, e_n\rangle$ and $\langle v_1, \ldots, v_n\rangle$ are bases for V such that $v_i = \sum_{k=1}^n P_{ki} e_k$. Let A be the matrix representing ϕ with respect to $\langle e_1, \ldots, e_n\rangle$ and B be the matrix representing ϕ with respect to $\langle v_1, \ldots, v_n\rangle$ then*

$$B = \overline{P}^T AP.$$

Proof. We compute

$$B_{ij} = \phi\left(\sum_{k=1}^n P_{ki} e_k, \sum_{l=1}^n P_{lj} e_l\right) = \sum_{k,l} (\overline{P}^T)_{ik} \phi(e_k, e_l) P_{lj} = [\overline{P}^T AP]_{ij}$$

as required. $\square$

Lemma (Polarisation Identity)**.** *A Hermitian form ϕ on a complex vector space V is determined by the function $\psi\colon V \to \mathbf{R}$; $v \mapsto \phi(v,v)$.*

Proof. It can be checked that

$$\phi(x,y) = \frac{1}{4}\left(\psi(x+y) - i\psi(x+iy) - \psi(x-y) + i\psi(x-iy)\right)$$

$\square$

Theorem (Hermitian version of Sylvester's Law of Inertia)**.** *Let V be a f.d. complex vector space and suppose that $\phi\colon V \times V \to \mathbf{C}$ is a Hermitian form on V. Then there is a basis $\langle v_1, \ldots, v_n\rangle$ of V with respect to which ϕ is represented by a matrix of the form*

$$\begin{pmatrix} I_s & 0 & 0 \\ 0 & -I_{r-s} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Moreover r and s depend only on ϕ not on the basis.

Notice that for such a basis $\phi(\sum \lambda_i v_i, \sum \lambda_i v_i) = \sum_{i=1}^s |\lambda_i|^2 - \sum_{j=s+1}^r |\lambda_j|^2$.

Sketch of Proof. This is nearly identical to the real case. For existence: if ϕ is identically zero then any basis will do. If not, then by the Polarisation Identity there is some $v_1 \in V$ such that $\phi(v_1, v_1) \neq 0$. By replacing v_1 by $\frac{v_1}{|\phi(v_1,v_1)|^{1/2}}$ we can assume that $\phi(v_1, v_1) = \pm 1$. Define $U := \ker \phi(v_1, -)\colon V \to \mathbf{C}$ a subspace of V of dimension $\dim V - 1$. Since $v_1 \notin U$, U is a complement to the span of v_1 in V. By induction on $\dim V$, there is a basis $\langle v_2, \ldots, v_n\rangle$ of U such that $\phi|_{U\times U}$ is represented by a matrix of the required form. Now $\langle v_1, v_2, \ldots, v_n\rangle$ is a basis for V that after suitable reordering works.

For uniqueness: r is the rank of the matrix representing ϕ with respect to any basis and s arises as the dimension of a maximal positive definite subspace as in the real symmetric case. □

7. Inner product spaces

7.1. **Definitions and basic properties.**

Definition. Let V be a vector space over $\mathbf{F}$. An *inner product* on V is a positive definite symmetric/Hermitian form ϕ on V. Usually instead of writing $\phi(x, y)$ we'll write (x, y). A vector space equipped with an inner product $(-,-)$ is called an *inner product space.*

Examples.

(1) The usual scalar product on $\mathbf{R}^n$ or $\mathbf{C}^n$: $(x, y) = \sum_{i=1}^n \overline{x_i} y_i$.

(2) Let $C([0, 1], \mathbf{F})$ be the space of continuous real/complex valued functions on $[0, 1]$ and define
$$(f, g) = \int_0^1 \overline{f(t)} g(t)\, dt.$$

(3) A weighted version of (2). Let $w\colon [0, 1] \to \mathbf{R}$ take only positive values and define
$$(f, g) = \int_0^1 w(t)\overline{f(t)} g(t)\, dt.$$

If V is an inner product space then we can define a norm $||\cdot||$ on V by $||v|| = (v, v)^{\frac{1}{2}}$. Note $||v|| \geqslant 0$ with equality if and only if $v = 0$. Note that the norm determines the inner product because of the polarisation identity.

Lemma (Cauchy–Schwarz inequality)**.** *Let V be an inner product space and take $v, w \in V$. Then $|(v, w)| \leqslant ||v||||w||$.*

Proof. Since $(-,-)$ is positive-definite,
$$0 \leqslant (v - \lambda w, v - \lambda w) = (v, v) - \lambda(v, w) - \overline{\lambda}(w, v) + |\lambda|^2(w, w)$$
for all $\lambda \in \mathbf{F}$. Now when $\lambda = \frac{(w,v)}{(w,w)}$ (the case $w = 0$ is clear) then we get
$$0 \leqslant (v, v) - \frac{2|(v, w)|^2}{(w, w)} + \frac{|(v, w)|^2}{(w, w)^2}(w, w) = (v, v) - \frac{|(v, w)|^2}{(w, w)}.$$
The inequality follows by multiplying by (w, w) rearranging and taking square roots. □

Corollary (Minkowski's inequality)**.** *Let V be an inner product space and take $v, w \in V$. Then $||v + w|| \leqslant ||v|| + ||w||$.*

Proof.
$$\begin{aligned} ||v + w||^2 &= (v + w, v + w) \\ &= ||v||^2 + (v, w) + (w, v) + ||w||^2 \\ &\leqslant ||v||^2 + 2||v||||w|| + ||w||^2 \\ &= (||v|| + ||w||)^2 \end{aligned}$$
Taking square roots gives the result. □

Lecture 21

Definition. Let V be an inner product space then $v, w \in V$ are said to be *orthogonal* if $(v, w) = 0$. A set $\{v_i \mid i \in I\}$ is *orthonormal* if $(v_i, v_j) = \delta_{ij}$ for $i, j \in I$. An *orthormal basis (o.n. basis)* for V is a basis for V that is orthonormal.

Suppose that V is a f.d. inner product space with o.n. basis $v_1, \ldots, v_n$. Then given $v \in V$, we can write $v = \sum_{i=1}^n \lambda_i v_i$. But then $(v_j, v) = \sum_{i=1}^n \lambda_i (v_j, v_i) = \lambda_j$. Thus $v = \sum_{i=1}^n (v_i, v) v_i$.

Lemma (Parseval's identity)**.** *Suppose that V is a f.d. inner product space with o.n basis $\langle v_1, \ldots, v_n \rangle$ then $(v, w) = \sum_{i=1}^n \overline{(v_i, v)}(v_i, w)$. In particular*

$$||v||^2 = \sum_{i=1}^n |(v_i, v)|^2.$$

Proof. $(v, w) = (\sum_{i=1}^n (v_i, v) v_i, \sum_{j=1}^n (v_j, w) v_j) = \sum_{i=1}^n \overline{(v_i, v)}(v_i, w)$. □

7.2. **Gram–Schmidt orthogonalisation.**

Theorem (Gram-Schmidt process)**.** *Let V be an inner product space and $e_1, e_2, \ldots$ be LI vectors. Then there is a sequence $v_1, v_2, \ldots$ of orthonormal vectors such that the sets $\{e_1, \ldots, e_k\}$ and $\{v_1, \ldots, v_k\}$ have the same span for each k.*

Proof. We proceed by induction on k. The case $k = 0$ is clear. Suppose we've found $v_1, \ldots, v_k$. Let

$$u_{k+1} = e_{k+1} - \sum_{i=1}^k (v_i, e_{k+1}) v_i.$$

Then for $j \leqslant k$,

$$(v_j, u_{k+1}) = (v_j, e_{k+1}) - \sum_{i=1}^k (v_i, e_{k+1})(v_j, v_i) = 0.$$

Since $\{v_1, \ldots, v_k\}$ and $\{e_1, \ldots, e_k\}$ span the same set, and $e_1, \ldots, e_{k+1}$ are LI, $\{v_1, \ldots, v_k, e_{k+1}\}$ are LI and so $u_{k+1} \neq 0$. Let $v_{k+1} = \frac{u_{k+1}}{||u_{k+1}||}$. □

Corollary. *Let V be a f.d. inner product space. Then any orthonormal sequence $v_1, \ldots, v_k$ can be extended to an orthonormal basis.*

Proof. Let $v_1, \ldots, v_k, x_{k+1}, \ldots, x_n$ be any basis of V extending $v_1, \ldots, v_k$. If we apply the Gram–Schmidt process to this basis we obtain an o.n. basis $w_1, \ldots, w_n$. Moreover one can check that $w_i = v_i$ for $1 \leqslant i \leqslant k$. □

Definition. Let V be an inner product space and let V_1, V_2 be subspaces of V. Then V is the *orthogonal (internal) direct sum* of V_1 and V_2, written $V = V_1 \perp V_2$, if

(1) $V = V_1 + V_2$;
(2) $V_1 \cap V_2 = 0$;
(3) $(v_1, v_2) = 0$ for all $v_1 \in V_1$ and $v_2 \in V_2$.

Note that condition (3) implies condition (2).

Definition. If $W \subset V$ is a subspace of an inner product space V then the *orthogonal complement* of W in V, written $W^\perp$, is the subspace of V

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

Corollary. *Let V be a f.d. inner product space and W a subspace of V. Then $V = W \perp W^\perp$.*

Proof. Of course if $w \in W$ and $w^\perp \in W^\perp$ then $(w, w^\perp) = 0$. So it remains to show that $V = W + W^\perp$. Let $w_1, \ldots, w_k$ be an o.n. basis of W and extend it to $w_1, \ldots, w_n$ an o.n. basis for V.

If $j > k$, then $(\sum_{i=1}^k \lambda_i w_i, w_j) = \sum \overline{\lambda_i}(w_i, w_j) = 0$ and so $w_j \in W^\perp$. Since $w_1, \ldots, w_n$ span V it follows that $V = W + W^\perp$. □

Notice that unlike general vector space complements, orthogonal complements are unique.

Definition. We can also define the *orthogonal (external) direct sum* of two inner product spaces V_1 and V_2 by endowing the vector space direct sum $V_1 \oplus V_2$ with the inner product

$$((v_1, v_2), (w_1, w_2)) = (v_1, w_1) + (v_2, w_2)$$

for $v_1, w_1 \in V_1$ and $v_2, w_2 \in V_2$.

Definition. Suppose that $V = U \oplus W$. Then we can define $\Pi \colon V \to W$ by $\Pi(u + w) = w$ for $u \in U$ and $w \in W$. We call Π a *projection map* onto W. If $U = W^\perp$ we call Π the *orthogonal projection* onto W.

Proposition. *Let V be a f.d. inner product space and $W \subset V$ be a subspace with o.n. basis $\langle e_1, \ldots, e_k \rangle$. Let Π be the orthogonal projection onto W. Then*

(a) $\Pi(v) = \sum_{i=1}^k (e_i, v)e_i$ *for each* $v \in V$*;*
(b) $||v - \Pi(v)|| \leqslant ||v - w||$ *for all* $w \in W$ *with equality if and only if* $\Pi(v) = w$*; that is* $\Pi(v)$ *is the closest point to* v *in* W*.*

Proof. (a) Put $w = \sum_{i=1}^k (e_i, v)e_i \in W$. Then

$$(e_j, v - w) = (e_j, v) - \sum_{i=1}^k (e_i, v)(e_j, e_i) = 0 \text{ for } 1 \leqslant j \leqslant k.$$

Thus $v - w \in W^\perp$. Now $v = w + (v - w)$ so $\Pi(v) = w$.

(b) If $x, y \in V$ are orthogonal then

$$||x + y||^2 = (x + y, x + y) = ||x||^2 + (x, y) + (y, x) + ||y||^2 = ||x||^2 + ||y||^2$$

so

$$||v - w||^2 = ||(v - \Pi(v)) + (\Pi(v) - w)||^2 = ||(v - \Pi(v)||^2 + ||(\Pi(v) - w)||^2$$

and $||v - w||^2 \geqslant ||v - \Pi(v)||^2$ with equality if and only if $||\Pi(v) - w||^2 = 0$ ie $\Pi(v) = w$. □

Lecture 22

7.3. Adjoints.

Lemma. *Suppose V and W are f.d. inner product spaces and $\alpha\colon V \to W$ is linear. Then there is a unique linear map $\alpha^*\colon W \to V$ such that $(\alpha(v), w) = (v, \alpha^*(w))$ for all $v \in V$ and $w \in W$.*

Proof. Let $\langle v_1, \ldots, v_m\rangle$ be an o.n. basis for V and $\langle w_1, \ldots, w_m\rangle$ be an o.n. basis for W and suppose that α is represented by the matrix A with respect to these bases. Then if $\alpha^*\colon W \to V$ satisfies $(\alpha(v), w) = (v, \alpha^*(w))$ for all $v \in V$ and $w \in W$. Then we can compute

$$(v_i, \alpha^*(w_j)) = (\alpha(v_i), w_j) = (\sum_k A_{ki}w_k, w_j) = \overline{A_{ji}}.$$

Thus $\alpha^*(w_j) = \sum \overline{A^T}_{kj}v_k$ ie α^* is represented by the matrix $\overline{A^T}$. In particular α^* is unique if it exists.

But to prove existence we can now take α^* to be the linear map represented by the matrix $\overline{A^T}$. Then

$$\begin{aligned}\left(\alpha\left(\sum_i \lambda_i v_i\right), \sum_j \mu_j w_j\right) &= \sum_{i,j} \overline{\lambda_i}\mu_j \left(\sum_k A_{ki}w_k, w_j\right) \\ &= \sum_{i,j} \overline{\lambda_i A_{ji}}\mu_j\end{aligned}$$

whereas

$$\begin{aligned}\left(\sum_i \lambda_i v_i, \sum_j \alpha^*\left(\mu_j w_j\right)\right) &= \sum_{i,j} \overline{\lambda_i}\mu_j \left(w_k, \sum_l \overline{A^T}_{lj} w_l\right) \\ &= \sum_{i,j} \overline{\lambda_i A_{ki}}\mu_j\end{aligned}$$

Thus $(\alpha(v), w) = (v, \alpha^*(w))$ for all $v \in V$ and $w \in W$ as required. □

Definition. We call the linear map α^* characterised by the lemma the *adjoint* of α.

We've seen that if α is represented by A with respect to some o.n. bases then α^* is represented by $\overline{A^T}$ with respect to the same bases.

Definition. Suppose that V is an inner product space. Then $\alpha \in \text{End}(V)$ is *self-adjoint* if $\alpha^* = \alpha$; i.e. if $(\alpha(v), w) = (v, \alpha(w))$ for all $v, w \in V$.

Thus if $V = \mathbf{R}^n$ with the standard inner product then a matrix is self-adjoint if and only if it is symmetric. If $V = \mathbf{C}^n$ with the standard inner product then a matrix is self-adjoint if and only if it is Hermitian.

Definition. If V is a real inner product space then we say that $\alpha \in \text{End}(V)$ is *orthogonal* if

$$(\alpha(v_1), \alpha(v_2)) = (v_1, v_2) \text{ for all } v_1, v_2 \in V.$$

By the polarisation identity α is orthogonal if and only if $||\alpha(v)|| = ||v||$ for all $v \in V$.

Note that a real square matrix is orthogonal (as an endomorphism of $\mathbf{R}^n$ with the standard inner product) if and only if its columns are orthonormal.

Lemma. *Suppose that V is a f.d. real inner product space. Let $\alpha \in \mathrm{End}(V)$. Then α is orthogonal if and only if α is invertible and $\alpha^* = \alpha^{-1}$.*

Proof. If $\alpha^* = \alpha^{-1}$ then $(v,v) = (v, \alpha^*\alpha(v)) = (\alpha(v), \alpha(v))$ for all $v \in V$ ie α is orthogonal.

Conversely, if α is orthogonal, let $v_1, \ldots, v_n$ be an o.n. basis for V. Then for each $1 \leqslant i,j \leqslant n$,

$$(v_i, v_j) = (\alpha(v_i), \alpha(v_j)) = (v_i, \alpha^*\alpha(v_j)).$$

Thus $\delta_{ij} = (v_i, v_j) = (v_i, \alpha^*\alpha(v_j))$ and $\alpha^*\alpha(v_j) = v_j$ as required. □

Corollary. *With notation as in the lemma, $\alpha \in \mathrm{End}(V)$ is orthogonal if and only if α is represnted by an orthogonal matrix with respect to any orthonormal basis.*

Proof. Let $\langle v_1, \ldots, v_n \rangle$ be an o.n. basis then α is represented by A with respect to this basis if and only if α^* is represented by A^T. Thus α is orthogonal if and only if A is invertible with inverse A^T i.e. A is orthogonal. □

Definition. If V is a f.d. real inner product space then

$$O(V) := \{\alpha \in \mathrm{End}(V) \mid \alpha \text{ is orthogonal}\}$$

forms a group under composition called the *orthogonal group* of V.

Definition. If V is a complex inner product space then we say that $\alpha \in \mathrm{End}(V)$ is *unitary* if

$$(\alpha(v_1), \alpha(v_2)) = (v_1, v_2) \text{ for all } v_1, v_2 \in V.$$

By the polarisation identity α is unitary if and only if $||\alpha(v)|| = ||v||$ for all $v \in V$.

Lemma. *Suppose that V is a f.d. complex inner product space. Let $\alpha \in \mathrm{End}(V)$. Then α is unitary if and only if α is invertible and $\alpha^* = \alpha^{-1}$.*

Proof. If $\alpha^* = \alpha^{-1}$ then $(v,v) = (v, \alpha^*\alpha(v)) = (\alpha(v), \alpha(v))$ for all $v \in V$ ie α is unitary.

Conversely, if α is unitary, let $v_1, \ldots, v_n$ be an o.n. basis for V. Then for each $1 \leqslant i,j \leqslant n$,

$$(v_i, v_j) = (\alpha(v_i), \alpha(v_j)) = (v_i, \alpha^*\alpha(v_j)).$$

Thus $\delta_{ij} = (v_i, v_j) = (v_i, \alpha^*\alpha(v_j))$ and $\alpha^*\alpha(v_j) = v_j$ as required. □

Corollary. *With notation as in the lemma, $\alpha \in \mathrm{End}(V)$ is unitary if and only if α is represnted by an unitary matrix A with respect to any orthonormal basis (ie $A^{-1} = \overline{A^T}$).*

Proof. Let $\langle v_1, \ldots, v_n \rangle$ be an o.n. basis then α is represented by A with respect to this basis if and only if α^* is represented by $\overline{A^T}$. Thus α is orthogonal if and only if A is invertible with inverse $\overline{A^T}$ i.e. A is unitary. □

Definition. If V is a f.d. complex inner product space then

$$U(V) := \{\alpha \in \mathrm{End}(V) \mid \alpha \text{ is unitary}\}$$

forms a group under composition called the *unitary group* of V.

Lecture 23

7.4. Spectral theory.

Lemma. *Suppose that V is an inner product space and $\alpha \in \mathrm{End}(V)$ is self-adjoint then*
(a) α has a real eigenvalue;
(b) all eigenvalues of α are real;
(c) eigenvectors of α with distinct eigenvalues are orthogonal.

Proof. (a) and (b) Suppose first that V is a complex inner product space. By the fundamental theorem of algebra α has an eigenvalue (since the mininal polynomial has a root). Suppose that $\alpha(v) = \lambda v$ with $v = V\backslash 0$ and $\lambda \in \mathbf{C}$. Then

$$\lambda(v,v) = (v, \lambda v) = (v, \alpha(v)) = (\alpha(v), v) = (\lambda v, v) = \overline{\lambda}(v,v).$$

Since $(v,v) \neq 0$ we can deduce $\lambda \in \mathbf{R}$.

Now, suppose that V is a real inner product space. Let $\langle v_1, \ldots, v_n \rangle$ be an o.n. basis. Then α is represented by a real symmetric matrix A. But A viewed as a complex matrix is also Hermitian so all its eigenvalues are real by the above. Finally, the eigenvalues of α are precisely the eigenvalues of A.

(c) Suppose $\alpha(v) = \lambda v$ and $\alpha(w) = \mu w$ with $\lambda \neq \mu \in \mathbf{R}$. Then

$$\lambda(v,w) = (\lambda v, w) = (\alpha(v), w) = (v, \alpha(w)) = (v, \mu(w)) = \mu(v,w).$$

Since $\lambda \neq \mu$ we must have $(v,w) = 0$. □

Theorem. *Let V be an inner product space and $\alpha \in End(V)$ self-adjont. Then V has an orthonormal basis of eigenvectors of α.*

Proof. By the lemma, α has a real eigenvalue λ, say. Thus we can find $v_1 \in V\backslash 0$ such that $\alpha(v_1) = \lambda v_1$. Let $U := \ker(v_1, -)\colon V \to \mathbf{F}$ the orthogonal complement of the span of v_1 in V.

If $u \in U$, then

$$(\alpha(u), v_1) = (u, \alpha(v_1)) = (u, \lambda v_1) = \lambda(u, v_1) = 0.$$

Thus $\alpha(u) \in U$ and α restricts to an element of $\mathrm{End}(U)$. Since $(\alpha(v), w) = (v, \alpha(w))$ for all $v, w \in V$ also for all $v, w \in U$ ie $\alpha|_U$ is also self-adjoint. By induction on $\dim V$ we can conclude that U has an o.n. basis of eigenvectors $\langle v_2, \ldots, v_n \rangle$ of $\alpha|_U$. Then $\langle \frac{v_1}{||v_1||}, v_2, \ldots, v_n \rangle$ is an o.n. basis for V consisting of eigenvectors of α. □

Corollary. *If V is an inner product space and $\alpha \in \mathrm{End}(V)$ is self adjoint then V is the orthogonal direct sum of its eigenspaces.*

Corollary. *Let $A \in \mathrm{Mat}_n(\mathbf{R})$ be a symmetric matrix. Then there is an orthogonal matrix P such that P^TAP is diagonal.*

Proof. Let $(-,-)$ be the standard inner product on $\mathbf{R}^n$. Then $A \in \mathrm{End}(\mathbf{R}^n)$ is self-adjoint so $\mathbf{R}^n$ has an o.n. basis $\langle e_1, \ldots, e_n \rangle$ consisting of eigenvectors of A. Let P be the matrix whose columns are given by $e_1, \ldots, e_n$. Then P is orthogonal and $P^TAP = P^{-1}AP$ is diagonal. □

Corollary. *Let V be a f.d. real inner product space and $\psi\colon V \times V \to \mathbf{R}$ a symmetric bilnear form. Then there is an orthonormal basis of V such that ψ is represented by a diagonal matrix.*

Proof. Let $\langle u_1, \ldots, u_n \rangle$ be any o.n. basis for V and suppose that A represents ψ with respect to this basis. Then A is symmetric and there is an orthogonal matrix P such that P^TAP is diagonal. Let $v_i = \sum_k P_{ki}u_k$. Then $\langle v_1, \ldots, v_n \rangle$ is an o.n. basis and ψ is represented by P^TAP with respect to it. $\square$

Remark. Note that in the proof the diagonal entries of P^TAP are the eigenvalues of A. Thus it is easy to see that the signature of ψ is given by

$$\# \text{ of positive eigenvalues of } A - \# \text{ of negative eigenvalues of } A.$$

Corollary. *Let V be a f.d. real vector space and let ϕ and ψ be symmetric bilinear forms on V. If ϕ is positive-definite there is a basis $v_1, \ldots, v_n$ for V with respect to which both forms are represented by a diagonal matrix.*

Proof. Use ϕ to make V into a real inner product space and then use the last corollary. $\square$

Corollary. *Let $A, B \in \mathrm{Mat}_n(\mathbf{R})$ be symmetric matrices such that A is postive definite (ie $v^TAv > 0$ for all $v \in \mathbf{R}^n \backslash 0$). Then there is an invertible matrix Q such that Q^TAQ and Q^TBQ are both diagonal.*

We can prove similar corollaries for f.d. complex inner product spaces. In particular:

(1) If $A \in \mathrm{Mat}_n(\mathbf{C})$ is Hermitian there is a unitary matrix U such that $\overline{U^T}AU$ is diagonal.
(2) If ψ is a Hermitian form on a complex inner product space then there is an orthonormal basis diagonalising ψ.
(3) If V is a complex vector space and ϕ and ψ are two Hermitian forms with ϕ positive definite then ϕ and ψ can be simultaneously diagonalised.
(4) If $A, B \in \mathrm{Mat}_n(\mathbf{C})$ are both Hermitian and A is positive definite (i.e. $\overline{v^T}Av > 0$ for all $v \in \mathbf{C}^n \backslash 0$) then there is an invertible matrix Q such that $\overline{Q^T}AQ$ and $\overline{Q^T}BQ$ are both diagonal.

Lecture 24

We can prove a similar diagonalisability theorem for unitary matrices. There is no direct analogue for real orthogonal matrices — because orthogonal matrices need not have an eigenvalue e.g. rotations in $\mathbf{R}^2$ — but see the last question of Example Sheet 4 for something close.

Theorem. *Let V be a f.d. complex inner product space and $\alpha \in \mathrm{End}(V)$ be unitary. Then V has an o.n. basis consisting of eigenvectors of α.*

Proof. By the fundamental theorem of algebra, α has an eigenvector v say. Let $W = \ker(v, -) \colon V \to \mathbf{C}$ a $\dim V - 1$ dimensional subspace. Then if $w \in W$,

$$(v, \alpha(w)) = (\alpha^{-1}v, w) = (\frac{1}{\lambda}v, w) = \lambda^{-1}(v, w) = 0.$$

Thus α restricts to a unitary endomorphism of W. By induction W has an o.n. basis consisting of eigenvectors of α. By adding $v/||v||$ to this basis of W we obtain a suitable basis of V. $\square$

8. Alternating forms

I believe that this short section is non-examinable.

Definition. Suppose that V is a vector space over $\mathbf{F}$. An bilinear form $\phi\colon V \times V \to \mathbf{F}$ is *alternating* if $\phi(v, v) = 0$ for all $v \in V$.

Lemma. *A bilinear form $\phi\colon V \times V \to \mathbf{F}$ is alternating if and only if it is skew-symmetric i.e $\phi(v, w) = -\phi(w, v)$ for all $v, w \in V$.*

Proof. If $v, w \in V$ then

$$\phi(v + w, v + w) = \phi(v, v) + \phi(v, w) + \phi(w, v) + \phi(w, w).$$

Thus if ϕ is alternating then $0 = 0 + \phi(v, w) + \phi(w, v) + 0$. That is ϕ is skew-symmetric.

Conversely, if ϕ is skew-symmetric and $v \in V$, then $\phi(v, v) = -\phi(v, v)$ and so $2\phi(v, v) = 0$. Since $2 \neq 0 \in \mathbf{F}$, we see that ϕ is alternating. □

Theorem. *If V is a f.d. vector space over $\mathbf{F}$ and $\phi\colon V \times V \to \mathbf{F}$ is an alternating bilinear form then there is a basis for V such that ϕ is represented by a block diagonal matrix with all block diagonal entries either*

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \textit{ or } \begin{pmatrix} 0 \end{pmatrix}.$$

In particular the rank of ϕ is even, and so if V has a non-degenerate alternating form then $\dim V$ *must be even.*

Proof. By induction on $\dim V$. If ϕ is identically zero then any basis is suitable. Otherwise, we can find $v, w \in V$ such that $\phi(v, w) = c \neq 0$. Let $v_1 = v/c$ and $v_2 = w$ so that $\phi(v_1, v_2) = 1$ and $\phi(v_2, v_1) = -1$ as required.

Now let $\alpha\colon V \to \mathbf{F}^2$ be the linear map given by $\alpha(v) = \begin{pmatrix} \phi(v_1, v) \\ \phi(v_2, v) \end{pmatrix}$ and define $U = \ker \alpha$. Since $\alpha(v_1) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ and $\alpha(v_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\operatorname{Im} \alpha = \mathbf{F}^2$ and so $\dim U = \dim V - 2$ by the rank-nullity theorem. Moreover U intersects the span of v_1 and v_2 trivially so that U is a complement to the span of v_1 and v_2.

Now if $u \in U$ then $\phi(u, u) = 0$ since $u \in U$. Thus ϕ restricts to an alternating form $\phi|_{U \times U}\colon U \times U \to \mathbf{F}$. By the induction hypothesis, U has a basis $\langle v_3, \ldots, v_n \rangle$ such that $\phi|_{U \times U}$ is represented by a matrix of the required form. It is straightforward to verify that $\langle v_1, \ldots, v_n \rangle$ is a suitable basis for V. □

Definition. If V is an $\mathbf{F}$-vector space equipped with an alternating bilinear form $\phi\colon V \times V \to \mathbf{F}$ then the *symplectic group* of V,

$$\operatorname{Sp}(V) := \{\alpha \in GL(V) \mid \phi(\alpha(v), \alpha(w)) = \phi(v, w) \text{ for all } v, w \in V\}.$$