Part IB of the Mathematical Tripos
of the University of Cambridge

Michaelmas 2012

Linear Algebra

Lectured by:
Prof. I. Grojnowski

Notes by:
Alex Chan

Comments and corrections should be sent to awlc2@cam.ac.uk.

Printed Friday, 11 January 2013.

Course schedule

Definition of a vector space (over $\mathbb{R}$ or $\mathbb{C}$), subspaces, the space spanned by a subset. Linear independence, bases, dimension. Direct sums and complementary subspaces. [3]

Linear maps, isomorphisms. Relation between rank and nullity. The space of linear maps from U to V, representation by matrices. Change of basis. Row rank and column rank. [4]

Determinant and trace of a square matrix. Determinant of a product of two matrices and of the inverse matrix. Determinant of an endomorphism. The adjugate matrix. [3]

Eigenvalues and eigenvectors. Diagonal and triangular forms. Characteristic and minimal polynomials. Cayley-Hamilton Theorem over $\mathbb{C}$. Algebraic and geometric multiplicity of eigenvalues. Statement and illustration of Jordan normal form. [4]

Dual of a finite-dimensional vector space, dual bases and maps. Matrix representation, rank and determinant of dual map. [2]

Bilinear forms. Matrix representation, change of basis. Symmetric forms and their link with quadratic forms. Diagonalisation of quadratic forms. Law of inertia, classification by rank and signature. Complex Hermitian forms. [4]

Inner product spaces, orthonormal sets, orthogonal projection, $V = W \oplus W^{\perp}$. Gram-Schmidt orthogonalisation. Adjoints. Diagonalisation of Hermitian matrices. Orthogonality of eigenvectors and properties of eigenvalues. [4]

Contents

1 Vector spaces

1.1 Definitions

5 Oct

We start by fixing a *field*, $\mathbb{F}$. We say that $\mathbb{F}$ is a field if:

- $\mathbb{F}$ is an abelian group under an operation called *addition*, $(+)$, with additive identity 0;
- $\mathbb{F}\backslash\{0\}$ is an abelian group under an operation called *multiplication*, $(\cdot)$, with multiplicative identity 1;
- Multiplication is distributive over addition; that is, $a\,(b+c) = ab + ac$ for all $a, b, c \in \mathbb{F}$.

Fields we've encountered before include the reals $\mathbb{R}$, the complex numbers $\mathbb{C}$, the ring of integers modulo p, $\mathbb{Z}/p = \mathbb{F}_p$, the rationals $\mathbb{Q}$, as well as $\mathbb{Q}(\sqrt{3}\,) = \left\{a + b\sqrt{3} : a, b \in \mathbb{Q}\right\}$, ...

Everything we will discuss works over any field, but it's best to have $\mathbb{R}$ and $\mathbb{C}$ in mind, since that's what we're most familiar with.

Definition. A *vector space* over $\mathbb{F}$ is a tuple $(V, +, \cdot)$ consisting of a set V, operations $+ : V \times V \to V$ (*vector addition*) and $\cdot : \mathbb{F} \times V \to V$ (*scalar multiplication*) such that

(i) $(V, +)$ is an abelian group, that is:

- Associative: for all $v_1, v_2, v_3 \in V$, $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$;
- Commutative: for all $v_1, v_2 \in V$, $v_1 + v_2 = v_2 + v_1$;
- Identity: there is some (unique) $0 \in V$ such that, for all $v \in V$, $0 + v = v = v + 0$;
- Inverse: for all $v \in V$, there is some $u \in V$ with $u + v = v + u = 0$. This inverse is unique, and often denoted $-v$.

(ii) Scalar multipication satisfies

- Associative: for all $\lambda_1, \lambda_2 \in \mathbb{F}$, $v \in V$, $\lambda_1 \cdot (\lambda_2 \cdot v) = (\lambda_1\lambda_2) \cdot v$;
- Identity: for all $v \in V$, the unit $1 \in \mathbb{F}$ acts by $1 \cdot v = v$;
- $\cdot$ distributes over $+_V$: for all $\lambda \in \mathbb{F}$, $v_1, v_2 \in V$, $\lambda\cdot(v_1 + v_2) = \lambda\cdot v_1 + \lambda\cdot v_2$;
- $+_\mathbb{F}$ distributes over $\cdot$: for all $\lambda_1, \lambda_2 \in \mathbb{F}$, $v \in V$, $(\lambda_1 + \lambda_2)\cdot v = \lambda_1\cdot v + \lambda_2\cdot v$;

We usually say "the vector space V" rather than $(V, +, \cdot)$.

Let's look at some examples:

Examples 1.1.

(i) $\{0\}$ is a vector space.

(ii) Vectors in the plane under vector addition form a vector space.

(iii) The space of n-tuples with entries in $\mathbb{F}$, denoted $\mathbb{F}^n = \big\{(a_1, \dots, a_n) : a_i \in \mathbb{F}\big\}$ with component-wise addition

$$(a_1, \dots, a_n) + (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n)$$

and scalar multiplication

$$\lambda \cdot (a_1, \dots, a_n) = (\lambda a_1, \dots, \lambda a_n)$$

Proving that this is a vector space is an exercise. It is also a special case of the next example.

(iv) Let X be any set, and $\mathbb{F}^X = \{f : X \to \mathbb{F}\}$ be the set of *all* functions $X \to \mathbb{F}$. This is a vector space, with addition defined pointwise:

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

and multiplication also defined pointwise:

$$(\lambda \cdot f)(x) = \lambda f(x)$$

if $\lambda \in \mathbb{F}$, $f, g \in \mathbb{F}^X$, $x \in X$. If $X = \{1, \dots, n\}$, then $\mathbb{F}^X = \mathbb{F}^n$ and we have the previous example.

Proof that $\mathbb{F}^X$ is a vector space.

- As $+$ in $\mathbb{F}$ is commutative, we have

$$(f + g)(x) = f(x) + g(x) = g(x) + f(x) = (g + f)(x),$$

so $f + g = g + f$. Similarly, f in $\mathbb{F}$ associative implies $f + (g + h) = (f + g) + h$, and that $(-f)(x) = -f(x)$ and $0(x) = 0$.

- Axioms for scalar multiplication follow from the relationship between $\cdot$ and $+$ in $\mathbb{F}$. Check this yourself! □

(v) $\mathbb{C}$ is a vector space over $\mathbb{R}$.

Lemma 1.2. *Let V be a vector space over $\mathbb{F}$.*

(i) For all $\lambda \in \mathbb{F}$, $\lambda \cdot 0 = 0$, and for all $v \in V$, $0 \cdot v = 0$.

(ii) Conversely, if $\lambda \cdot v = 0$ and $\lambda \in \mathbb{F}$ has $\lambda \neq 0$, then $v = 0$.

(iii) For all $v \in V$, $-1 \cdot v = -v$.

Proof.

(i) $\lambda \cdot 0 = \lambda \cdot (0 + 0) = \lambda \cdot 0 + \lambda \cdot 0 \implies \lambda \cdot 0 = 0$.
$0 \cdot v = (0 + 0) \cdot v = 0 \cdot v + 0 \cdot v \implies 0 \cdot v = 0$.

(ii) As $\lambda \in \mathbb{F}$, $\lambda \neq 0$, there exists $\lambda^{-1} \in \mathbb{F}$ such that $\lambda^{-1}\lambda = 1$, so $v = (\lambda^{-1}\lambda) \cdot v = \lambda^{-1}(\lambda \cdot v)$, hence if $\lambda \cdot v = 0$, we get $v = \lambda^{-1} \cdot 0 = 0$ by (i).

(iii) $0 = 0 \cdot v = (1 + (-1)) \cdot v = 1 \cdot v + (-1 \cdot v) = v + (-1 \cdot v) \implies -1 \cdot v = -v$. □

We will write λv rather than $\lambda \cdot v$ from now on, as the lemma means this will not cause any confusion.

1.2 Subspaces

Definition. Let V be a vector space over $\mathbb{F}$. A subset $U \subseteq V$ is a *vector subspace* (or just a *subspace*), written $U \leq V$, if the following holds:

(i) $0 \in U$;
(ii) If $u_1, u_2 \in U$, then $u_1 + u_2 \in U$;
(iii) If $u \in U$, $\lambda \in \mathbb{F}$, then $\lambda u \in U$.

Equivalently, U is a subspace if $U \subseteq V$, $U \neq \emptyset$ (U is non-empty) and for all $u, v \in U$, $\lambda, \mu \in \mathbb{F}$, $\lambda u + \mu v \in U$.

Lemma 1.3. *If V is a vector space over $\mathbb{F}$ and $U \leq V$, then U is a vector space over $\mathbb{F}$ under the restriction of the operations $+$ and $\cdot$ on V to U. (Proof is an exercise.)*

Examples 1.4.

(i) $\{0\}$ and V are always subspaces of V.

(ii) $\{(r_1, \ldots, r_n, 0, \ldots, 0) : r_i \in \mathbb{R}\} \subseteq \mathbb{R}^{n+m}$ is a subspace of $\mathbb{R}^{n+m}$.

(iii) The following are all subspaces of sets of functions:

$$\begin{aligned} C^1(\mathbb{R}) &= \left\{f : \mathbb{R} \to \mathbb{R} \mid f \text{ continuous and differentiable}\right\} \\ &\subseteq C(\mathbb{R}) = \left\{f : \mathbb{R} \to \mathbb{R} \mid f \text{ continuous}\right\} \\ &\subseteq \mathbb{R}^{\mathbb{R}} = \{f : \mathbb{R} \to \mathbb{R}\}. \end{aligned}$$

Proof. f, g continuous implies $f + g$ is, and λf is, for $\lambda \in \mathbb{R}$; the zero function is continuous, so $C(\mathbb{R})$ is a subspace of $\mathbb{R}^{\mathbb{R}}$, similarly for $C^1(\mathbb{R})$. $\square$

(iv) Let X be any set, and write

$$\mathbb{F}[X] = (\mathbb{F}^X)_{\text{fin}} = \left\{f : X \to \mathbb{F} \mid f(x) \neq 0 \text{ for only finitely many } x \in X\right\}.$$

This is the set of *finitely supported functions*, which is is a subspace of $\mathbb{F}^X$.

Proof that this is a subspace. $f(x) = 0 \implies \lambda\, f(x) = 0$, so if $f \in (\mathbb{F}^X)_{\text{fin}}$, then so is λf. Similarly,

$$(f+g)^{-1}(\mathbb{F}\backslash\{0\}) \subseteq f^{-1}(\mathbb{F}\backslash\{0\}) \cup g^{-1}(\mathbb{F}\backslash\{0\})$$

and if these two are finite, so is the LHS. $\square$

Special case. Consider the case $X = \mathbb{N}$, so

$$\mathbb{F}[\mathbb{N}] = (\mathbb{F}^{\mathbb{N}})_{\text{fin}} = \left\{(\lambda_0, \lambda_1, \ldots) \mid \text{only finitely many } \lambda_i \text{ are non-zero}\right\}.$$

We write x^i for the function which sends $i \mapsto 1$, $j \mapsto 0$ if $j \neq i$; that is, for the tuple $(0, \ldots, 0, 1, 0, \ldots)$ in the ith place. Thus

$$\mathbb{F}[\mathbb{N}] = \left\{\sum \lambda_i \mid \text{only finitely many } \lambda_i \text{ non-zero}\right\}.$$

Note that we can do better than a vector space here; we can define multiplication by

$$\left(\sum \lambda_i\, x^i\right)\left(\sum \mu_j\, x^j\right) = \sum \lambda_i\, \mu_j \cdot x^{i+j}.$$

This is still in $\mathbb{F}[\mathbb{N}]$. It is more usual to denote this $\mathbb{F}[x]$, the polynomials in x over $\mathbb{F}$ (and this is a formal definition of the *polynomial ring*).

1.3 Bases

8 Oct

Definition. Suppose V is a vector space over $\mathbb{F}$, and $S \subseteq V$ is a subset of V. Then v is a *linear combination* of elements of S if there is some $n > 0$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$, $v_1, \ldots, v_n \in S$ such that $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$ or if $v = 0$.

Write $\langle S \rangle$ for the *span* of S, the set of all linear combinations of elements of S.

Notice that it is important in the definition to use only finitely many elements – infinite sums do not make sense in arbitrary vector spaces.

We will see later why it is convenient notation to say that 0 is a linear combination of $n = 0$ elements of S.

Example 1.5. $\langle \emptyset \rangle = \{0\}$.

Lemma 1.6.

(i) $\langle S \rangle$ *is a subspace of* V.

(ii) If $W \leq V$ *is a subspace, and* $S \subseteq W$, *then* $\langle S \rangle \leq W$; *that is,* $\langle S \rangle$ *is the smallest subset of* V *containing* S.

Proof. (i) is immediate from the definition. (ii) is immediate, by (i) applied to W. □

Definition. We say that S *spans* V if $\langle S \rangle = V$.

Example 1.7. The set $\{(1,0,0),(0,1,0),(1,1,0),(7,8,0)\}$ spans $W = \big\{(x,y,z) \mid z = 0\big\} \leq \mathbb{R}^3$.

Definition. Let $v_1, \ldots, v_n$ be a sequence of elements in V. We say they are *linearly dependent* if there exist $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$, not all zero, such that

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

which we call a *linear relation* among the v_i. We say that $v_1, \ldots, v_n$ are *linearly independent* if they are not linearly dependent; that is, if there is no linear relation among them, or equivalently if

$$\sum_{i=1}^{n} \lambda_i\, v_i = 0 \implies \lambda_i = 0 \text{ for all } i.$$

We say that a subset $S \subseteq V$ is *linearly independent* if every finite sequence of distinct elements in S is linearly independent.

Note that if $v_1, \dots, v_n$ is linearly independent, then so is every reordering $v_{\pi(1)}, \dots, v_{\pi(n)}$.

- If $v_1, \dots, v_n$ are linearly independent, and $v_{i_1}, \dots, v_{i_k}$ is a subsequence, then the subsequence is also linearly independent.
- If some $v_i = \mathbf{0}$, then $1 \cdot \mathbf{0} = \mathbf{0}$ is a linear relation, so $v_1, \dots, v_n$ is not linearly independent.
- If $v_i = v_j$ for some $i \neq j$, then $1 \cdot v_i + (-1)\, v_j = 0$ is a linear relation, so the sequence isn't linearly independent.
- If $|S| < \infty$, say $S = \{v_1, \dots, v_n\}$, then S is linearly independent if and only if $v_1, \dots, v_n$ are linearly independent.

Example 1.8. Let $V = \mathbb{R}^3$, $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$, and then

$$\lambda_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ 0 \end{pmatrix}$$

is zero if and only if $\lambda_1 = \lambda_2 = 0$, and so S is linearly independent.

Exercises:

(i) Show that $v_1, v_2 \in V$ are linearly dependent if and only if $v_1 = 0$ *or* $v_2 = \lambda v_1$ for some $\lambda \in \mathbb{F}$.

(ii) Let $S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \right\}$, then

$$\lambda_1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 1 \\ 1 & 0 & 0 \end{pmatrix}}_{A} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix},$$

so linear independence of S is the same as $A\lambda = 0 \implies \lambda = 0$. Show that in this example, there are no non-zero solutions.

(iii) If $S \subseteq \mathbb{F}^n$, $S = \{v_1, \dots, v_m\}$, then show that finding a relation of linear dependence $\sum_{i=1}^m \lambda_i\, v_i$ is equivalent to solving $A\lambda = 0$, where $A = (v_1 \ \dots \ v_m)$ is an $n \times m$ matrix whose columns are the v_i.

(iv) Hence show that every collection of four vectors in $\mathbb{R}^3$ has a relation of linear dependence.

Definition. The set $S \subseteq V$ is a *basis* for V if

(i) S is linearly independent and;

(ii) S spans V.

Remark. This is slightly the wrong notion. We should order S, but we'll deal with this later.

Examples 1.9.

(i) By convention, the vector space $\{0\}$ has $\emptyset$ as a basis.

(ii) $S = \{e_1, \ldots, e_n\}$, where e_i is a vector of all zeroes except for a one in the ith position, is a basis of $\mathbb{F}^n$ called the *standard basis*.

(iii) $\mathbb{F}[x] = \mathbb{F}[\mathbb{N}] = (\mathbb{F}^{\mathbb{N}})_{\text{fin}}$ has basis $\{1, x, x^2, \ldots\}$.
More generally, $\mathbb{F}[X]$ has $\{\delta_x \mid x \in X\}$ as a basis, where

$$\delta_x(y) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise,} \end{cases}$$

so $\mathbb{F}[X]$ is, formally, the set of linear combinations of elements of X.
For amusement: $\mathbb{F}[\mathbb{N}] \leq \mathbb{F}^{\mathbb{N}}$, and $1, x, x^2, \ldots$ are linearly independent in $\mathbb{F}^{\mathbb{N}}$ as they are linearly independent in $\mathbb{F}[\mathbb{N}]$, but they do not span $\mathbb{F}^{\mathbb{N}}$, as $(1, 1, 1, \ldots) \notin \mathbb{F}[\mathbb{N}]$.
Show that if a basis of $\mathbb{F}^{\mathbb{N}}$ exists, then it is uncountable.

Lemma 1.10. *A set S is a basis of V if and only if every vector $v \in V$ can be written uniquely as a linear combination of elements of S.*

Proof. ($\Leftarrow$) Writing v as a linear combination of elements of S for every $v \in V$ means that $\langle S \rangle = V$. Uniquely means that, in particular, 0 can be written uniquely, and so S is linearly independent.

($\Rightarrow$) If $v = \sum_{i=1}^n \lambda_i v_i = \sum_{i=1}^n \mu_i v_i$, where $v_i \in S$ and $i = 1, \ldots, n$, then $\sum_{i=1}^n (\lambda_i - \mu_i) v_i = 0$, and since the v_i are linearly independent, $\lambda_i = \mu_i$ for all i. $\square$

Observe: if S is a basis of V, $|S| = d$ and $|\mathbb{F}| = q < \infty$ (for example, $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$, and $q = p$), then the lemma gives $|V| = q^d$, which implies that d is the same, regardless of choice of basis for V, that is every basis of V has the same size. In fact, this is true when $\mathbb{F} = \mathbb{R}$ or indeed when $\mathbb{F}$ is arbitrary, which means we must give a proof without counting. We will now slowly show this, showing that the language of vector spaces reduces the proof to a statement about matrices – Gaussian elimination (row reduction) – we're already familiar with.

Definition. V is *finite dimensional* if there exists a finite set S which spans V.

10 Oct

Theorem 1.11

Let V be a vector space over $\mathbb{F}$, and let S span V. If S is finite, then S has a subset which is a basis for V. In particular, if V is finite dimensional, then V has a basis.

Proof. If S is linearly independent, then we're done. Otherwise, there exists a relation of linear dependence, $\sum_{i=1}^n c_i v_i = 0$, where not all c_i are zero (for $c_i \in \mathbb{F}$). Suppose $c_{i_0} \neq 0$, then we get $c_{i_0} v_{i_0} = -\sum_{j \neq i_0} c_j v_j$, so $v_{i_0} = -\sum c_j v_j / c_{i_0}$, and hence we claim $\langle v_1, \ldots, v_m \rangle = \langle v_1, \ldots, v_{i_0 - 1}, v_{i_0 + 1}, \ldots, v_m \rangle$ (proof is an exercise). So removing v_{i_0} doesn't change the span. We repeat this process, continuing to remove elements until we have a basis. $\square$

Remark. If $S = \{0\}$, say with $V = \{0\}$, then the proof says remove 0 from the set S to get $\emptyset$, which is why it is convenient to say that $\emptyset$ is a basis of $\{0\}$.

Theorem 1.12

Let V be a vector space over $\mathbb{F}$, and V finite dimensional. If $v_1, \ldots, v_r$ are linearly independent vectors, then there exist elements $v_{r+1}, \ldots, v_n \in V$ such that $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$ is a basis.

That is, any linearly independent set can be extended to a basis of V.

Remark. This theorem is true without the assumption that V is finite dimensional – any vector space has a basis. The proof is similar to what we give below, plus a bit of fiddling with the axiom of choice. The interesting theorems in this course are about finite dimensional vector spaces, so you're not missing much by this omission.

First, we prove a lemma.

Lemma 1.13. *Let $v_1, \ldots, v_m$ be linearly independent, and $v \in V$. Then $v \notin \langle v_1, \ldots, v_m \rangle$ if and only if $v_1, \ldots, v_m, v$ are linearly independent.*

Proof. ($\Leftarrow$) If $v \in \langle v_1, \ldots, v_m \rangle$, then $v = \sum_{i=1}^m c_i v_i$ for some $c_i \in \mathbb{F}$, so $\sum_{i=1}^m c_i v_i + (-1) \cdot v$ is a non-trivial relation of linear dependence.

($\Rightarrow$) Conversely, if $v_1, \ldots, v_m, v$ are linearly dependent, then there exist c_i, b such that $\sum c_i v_i + bv = 0$, with not all c_i, b zero. Then if $b = 0$, we get $\sum c_i v_i = 0$, which is a non-trivial relation on the linearly independent v_i, which is not possible, so $b \neq 0$. So $v = -\sum c_i v_i / b$ and $v \in \langle v_1, \ldots, v_m \rangle$. □

Proof of theorem 1.12. Since V is finite dimensional, there is a finite spanning set $S = \{w_1, \ldots, w_d\}$. Now, if $w_i \in \langle v_1, \ldots, v_r \rangle$ for all i, then $V = \langle w_1, \ldots, w_d \rangle \subseteq \langle v_1, \ldots, v_r \rangle$, so in this case $v_1, \ldots, v_r$ is already a basis.

Otherwise, there is some $w_i \notin \langle v_1, \ldots, v_r \rangle$. But then the lemma implies that $v_1, \ldots, v_r, w_i$ is linearly independent.

We repeat this process, adding elements in S, till we have a basis. □

Theorem 1.14

Let V be a vector space over $\mathbb{F}$. Let $S = \{v_1, \ldots, v_m\}$ span V and $L = \{w_1, \ldots, w_n\}$ be linearly independent. Then $m \geq n$.

In particular, if $\mathfrak{B}_1, \mathfrak{B}_2$ are two bases of V, then $|\mathfrak{B}_1| = |\mathfrak{B}_2|$.

Proof. As the v_k's span V, we can write each w_i as a linear combination of the v_k's, $w_i = \sum_{k=1}^m c_{ki}\, v_k$, for some $c_{ki} \in \mathbb{F}$. Now we know the w_i's are linearly independent, which means $\sum_i \lambda_i\, w_i = 0 \implies \lambda_i = 0$ for all i. But

$$\textstyle\sum_i \lambda_i\, w_i = \sum_i \lambda_i \left(\sum_k c_{ki}\, v_k\right) = \sum_k \left(\sum_i c_{ki}\, \lambda_i\right) v_k.$$

We write $C = (c_{ki})$ for the $m \times n$ matrix formed by the coefficients c_{ki}. Observe that the rules of matrix multiplication are such that the coefficient of v_k in $\sum \lambda_i\, w_i$ is the kth entry of the column vector $C\lambda$.

If $m < n$, we learned in *Vectors & Matrices* that there is a non-trivial solution $\lambda \neq 0$. (We have m linear equations in n variables, so a non-zero solution exists; the proof is by row reduction.) This contradicts the w_i's as linearly independent. So $m \geq n$.

Now, if $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are bases, then apply this to $S = \mathfrak{B}_1$, $L = \mathfrak{B}_2$ to get $|\mathfrak{B}_1| \geq |\mathfrak{B}_2|$. Similarly apply this $S = \mathfrak{B}_2$, $L = \mathfrak{B}_1$ to get $|\mathfrak{B}_2| \geq |\mathfrak{B}_1|$, and so $|\mathfrak{B}_1| = |\mathfrak{B}_2|$. $\square$

Definition. Let V be a vector space over a field $\mathbb{F}$. Then the *dimension* of V, denoted by $\dim V$, is the number of elements in a basis of V.

Example 1.15. $\dim \mathbb{F}^n = n$, as $e_1, \ldots, e_n$ is a basis, called *the standard basis*, where $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$

Corollary 1.16.

(i) If S spans V, then $|S| \geq \dim V$, with equality if and only if S is a basis.

(ii) If $L = \{v_1, \ldots, v_k\}$ is linearly independent, then $|L| \leq \dim V$, with equality if and only if L is a basis.

Proof. Immediate. Theorem 1.11 implies (i) and theorem 1.12 implies (ii). $\square$

Lemma 1.17. *Let $W \leq V$, and V be finite dimensional. Then W is finite dimensional, and* $\dim W \leq \dim V$. *Moreover,* $\dim W = \dim V$ *if and only if $W = V$.*

Proof. The subtle point is to show that W is finite dimensional.

Let $w_1, \ldots, w_r$ be linearly independent vectors in W. Then they are linearly independent when considered as vectors in V, so $r \leq \dim V$ by our theorem. If $\langle w_1, \ldots, w_r \rangle \neq W$, then there is some $w \in W$ with $w \notin \langle w_1, \ldots, w_r \rangle$, and so by lemma 1.13, $w_1, \ldots, w_r, w$ is linearly independent, and $r + 1 \leq \dim V$.

Continue in this way finding linearly independent vectors in W, and we must stop after at most $(\dim V)$ steps. When we stop, we have a finite basis of W, so W is finite dimensional, and the rest of the theorem is immediate. $\square$

Lemma 1.18. *Let V be finite dimensional and S any spanning set. Then there is a* finite *subset S' of S which still spans V, and hence a finite subset of that which is a basis.*

Proof. As V is finite dimensional, there is a finite spanning set $\{v_1, \ldots, v_n\}$. Now, as S spans V, we can write each v_i as a *finite* linear combination of elements of S.

But when you do this, you use only finitely many elements of S for each i. Hence as there are only finitely many v_i (there are n of them!), this only uses finitely many elements of S. We call this finite subset S'. By construction, $V = \langle v_1, \ldots, v_n \rangle \subseteq \langle S' \rangle$. $\square$

1.4 Linear maps and matrices

Definition. Let V and W be vector spaces over $\mathbb{F}$, and $\varphi : V \to W$ a map. We say that φ is *linear* if

(i) φ is a homomorphism of abelian groups; that is, $\varphi(0) = 0$ and for all $v_1, v_2 \in$

V, we have $\varphi(v_1 + v_2) = \varphi(v_1) + \varphi(v_2)$.

(ii) φ respects scalar multiplication; that is, $\varphi(\lambda v) = \lambda\,\varphi(v)$ for all $\lambda \in \mathbb{F}, v \in V$.

Combining these two conditions, we see that a map φ is linear if and only if

$$\varphi(\lambda_1\, v_1 + \lambda_2\, v_2) = \lambda_1\, \varphi(v_1) + \lambda_2\, \varphi(v_2)$$

for all $\lambda_1, \lambda_2 \in \mathbb{F}$, $v_1, v_2 \in V$.

Definition. We write $\mathcal{L}(V, W)$ to be the set of linear maps from V to W; that is, $\mathcal{L}(V, W) = \big\{\varphi : V \to W \mid \varphi \text{ linear}\big\}$.

A linear map $\varphi : V \to W$ is an *isomorphism* if there is a linear map $\psi : W \to V$ such that $\varphi\psi = 1_W$ and $\psi\varphi = 1_V$.

Notice that if φ is an isomorphism, then in particular φ is a bijection on sets. The converse also holds:

Lemma 1.19. *A linear map φ is an isomorphism if φ is a bijection; that is, if φ^{-1} exists as a map of sets.*

Proof. We must show that $\varphi^{-1} : W \to V$ is linear; that is,

$$\varphi^{-1}(a_1 w_1 + a_2 w_2) = a_1\, \varphi^{-1}(w_1) + a_2\, \varphi^{-1}(w_2). \tag{$*$}$$

But we have

$$\varphi\left(a_1\, \varphi^{-1}(w_1) + a_2\, \varphi^{-1}(w_2)\right) = a_1\, \varphi(\varphi^{-1}(w_1)) + a_2\, \varphi(\varphi^{-1}(w_2)) = a_1\, w_1 + a_2\, w_2,$$

as φ is linear. Now apply φ^{-1} to get $(*)$. □

Lemma 1.20. *If $\varphi : V \to W$ is a vector space isomorphism, then* $\dim V = \dim W$.

Proof. Let $b_1, \ldots, b_n$ be a basis of V. We claim that $\varphi(b_1), \ldots, \varphi(b_n)$ is a basis of W. 12 Oct
First we check linear independence: Suppose

$$0 = \sum_{i=1}^{n} \lambda_i\, \varphi(b_i) = \varphi\left(\sum_{i=1}^{n} \lambda_i\, b_i\right).$$

As φ is injective, so $\sum \lambda_i\, b_i = 0$, and hence as the b_i are linearly independent, $\lambda_i = 0$ for $i = 1, \ldots, n$. So $\varphi(b_i)$ are linearly independent.

Then we check they span: since φ is surjective, for all $w \in W$, we have $w = \varphi(v)$ for some $v \in V$. But $v = \sum \lambda_i\, b_i$ for some $\lambda_i \in \mathbb{F}$, as the b_i span V. But then $w = \varphi(v) = \sum \lambda_i\, \varphi(b_i)$, and the $\varphi(b_i)$ span W.

Since they both have a basis of the same size, it follows that $\dim V = \dim W$. □

Definition. If $b_1, \ldots, b_n$ are a basis of V, and $v = \sum_i \lambda_i\, v_i$, we say $\lambda_1, \ldots, \lambda_n$ are the *coordinates* of v with respect to the basis $b_1, \ldots, b_n$.

Here is another view of what the coordinates of a vector mean:

Proposition 1.21. *Let V be a finite-dimensional vector space over $\mathbb{F}$, with* $\dim V = n$. *Then there is a bijection*

$$\{\textit{ordered bases } b_1, \ldots, b_n \textit{ of } V\} \xrightarrow{\sim} \left\{\varphi : \mathbb{F}^n \xrightarrow{\sim} V\right\}.$$

The idea of the proposition is that coordinates of a vector with respect to a basis define a point in $\mathbb{F}^n$, and hence a choice of a basis is a choice of an isomorphism of our vector space V with $\mathbb{F}^n$.

Proof. Given an ordered basis $b_1, \dots, b_n$ of V, call it $\mathfrak{B}$, we can write every vector $v \in V$ as $v = \sum \lambda_i\, b_i$ for unique $\lambda_i, \dots, \lambda_n \in \mathbb{F}$. Define $\alpha_{\mathfrak{B}} : V \to \mathbb{F}^n$ by

$$\alpha_{\mathfrak{B}}(v) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \sum_{i=1}^{n} \lambda_i\, e_i,$$

where $\{e_i\}$ is the standard basis of $\mathbb{F}^n$.

It is clear that $\alpha_{\mathfrak{B}}$ is well-defined, linear and an isomorphism, and the inverse sends $(\lambda_1, \dots, \lambda_n) \mapsto \sum \lambda_i\, b_i$.

This defines a map $\{\text{ordered bases } \mathfrak{B}\} \to \left\{\alpha : V \xrightarrow{\sim} \mathbb{F}^n\right\}$ taking $\mathfrak{B} \mapsto \alpha_{\mathfrak{B}}$.

To see that this map is a bijection, suppose we are given $\alpha : V \to \mathbb{F}^n$ an isomorphism. Then $\alpha^{-1} : \mathbb{F}^n \to V$ is also an isomorphism, and we define $b_i = \alpha^{-1}(e_i)$. The proof of the previous lemma showed that $b_1, \dots, b_n$ is a basis of V. It is clear that for this ordered basis $\mathfrak{B}$, $\alpha_{\mathfrak{B}} = \alpha$. $\square$

Let V and W be finite dimensional vector spaces over $\mathbb{F}$, and choose bases $v_1, \dots, v_n$ and $w_1, \dots, w_m$ of V and W, respectively. Then we have the diagram:

$$\begin{tikzcd} \mathbb{F}^n \arrow[d, "\cong"'] & \mathbb{F}^m \arrow[d, "\cong"] \\ V \arrow[r, "\alpha"'] & W \end{tikzcd}$$

Now, suppose $\alpha : V \to W$ is a linear map. As α is linear, and every $v \in V$ can be written as $v = \sum \lambda_i\, v_i$ for some $\lambda_1, \dots, \lambda_n$, we have

$$\alpha(v) = \alpha\left(\sum_{i=1}^{n} \lambda_i\, v_i\right) = \sum_{i=1}^{n} \lambda_i\, \alpha(v_i),$$

so α is determined by its values $\alpha(v_1), \dots, \alpha(v_n)$. But then write each $\alpha(v_i)$ as a sum of basis elements $w_1, \dots, w_m$

$$\alpha(v_j) = \sum_{i=1}^{m} a_{ij}\, w_i \qquad j = 1, \dots, m$$

for some $a_{ij} \in \mathbb{F}$.

Hence, if $(\lambda_1, \dots, \lambda_n)$ are the coordinates of $v \in V$, with respect to a basis $v_1, \dots, v_n$; that is, if $v = \sum \lambda_i\, v_i$, then

$$\alpha(v) = \alpha\left(\sum_{j=1}^{n} \lambda_j\, v_j\right) = \sum_{i,j} a_{ij}\, \lambda_j\, w_i;$$

that is,

$$\begin{pmatrix} \sum a_{1j}\, \lambda_j \\ \sum a_{2j}\, \lambda_j \\ \vdots \\ \sum a_{mj}\, \lambda_j \end{pmatrix} = A \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}$$

are the coordinates of $\alpha(v)$ with respect to $w_1, \ldots, w_m$.

That is, by choosing bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ of V and W, respectively, every linear map $\alpha : V \to W$ determines a matrix $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$.

Conversely, given $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$, we can define

$$\alpha\left(\sum_{i=1}^{n} \lambda_i \, v_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}\,\lambda_j\, w_i,$$

which is a *well-defined* linear map $\alpha : V \to W$, and these constructions are inverse, and so we've proved the following theorem:

Theorem 1.22

A choice of bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ of vector spaces V and W defines an isomorphism $\mathcal{L}(V, W) \xrightarrow{\sim} \mathrm{Mat}_{m,n}(\mathbb{F})$.

Remark. Actually, $\mathcal{L}(V, W)$ is a vector space. The vector space structure is given by defining, for $a, b \in \mathbb{F}$, $\alpha, \beta \in \mathcal{L}(V, W)$,

$$(a\alpha + b\beta)\,(v) = a\,\alpha(v) + b\,\beta(v).$$

Also, $\mathrm{Mat}_{m,n}(\mathbb{F})$ is a vector space over $\mathbb{F}$, and these maps $\mathcal{L}(V, W) \rightleftarrows \mathrm{Mat}_{m,n}(\mathbb{F})$ are vector space isomorphisms.

The choice of bases for V and W define isomorphisms with $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively so that the following diagram commutes:

$$\begin{tikzcd}
\mathbb{F}^n \arrow[r, "A"] \arrow[d, "\cong"'] & \mathbb{F}^m \arrow[d, "\cong"] \\
V \arrow[r, "\alpha"'] & W
\end{tikzcd}$$

We say a diagram *commutes* if every directed path through the diagram with the same start and end vertices leads to the same result by composition. This is convenient short hand language for a bunch of linear equations – that the coordinates of the different maps that you get by composing maps in the different manners agree.

Corollary 1.23. $\dim \mathcal{L}(V, W) = \dim \mathrm{Mat}_{m,n}(\mathbb{F}) = nm = \dim V \dim W$.

Lemma 1.24. *Let $\alpha : V \to W$, $\beta : W \to U$ be linear maps of vector spaces U, V, W.*

(i) $\beta\alpha : V \to U$ is linear.

(ii) If $v_1, \ldots, v_n$ is a basis of V,
$w_1, \ldots, w_m$ is a basis of W,
$u_1, \ldots, u_r$ is a basis of U,
and $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$ is the matrix of α with respect to the v_i, w_j bases, and
$B \in \mathrm{Mat}_{r,m}(\mathbb{F})$ is the matrix of β with respect to the w_j, u_k bases,
then the matrix of $\beta\alpha : V \to U$ with respect to the v_i, u_k bases is BA.

Proof.

(i) Exercise.

(ii) We have from our earlier work

$$\alpha(v_j) = \sum_{i=1}^{m} a_{ij}\, w_i \qquad \text{and} \qquad \beta(w_i) = \sum_{k=1}^{r} b_{ki}\, u_k.$$

Now we have

$$(\beta\alpha)(v_j) = \beta\left(\sum_{i=1}^{m} a_{ij}\, w_i\right) = \sum_{k=1}^{r}\sum_{i=1}^{m} a_{ij}\, b_{ki}\, u_k,$$

and so the coefficient of u_k is $\sum_{i,k} b_{ki}\, a_{ij} = (BA)_{kj}$. □

Definition. A linear map $\varphi : V \to V$ is an *automorphism* if it is an isomorphism. The set of automorphisms forms a group, and is denoted

$$\begin{aligned}\mathrm{GL}(V) &= \{\varphi : V \to V \mid \varphi \text{ a linear isomorphism}\}\\ &= \{\varphi \in \mathcal{L}(V,V) \mid \varphi \text{ an isomorphism}\}\end{aligned}$$

Example 1.25. We write $\mathrm{GL}_n(\mathbb{F}) = \{\varphi : \mathbb{F}^n \to \mathbb{F}^n, \varphi \text{ isomorphism}\} = \mathrm{GL}(\mathbb{F}^n)$.

Exercise: Show that if $\varphi : V \xrightarrow{\sim} W$ is an isomorphism, then it induces an isomorphism
15 Oct of groups $\mathrm{GL}(V) \cong \mathrm{GL}(W)$, so $\mathrm{GL}(V) \cong \mathrm{GL}_{\dim V}(\mathbb{F})$.

Lemma 1.26. *Let $v_1, \ldots, v_n$ be a basis of V and $\varphi : V \to V$ be an isomorphism; that is, let $\varphi \in \mathrm{GL}(V)$. Then we showed that $\varphi(v_1), \ldots, \varphi(v_n)$ is also a basis of V and hence*

(i) If $v_1 = \varphi(v_1), \ldots, v_n = \varphi(v_n)$, then $\varphi = \mathrm{id}_V$. In other words, we get the same ordered basis if and only if φ is the identity map.

(ii) If $v'_1, \ldots, v'_n$ is another basis of V, then the linear map $\varphi : V \to V$ defined by

$$\varphi\left(\sum \lambda_i\, v_i\right) = \sum \lambda_i\, v'_i$$

(that is, the map sending $v_i \mapsto v'_i$) is an isomorphism.

Proof. Define its inverse $\psi : V \to V$ by $v'_i \mapsto v_i$; that is,

$$\psi\left(\sum \lambda_i\, v'_i\right) = \sum \lambda_i\, v_i.$$

Then it is clear $\varphi\psi = \psi\varphi = \mathrm{id}_V : V \to V$. □

So (i) and (ii) say that:

Proposition 1.27. $\mathrm{GL}(V)$ *acts simply and transitively on the set of bases; that is, given $v'_1, \ldots, v'_n$ a basis, there is a unique $\varphi \in \mathrm{GL}(V)$ such that $\varphi(v_1) = v'_1, \ldots, \varphi(v_n) = v'_n$.*

Corollary 1.28. $\left|\mathrm{GL}_n(\mathbb{F}_p)\right| = (p^n - 1)(p^n - p) \cdots \left(p^n - p^{n-1}\right)$.

Proof. It is enough to count ordered bases of $\mathbb{F}_p^n$, which is done by proceeding as follows:

Choose v_1, which can be any non-zero element, so we have $p^n - 1$ choices.
Choose v_2, any non-zero element not a multiple of v_1, so $p^n - p$ choices.
Choose v_3, any non-zero element not in $\langle v_1, v_2\rangle$, so $p^n - p^2$ choices.

$\vdots$

Choose v_n, any non-zero element not in $\langle v_1, \ldots, v_{n-1}\rangle$, so $p^n - p^{n-1}$ choices. □

Example 1.29. $\left|\mathrm{GL}_2(\mathbb{F}_p)\right| = p\,(p-1)^2\,(p+1)$.

Remark. We could express the same proof by saying that a matrix $A \in \mathrm{Mat}_n(\mathbb{F}_p)$ is invertible if and only if all of its columns are linearly independent, and the proof works by picking each column in turn.

Let $v_1, \dots, v_n$ be a basis of V, and $\varphi \in \mathrm{GL}(V)$. Then $v_i' = \varphi(v_i)$ is a new basis of V. Let A be the matrix of φ with respect to the original basis $v_1, \dots, v_n$ for both source and target $\varphi : V \to V$. Then

$$\varphi(v_i) = v_i' = \sum_j a_{ji}\, v_j,$$

so the columns of A are the coordinates of the new basis in terms of the old.

We can also express this by saying the following diagram commutes:

$$\begin{tikzcd} \mathbb{F}^n \arrow[r, "\cong"] \arrow[d, dotted, "A"'] & V \arrow[d, "\varphi"] \\ \mathbb{F}^n \arrow[r, "\cong"'] & V \end{tikzcd} \qquad \begin{array}{c} e_i \mapsto v_i \\ \\ e_i \mapsto v_i \end{array}$$

Conversely, if $v_1, \dots, v_n$ is a basis of V, and $v_1', \dots, v_n'$ is another basis, then we can define $\varphi : V \to V$ by $\varphi(v_i) = v_i'$, and we can express this by saying the following diagram commutes.

$$\begin{tikzcd} \mathbb{F}^n \arrow[r, "\cong"] \arrow[dr, "\cong"'] & V \arrow[d, dotted, "\varphi"] \\ & V \end{tikzcd} \qquad \begin{tikzcd} e_i \arrow[r, mapsto] \arrow[dr] & v_i \arrow[d, dotted, mapsto, "\varphi"] \\ & v_i' \end{tikzcd}$$

This is just language meant to clarify the relation between changing bases, and bases as giving isomorphisms with a fixed $\mathbb{F}^n$. If it instead confuses you, feel free to ignore it. In contrast, here is a practical and important question about bases and linear maps, which you can't ignore:

Consider a linear map $\alpha : V \to W$. Let $v_1, \dots, v_n$ be a basis of V, $w_1, \dots, w_n$ of W, and A be the matrix of α. If we have new bases $v_1', \dots, v_n'$ and $w_1', \dots, w_n'$, then we get a new matrix of α with respect to this basis. *What is the matrix with respect to these new bases?* We write

$$v_i' = \sum_j p_{ji}\, v_j \qquad w_i' = \sum_j q_{ji}\, w_j.$$

Exercise 1.30. Show that $w_i' = \sum_j q_{ji}\, w_j$ if and only if $w_i = \sum_j (Q^{-1})_{ji}\, w_j'$, where $Q = (q_{ab})$.

Then we have

$$\alpha(v_i') = \sum_j p_{ji}\, \alpha(v_j) = \sum_{j,k} p_{ji}\, a_{kj}\, w_k = \sum_{j,k,l} p_{ji}\, a_{kj}\, (Q^{-1})_{lk}\, w_l' = \sum_l (Q^{-1}AP)_{li}\, w_l',$$

and we see that the matrix is $Q^{-1}AP$.

Finally, a definition.

Definition. (i) Two matrices $A, B \in \mathrm{Mat}_{m,n}(\mathbb{F})$ are said to be *equivalent* if they represent the same linear map $\mathbb{F}^n \to F^m$ with respect to different bases, that is there exist $P \in GL_n(\mathbb{F}), Q \in GL_m(\mathbb{F})$ such that

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

(ii) The linear maps $\alpha : V \to W$ and $\beta : V \to W$ are *equivalent* if their matrices look the same after an appropriate choice of bases; that is, if there exists an isomorphism $p \in \mathrm{GL}(V)$, $q \in \mathrm{GL}(W)$ such that the following diagram commutes:

$$\begin{tikzcd} V \arrow[r, "\alpha"] & W \\ V \arrow[u, "p"] \arrow[r, "\beta"'] & W \arrow[u, "q"'] \end{tikzcd}$$

That is to say, if $q^{-1}\alpha p = \beta$.

1.5 Conservation of dimension: the Rank-nullity theorem

Definition. For a linear map $\alpha : V \to W$, we define the *kernel* to be the set of all elements that are mapped to zero

$$\ker \alpha = \big\{x \in V : \alpha(x) = 0\big\} = K \leq V$$

and the *image* to be the points in W which we can reach from V

$$\operatorname{Im} \alpha = \alpha(V) = \big\{\alpha(v) : v \in V\big\} \leq W.$$

Proving that these are subspaces is left as an exercise.

We then say that $r(\alpha) = \dim \operatorname{Im} \alpha$ is the *rank* and $n(\alpha) = \dim \ker \alpha$ is the *nullity*.

Theorem 1.31: Rank-nullity theorem

For a linear map $\alpha : V \to W$, where V is finite dimensional, we have

$$r(\alpha) + n(\alpha) = \dim \operatorname{Im} \alpha + \dim \ker \alpha = \dim V.$$

Proof. Let $v_1, \dots, v_d$ be a basis of $\ker \alpha$, and extend it to a basis of V, say, $v_1, \dots, v_d, v_{d+1}, \dots, v_n$. We show the following claim, which implies the theorem immediately:

Claim. $\alpha(v_{d+1}), \dots, \alpha(v_n)$ is a basis of $\operatorname{Im} \alpha$.

Proof of claim. Span: if $w \in \operatorname{Im} \alpha$, then $w = \alpha(v)$ for some $v \in V$. But $v_1, \dots, v_n$ is a basis, so there are some $\lambda_1, \dots, \lambda_n \in \mathbb{F}$ with $v = \sum \lambda_i v_i$. Then

$$\alpha(v) = \alpha\left(\sum \lambda_i v_i\right) = \sum_{i=d+1}^{n} \lambda_i \alpha(v_i)$$

as $\alpha(v_1) = \cdots = \alpha(v_d) = 0$; that is, $\alpha(v_{d+1}), \dots, \alpha(v_n)$ span $\operatorname{Im} \alpha$.

Linear independence: we have

$$\sum_{i=d+1}^{n} \lambda_i\, \alpha(v_i) = 0 \implies \alpha\left(\sum_{i=d+1}^{n} \lambda_i\, v_i\right) = 0.$$

And hence $\sum_{i=d+1}^{n} \lambda_i\, v_i \in \ker \alpha$. But $\ker \alpha$ has basis $v_1, \dots, v_d$, and so there are $\mu_1, \dots, \mu_d \in \mathbb{F}$ such that

$$\sum_{i=d+1}^{n} \lambda_i\, v_i = \sum_{i=1}^{d} \mu_i\, v_i.$$

But this is a relation of linear dependence on $v_1, \dots, v_n$, which is a basis of V, so we must have

$$-\mu_1 = -\mu_2 = \cdots = -\mu_d = \underbrace{\lambda_{d+1} = \cdots = \lambda_n}_{\text{hence linearly independent}} = 0. \qquad \square$$

Corollary 1.32. *Let $\alpha : V \to W$ be a linear map between finite dimensional spaces V and W. If $\dim V = \dim W$, then $\alpha : V \to W$ is an isomorphism if and only if α is injective, and if and only if α is surjective.*

Proof. The map α is injective if and only if $\dim \ker \alpha = 0$, and so $\dim \operatorname{Im} \alpha = \dim V$ (which is $\dim W$ here), which is true if and only if α is surjective. $\square$

Remark. If $v_1, \dots, v_n$ is a basis for V, $w_1, \dots, w_m$ is a basis for W and A is the matrix of 17 Oct
the linear map $\alpha : V \to W$, then $\operatorname{Im} \alpha \xrightarrow{\sim} \langle \text{column space of } A \rangle$, $\ker \alpha \xrightarrow{\sim} \ker A$, and the isomorphism is induced by the choice of bases for V and W, that is by the isomorphisms $W \xrightarrow{\sim} \mathbb{F}^m$, $V \xleftarrow{\sim} \mathbb{F}^n$.

Remark. You'll notice that the rank-nullity theorem follows easily from our basic results about how linearly independent sets extend to bases. You'll recall that these results in turn depended on row and column reduction of matrices. We'll now show that in turn they imply the basic results about row and column reduction – the first third of this course is really just learning fancy language in which to rephrase Gaussian elimination.

The language will be useful in future years, especially when you learn geometry. However it doesn't really help when you are trying to solve linear equations – that is, finding the kernel of a linear transformation. For that, there's not much you can say other than: write the linear map in terms of a basis, as a matrix, and row and column reduce!

Theorem 1.33

(i) Let $A \in \operatorname{Mat}_{m,n}(\mathbb{F})$. Then A is equivalent to

$$B = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \ddots & \ddots & \vdots \\ \vdots & 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & 0 & 0 & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

that is, there exist invertible $P \in \mathrm{GL}_m(\mathbb{F})$, $Q \in \mathrm{GL}_n(\mathbb{F})$ such that $B = Q^{-1}AP$.

(ii) The matrix B is well defined. That is, if A is equivalent to another matrix B' of the same form, then $B' = B$.

Part (ii) of the theorem is clunkily phrased. We'll phrase it better in a moment by saying that the number of ones is the rank of A, and equivalent matrices have the same rank.

Proof 1 of theorem 1.33.

(i) Let $V = \mathbb{F}^n$, $W = \mathbb{F}^m$ and $\alpha : V \to W$ be the linear map taking $x \mapsto Ax$. Define $d = \dim \ker \alpha$. Choose a basis $y_1, \dots, y_d$ of $\ker \alpha$, and extend this to a basis $v_1, \dots, v_{n-d}, y_1, \dots, y_d$ of V.

   Then by the proof of the rank-nullity theorem, $\alpha(v_i) = w_i$, for $1 \le i \le n-d$, are linearly independent in W, and we can extend this to a basis $w_1, \dots, w_m$ of W. But then with respect to these new bases of V and W, the matrix of α is just B, as desired.

(ii) The number of one's ($n-d$ here) in this matrix equals the rank of B. By definition,

$$\begin{aligned} r(A) &= \text{column rank of } A \\ &= \dim \operatorname{Im} \alpha \\ &= \dim(\text{subspace spanned by columns}) \end{aligned}$$

   So to finish the proof, we need a lemma.

Lemma 1.34. *If $\alpha, \beta : V \to W$ are equivalent linear maps, then*

$$\dim \ker \alpha = \dim \ker \beta \qquad \dim \operatorname{Im} \alpha = \dim \operatorname{Im} \beta$$

Proof of lemma. Recall $\alpha, \beta : V \to W$ are equivalent if there are some $p, q \in \mathrm{GL}(V) \times \mathrm{GL}(W)$ such that $\beta = q^{-1}\alpha p$.

$$\begin{tikzcd} V \arrow[r, "\alpha"] & W \\ V \arrow[u, "p"] \arrow[r, "\beta"'] & W \arrow[u, "q"'] \end{tikzcd}$$

Claim. $x \in \ker \beta \iff px \in \ker \alpha$.

Proof. $\beta(x) = q^{-1}\alpha p(x)$. As q is an isomorphism, $q^{-1}(\alpha(p(x))) = 0 \iff \alpha(p(x)) = 0$; that is, the restriction of p to $\ker \beta$ maps $\ker \beta$ to $\ker \alpha$; that is, $p : \ker \beta \xrightarrow{\sim} \ker \alpha$, and this is an isomorphism, as p^{-1} exists on V. (So $p^{-1}y \in \ker \beta \iff y \in \ker \alpha$.)

Similarly, you can show that q induces an isomorphism $q : \operatorname{Im} \beta \xrightarrow{\sim} \operatorname{Im} \alpha$. ☐

Note that the rank-nullity theorem implies that in the lemma, if we know $\dim \ker \alpha = \dim \ker \beta$, then you know $\dim \operatorname{Im} \alpha = \dim \operatorname{Im} \beta$, but we didn't need to use this.

Theorem 1.35: Previous theorem restated

The $\mathrm{GL}(V) \times \mathrm{GL}(W)$ orbits on $\mathcal{L}(V, W)$ are in bijection with

$$\{r : 0 \le r \le \min(\dim V, \dim W)\}$$

under the map taking $\alpha : V \to W$ to $\operatorname{rank}(\alpha) = \dim \operatorname{Im} \alpha$.

Here $\mathrm{GL}(V) \times \mathrm{GL}(W)$ acts on $\mathcal{L}(V, W)$ by $(q, p) \cdot \beta = q\beta p^{-1}$.

Hard exercise.

(i) What are the orbits of $\mathrm{GL}(V) \times \mathrm{GL}(W) \times \mathrm{GL}(U)$ on the set $\mathcal{L}(V, W) \times \mathcal{L}(W, U) = \{\alpha : V \to W, \beta : W \to V \text{ linear}\}$?
(ii) What are the orbits of $\mathrm{GL}(V) \times \mathrm{GL}(W)$ on $\mathcal{L}(V, W) \times \mathcal{L}(W, V)$?

You won't be able to do part (ii) of the exercise before the next chapter, when you learn Jordan normal form. It's worthwhile trying to do them then.

Proof 2 of theorem 1.33.

(ii) As before, no theorems were used.
(i) We'll write an algorithm to find P and Q explicitly:

Step 1: If top left $a_{11} \neq 0$, then we can clear all of the first column by row operations, and all of the first row by column operations.

Let's remember what this means.

Let E_{ij} be the matrix with a 1 in the (i, j)'th position, and zeros elsewhere. Recall that for $i \neq j$, $\left(I + \alpha E_{ij}\right) A$ is a new matrix, whose ith row is the ith row of $A + \alpha \cdot (i\text{th row of } A)$. This is an elementary row operation.

Similarly $A\left(I + \alpha E_{ij}\right)$ is an elementary column operation. As an exercise, state this precisely, as we did for the rows.

We have

$$E'_m\, E'_{m-1} \cdots E'_1\, A\, E_1 \cdots E_n = \begin{pmatrix} a_{11} & 0 \\ 0 & A' \end{pmatrix}$$

where

$$E'_i = I - \frac{a_{i1}}{a_{11}}\, E_{i1}, \qquad E_j = I - \frac{a_{1j}}{a_{11}}\, E_{1j}.$$

Step 2: if $a_{11} = 0$, either $A = 0$, in which case we are done, or there is some $a_{ij} \neq 0$.

Consider the matrix s_{ij}, which is the identity matrix with the ith row and the jth row swapped, for example $s_{12} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Exercise. $s_{ij}\, A$ is the matrix A with the ith row and the jth row swapped, As_{ij} is the matrix A with the ith and the jth column swapped.

Hence $s_{i1}\, As_{j1}$ has $(1, 1)$ entry $a_{ij} \neq 0$.

Now go back to step 1 with this matrix instead of A.

Step 3: multiply by the diagonal matrix with ones along the diagonal except for the $(1,)$, position, where it is a_{11}^{-1}.

Note it doesn't matter whether we multiply on the left or the right, we get a matrix of the form

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

Step 4: Repeat this algorithm for A''.

When the algorithm finishes, we end up with a diagonal matrix B with some ones on the diagonal, then zeros, and we have written it as a product

$$\underbrace{\begin{pmatrix} \text{row opps} \\ \text{for col } n \end{pmatrix} * \cdots * \begin{pmatrix} \text{row opps} \\ \text{for col } 1 \end{pmatrix}}_{Q} *A* \underbrace{\begin{bmatrix} \text{row opps} \\ \text{for row } 1 \end{bmatrix} * \cdots * \begin{bmatrix} \text{row opps} \\ \text{for row } n \end{bmatrix}}_{P}$$

where each $*$ is either s_{ij} or 1 times an invertible diagonal matrix (which is mostly ones, but in the i'th place is a_{ii}^{-1}).

But this is precisely writing this as a product $Q^{-1}AP$. □

Corollary 1.36. *Another direct proof of the rank-nullity theorem.*

Proof. (ii) showed that $\dim \ker A = \dim \ker B$ and $\dim \operatorname{Im} A = \dim \operatorname{Im} B$ if A and B are equivalent, by (i) of the theorem, it is enough to the show rank/nullity for B in the special form above. But here is it obvious. □

Remark. Notice that this proof really is just the Gaussian elimination argument you learned last year. We used this to prove the theorem ? on bases. So now that we've written the proof here, the course really is self contained. It's better to think that everything we've been doing as dressing up this algorithm in coordinate independent language.

In particular, we have given coordinate independent meaning to the kernel and column space of a matrix, and hence to its column rank. We should also give a coordinate independent meaning for the row space and row rank, for the transposed matrix A^{T}, and show that column rank equals row rank. This will happen in chapter 4.

1.6 Sums and intersections of subspaces

Lemma 1.37. *Let V be a vector space over $\mathbb{F}$, and $U_i \leq V$ subspaces. Then $U = \bigcap U_i$ is a subspace.*

Proof. Since $0 \in U_i$ for all i, certainly $0 \in \bigcap U_i$. And if $u, v \in U$, then $u, v \in U_i$ for all i, so $\lambda u + \mu v \in U_i$ for all i, and hence $\lambda u + \mu v \in U$. □ 19 Oct

By contrast, the union $U_1 \cup U_2$ is not a subspace unless $U_1 \subseteq U_2$ or $U_2 \subseteq U_1$.

Definition. Let $U_1, \ldots, U_r \leq V$ be subspaces. The *sum of the U_i* is the subspace denoted

$$\begin{aligned}\sum_{i=1}^{r} U_i &= U_1 + \cdots + U_r \\ &= \{u_1 + u_2 + \cdots + u_r \mid u_i \in U_i\} \\ &= \langle U_1, \ldots, U_r \rangle ,\end{aligned}$$

which is the span of $\bigcup_{i=1}^{r} U_i$.

Exercise: prove the two equalities in the definition.

Definition. The set of d-dimensional subspaces of V, $\{U \mid U \leq V, \dim U = d\}$ is called the *Grassmannian of d-planes in* V, denoted $Gr_d(V)$.

Example 1.38. We have

$$Gr_1(\mathbb{F}^2) = \left\{\text{lines } L \text{ in } \mathbb{F}^2\right\} = \mathbb{F} \cup \{\infty\},$$

as $L = \langle \lambda e_1 + \mu e_2 \rangle$. If $\lambda \neq 0$, we get $L = \langle e_1 + \gamma e_2 \rangle$, where $\gamma = \mu/\lambda \in \mathbb{F}$. If $\lambda = 0$, then $L = \langle e_2 \rangle$, which we think of as ∞.

If $\mathbb{F} = \mathbb{R}$, then this is $\mathbb{R} \cup \{\infty\}$, the circle. If $\mathbb{F} = \mathbb{C}$ then this is $\mathbb{C} \cup \{\infty\}$, the Riemann sphere.

Theorem 1.39

Suppose $U_1, U_2 \leq V$ and U_i finite dimensional. Then

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

Proof 1. Pick a basis $v_1, \ldots, v_d$ of $U_1 \cap U_2$. Extend it to a basis $v_1, \ldots, v_d, w_1, \ldots, w_r$ of U_1 and a basis $v_1, \ldots, v_d, y_1, \ldots, y_s$ of U_2.

Claim. $\{v_1, \ldots, v_d, w_1, \ldots, w_r, y_1, \ldots, y_s\}$ is a basis of $U_1 + U_2$. The claim implies the theorem immediately.

Proof of claim. Span: an element of $U_1 + U_2$ can be written $x + y$ for $x \in U_1$, $y \in U_2$, and so

$$x = \sum \lambda_i v_i + \sum \mu_j w_j \qquad y = \sum \alpha_i v_i + \sum \beta_k y_k$$

Combining these two, we have

$$x + y = \sum (\lambda_i + \alpha_i) v_i + \sum \mu_j w_j + \sum \beta_k y_k$$

Linear independence is obvious, but messy to write: if

$$\sum \alpha_i v_i + \sum \beta_j w_j + \sum \gamma_k y_k = 0,$$

then

$$\underbrace{\sum \alpha_i v_i + \sum \beta_j w_j}_{\in U_1} = \underbrace{- \sum \gamma_k y_k}_{\in U_2},$$

hence $\sum \gamma_k y_k \in U_1 \cap U_2$, and hence $\sum \gamma_k y_k = \sum \theta_i v_i$ for some θ_i, as $v_1, \dots, v_d$ is a basis of $U_1 \cap U_2$. But v_i, y_k are linearly independent, so $\gamma_k = \theta_i = 0$ for all i, k. Thus $\sum \alpha_i v_i + \sum \beta_j w_j = 0$, but as v_i, w_j are linearly independent, we have $\alpha_i = \beta_j = 0$ for all i, j. □

We can rephrase this by introducing more notation. Suppose $U_i \leq V$, and we say that $U = \sum U_i$ is a *direct sum* if every $u \in U$ can be written *uniquely* as $u = u_1 + \cdots + u_k$, for some $u_i \in U$.

Lemma 1.40. *$U_1 + U_2$ is a direct sum if and only if $U_1 \cap U_2 = \{0\}$.*

Proof. ($\Rightarrow$) Suppose $v \in U_1 \cap U_2$. Then

$$v = \underset{\in U_1}{v} + 0 = 0 + \underset{\in U_2}{v},$$

which is two ways of writing v, so uniqueness gives that $v = 0$.

($\Leftarrow$) If $u_1 + u_2 = u_1' + u_2'$, for $u_i, u_i' \in U_i$, then $\underset{\in U_1}{u_1 - u_1'} = \underset{\in U_2}{u_2 - u_2'}$.

This is in $U_1 \cap U_2 = \{0\}$, and so $u_1 = u_1'$ and $u_2 = u_2'$, and sums are unique. □

Definition. Let $U \leq V$. A *complement* to U is a subspace $W \leq V$ such that $W + U = V$ and $W \cap U = \{0\}$.

Example 1.41. Let $V = \mathbb{R}^2$, and U be the line spanned by e_1. Any line different from U is a complement to U; that is, $W = \langle e_2 + \lambda e_1 \rangle$ is a complement to U, for any $\lambda \in \mathbb{F}$.

In particular, complements are *not* unique. But they always exist:

Lemma 1.42. *Let $U \leq V$ and U finite dimensional. Then a complement to U exists.*

Proof. We've seen that U is finite dimensional. Choose $v_1, \dots, v_d$ as a basis of V, and extend it by $w_1, \dots, w_r$ to a basis $v_1, \dots, v_d, w_1, \dots, w_r$ of V.

Then $W = \langle w_1, \dots, w_r \rangle$ is a complement. □

Exercise 1.43. Show that if W' is another complement to U, then there exists a unique $\varphi : W \to U$ linear, such that $W' = \big\{w + \varphi(w) \mid w \in W\big\}$, and conversely. In other words, show that there is a bijection from the set of complements of U to $L(W, U)$.

Lemma 1.44. *If $U_1, \dots U_r \leq U$ are such that $U_1 + \dots U_r$ is a direct sum, show that* $\dim(U_1 + \cdots + U_r) = \dim U_1 + \cdots + \dim U_r$.

Proof. Exercise. Show that a union of bases for U_i is a basis for $\sum U_i$.

Now let $U_1, U_2 \leq V$ be finite dimensional subspaces of V. Choose $W_1 \leq U_1$ a complement to $U_1 \cap U_2$ in U_1, and $W_2 \leq U_2$ a complement to $U_1 \cap U_2$ in U_2. Then

Corollary 1.45.

$$U_1 + U_2 = (U_1 \cap U_2) + W_1 + W_2$$

is a direct sum, and the previous lemma gives another proof that

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

Once more, let's look at this:

> **Definition.** Let V_1, V_2 be two vector spaces over $\mathbb{F}$. Then define $V_1 \oplus V_2$, the *direct sum of* V_1 *and* V_2 to be the product set $V_1 \times V_2$, with vector space structure
>
> $$(v_1, v_2) + (w_1, w_2) = (v_1 + w_1, v_2 + w_2) \qquad \lambda\,(v_1, v_2) = (\lambda v_1, \lambda v_2)\,.$$

> **Exercises 1.46.**
>
> (i) Show that $V_1 \oplus V_2$ is a vector space. Consider the linear maps
>
> $$\begin{aligned} i_1 &: V_1 \hookrightarrow V_1 \oplus V_2 \text{ taking } v_1 \mapsto (v_1, 0) \\ i_2 &: V_2 \hookrightarrow V_1 \oplus V_2 \text{ taking } v_2 \mapsto (0, v_2) \end{aligned}$$
>
> These makes $V_1 \cong i(V_1)$ and $V_2 \cong i(V_2)$ subspaces of $V_1 \oplus V_2$ such that $iV_1 \cap iV_2 = \{0\}$, and so $V_1 \oplus V_2 = iV_1 + iV_2$, so it is a direct sum.
>
> (ii) Show that $\underbrace{\mathbb{F} \oplus \cdots \oplus \mathbb{F}}_{n \text{ times}} = \mathbb{F}^n$.

Once more let $U_1, U_2 \leq V$ be subspaces of V. Consider U_1, U_2 as vector spaces in their own right, and form $U_1 \oplus U_2$, a new vector space. (This is no longer a subspace of V.)

Lemma 1.47. *Consider the linear map* $U_1 \oplus U_2 \xrightarrow{\pi} V$ *taking* $(u_1, u_2) \mapsto u_1 + u_2$.

(i) This is linear.

(ii) $\ker \pi = \{(-w, w) \mid w \in U_1 \cap U_2\}$.

(iii) $\operatorname{Im} \pi = U_1 + U_2 \leq V$.

Proof. Exercise.

Corollary 1.48. *Show that the rank-nullity theorem implies*

$$\underbrace{\dim \ker \pi}_{=\dim U_1 \cap U_2} + \underbrace{\dim \operatorname{Im} \pi}_{=\dim(U_1+U_2)} = \dim(U_1 \oplus U_2) = \dim U_1 + \dim U_2.$$

This is our slickest proof yet. All three proofs are really the same – they ended up reducing to Gaussian elimination – but the advantage of this formulation is we never have to mention bases. Not only is it the cleanest proof, it actually makes it easier to calculate. It is certainly helpful for part (ii) of the following exercise.

> **Exercise 1.49.** Let $V = \mathbb{R}^n$ and $U_1, U_2 \leq \mathbb{R}^n$. Let U_1 have a basis $v_1, \ldots, v_r$ and U_2 have a basis $w_1, \ldots, w_s$.
>
> (i) Find a basis for $U_1 + U_2$.
>
> (ii) Find a basis for $U_1 \cap U_2$.

2 Endomorphisms

In this chapter, unless stated otherwise, we take V to be a vector space over a field $\mathbb{F}$, and $\alpha : V \to V$ to be a linear map.

Definition. An *endomorphism* of V is a linear map from V to V. We write $\operatorname{End}(V) = \mathcal{L}(V, V)$ to denote the set of endomorphisms of V:

$$\operatorname{End}(V) = \{\alpha : V \to V, \alpha \text{ linear}\}.$$

The set $\operatorname{End}(V)$ is an algebra: as well as being a vector space over $\mathbb{F}$, we can also multiply elements of it – if $\alpha, \beta \in \operatorname{End}(V)$, then $\alpha\beta \in \operatorname{End}(V)$, i.e. product is *composition* of linear maps.

Recall we have also defined

$$\operatorname{GL}(V) = \big\{\alpha \in \operatorname{End}(V) : \alpha \text{ invertible}\big\}.$$

Fix a basis $b_1, \dots, b_n$ of V and use it as the basis for both the source and target of $\alpha : V \to V$. Then α defines a matrix $A \in \operatorname{Mat}_n(\mathbb{F})$, by $\alpha(b_j) = \sum_i a_{ij}\, b_i$. If $b_1', \dots, b_n'$ is another basis, with change of basis matrix P, then the matrix of α with respect to the new basis is PAP^{-1}.

$$\begin{tikzcd} V \arrow[r, "\alpha"] \arrow[d, "\cong"'] & V \arrow[d, "\cong"] \\ \mathbb{F}^n \arrow[r, "A"'] & \mathbb{F}^n \end{tikzcd}$$

Hence the properties of $\alpha : V \to V$ which don't depend on choice of basis are the properties of the matrix A which are also the properties of all *conjugate* matrices PAP^{-1}.

These are the properties of the set of $\operatorname{GL}(V)$ orbits on $\operatorname{End}(V) = \mathcal{L}(V, V)$, where $\operatorname{GL}(V)$ acts on $\operatorname{End}(V)$, by $(g, \alpha) \mapsto g\alpha g^{-1}$.

In the next two chapters we will determine the set of orbits. This is called the theory of *Jordan normal forms*, and is quite involved.

Contrast this with the properties of a linear map $\alpha : V \to W$ which don't depend on the choice of basis of both V and W; that is, the determination of the $\operatorname{GL}(V) \times \operatorname{GL}(W)$ orbits on $\mathcal{L}(V, W)$. In chapter 1, we've seen that the only property of a linear map which doesn't depend on the choices of a basis is its rank – equivalently that the set of orbits is isomorphic to $\{i \mid 0 \le i \le \min(\dim V, \dim W)\}$.

We begin by defining the determinant, which is a property of an endomorphism which doesn't depend on the choice of a basis.

2.1 Determinants

Definition. We define the map $\det : \operatorname{Mat}_n(\mathbb{F}) \to \mathbb{F}$ by

$$\det A = \sum_{\sigma \in S_n} \epsilon(\sigma)\, a_{1,\sigma(1)} \dots a_{n,\sigma(n)}.$$

Recall that S_n is the group of permutations of $\{1, \dots, n\}$. Any $\sigma \in S_n$ can be written as a product of transpositions (ij).

Then $\epsilon : S_n \to \{\pm 1\}$ is a group homomorphism taking

$$\epsilon(\sigma) = \begin{cases} +1 & \text{if number of permutations is even,} \\ -1 & \text{if number of permutations is odd.} \end{cases}$$

In class, we had a nice interlude here on drawing pictures for symmetric group elements as braids, composition as concatenating pictures of braids, and how $\epsilon(w)$ is the parity of the number of crossings in any picture of w. This was just too unpleasant to type up; sorry!

> **Example 2.1.** We can calculate det by hand for small values of n:
>
> $$\det \begin{pmatrix} a_{11} \end{pmatrix} = a_{11}$$
>
> $$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}$$
>
> $$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{matrix} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ -a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} \end{matrix}$$

The complexity of these expressions grows nastily; when calculating determinants it's usually better to use a different technique rather than directly using the definition.

Lemma 2.2. *If A is upper triangular, that is, if $a_{ij} = 0$ for all $i > j$, then* $\det A = a_{11} \dots a_{nn}$.

Proof. From the definition of determinant:

$$\det A = \sum_{\sigma \in S_n} \epsilon(\sigma)\, a_{1,\sigma(1)} \dots a_{n,\sigma(n)}.$$

If a product contributes, then we must have $\sigma(i) \le i$ for all $i = 1, \dots, n$. Hence $\sigma(1) = 1$, $\sigma(2) = 2$, and so on until $\sigma(n) = n$. Thus the only term that contributes is the identity, $\sigma = \mathrm{id}$, and $\det A = a_{11} \dots a_{nn}$. ☐

Lemma 2.3. $\det A^\mathsf{T} = \det A$, *where* $(A^\mathsf{T})_{ij} = A_{ji}$ *is the* transpose.

Proof. From the definition of determinant, we have

$$\begin{aligned} \det A^\mathsf{T} &= \sum_{\sigma \in S_n} \epsilon(\sigma)\, a_{\sigma(1),1} \dots a_{\sigma(n),n} \\ &= \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^n a_{\sigma(i),i} \end{aligned}$$

Now $\prod_{i=1}^n a_{\sigma(i),i} = \prod_{i=1}^n a_{i,\sigma^{-1}(i)}$, since they contain the same factors but in a different order. We relabel the indices accordingly:

$$= \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{k=1}^n a_{k,\sigma^{-1}(k)}$$

Now since ϵ is a group homomorphism, we have $\epsilon(\sigma \cdot \sigma^{-1}) = \epsilon(\iota) = 1$, and thus $\epsilon(\sigma) = \epsilon(\sigma^{-1})$. We also note that just as σ runs through $\{1, \dots, n\}$, so does σ^{-1}. We thus have

$$= \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{k=1}^n a_{k,\sigma(k)} = \det A.$$ ☐

Writing v_i for the ith column of A, we can consider A as an n-tuple of column vectors, $A = (v_1, \dots, v_n)$. Then $\mathrm{Mat}_n(\mathbb{F}) \cong \mathbb{F}^n \times \cdots \times \mathbb{F}^n$, and det is a function $\mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$.

Proposition 2.4. *The function* $\det : \mathrm{Mat}_n(\mathbb{F}) \to \mathbb{F}$ *is* multilinear*; that is, it is linear in each column of the matrix separately, so:*

$$\begin{aligned} \det(v_1, \dots, \lambda_i\, v_i, \dots, v_n) &= \lambda_i \det(v_1, \dots, v_i, \dots, v_n) \\ \det(v_1, \dots, v_i' + v_i'', \dots, v_n) &= \det(v_1, \dots, v_i', \dots, v_n) + \det(v_1, \dots, v_i'', \dots, v_n). \end{aligned}$$

We can combine this into the single condition

$$\begin{aligned} \det(v_1, \dots, \lambda_i'\, v_i' + \lambda_i''\, v_i'', \dots, v_n) = \lambda_i' \det(v_1, \dots, v_i', \dots, v_n) \\ + \lambda_i'' \det(v_1, \dots, v_i'', \dots, v_n). \end{aligned}$$

Proof. Immediate from the definition: $\det A$ is a sum of terms $a_{1,\sigma(1)}, \dots, a_{n,\sigma(n)}$, each of which contains only one factor from the ith column: $a_{\sigma^{-1}(i),i}$. If this term is $\lambda_i'\, a_{\sigma^{-1}(i),i} + \lambda_i''\, a''_{\sigma^{-1}(i),i}$, then the determinant expands as claims. □

Example 2.5. If we split a matrix along a single column, such as below, then $\det(A) = \det A' + \det A''$.

$$\det \begin{pmatrix} 1 & 7 & 1 \\ 3 & 4 & 1 \\ 2 & 3 & 0 \end{pmatrix} = \det \begin{pmatrix} 1 & 3 & 1 \\ 3 & 2 & 1 \\ 2 & 1 & 0 \end{pmatrix} + \det \begin{pmatrix} 1 & 4 & 1 \\ 3 & 2 & 1 \\ 2 & 2 & 0 \end{pmatrix}$$

Observe how the first and third columns remain the same, and only the second column changes. (Don't get confused: note that $\det(A + B) \neq \det A + \det B$ for general A and B.)

Corollary 2.6. $\det(\lambda A) = \lambda^n A$.

Proof. This follows immediately from the definition, or from applying the result of proposition 2.4 multiple times. □

Proposition 2.7. *If two columns of A are the same, then* $\det A = 0$.

Proof. Suppose v_i and v_j are the same. Let $\tau = (i\,j)$ be the transposition in S_n which swaps i and j. Then $S_n = A_n \coprod A_n\tau$, where $A_n = \ker \epsilon : S_n \to \{\pm 1\}$. We will prove the result by splitting the sum

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

into a sum over these two cosets for A_n, observing that for all $\sigma \in A_n$, $\epsilon(\sigma) = 1$ and $\epsilon(\sigma\tau) = -1$.

Now, for all $\sigma \in A_n$ we have

$$a_{1,\sigma(1)} \dots a_{n,\sigma(n)} = a_{1,\tau\sigma(1)} \dots a_{n,\tau\sigma(n)},$$

as if $\sigma(k) \notin \{i, j\}$, then $\tau\sigma(k) = \sigma(k)$, and if $\sigma(k) = i$, then

$$a_{k,\tau\sigma(k)} = a_{k,\tau(i)} = a_{k,j} = a_{k,i} = a_{k,\sigma(k)},$$

and similarly if $\sigma(k) = j$. Hence

$$\det A = \sum_{\sigma \in A_n} \left(\prod_{i=1}^{n} a_{\sigma(i),i} - \prod_{i=1}^{n} a_{\sigma\tau(i),i} \right) = 0. \qquad \square$$

Proposition 2.8. *If I is the identity matrix, then* $\det I = 1$

Proof. Immediate.

Theorem 2.9

These three properties characterise the function det.

Before proving this, we need some language.

Definition. A function $f : \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$ is a *volume form* on $\mathbb{F}^n$ if

(i) It is *multilinear*, that is, if

$$f(v_1, \ldots, \lambda_i v_i, \ldots, v_n) = \lambda_i f(v_1, \ldots, v_i, \ldots, v_n)$$
$$f(v_1, \ldots, v_i + v_i', \ldots, v_n) = f(v_1, \ldots, v_i, \ldots, v_n) + f(v_1, \ldots, v_i', \ldots, v_n).$$

We saw earlier that we can write this in a single condition:

$$\begin{aligned} f(v_1, \ldots, \lambda_i v_i + \lambda_i' v_i', \ldots, v_n) = \lambda_i f(v_1, \ldots, v_i, \ldots, v_n) \\ + \lambda_i' f(v_1, \ldots, v_i', \ldots, v_n). \end{aligned}$$

(ii) It is *alternating*; that is, whenever $i \neq j$ and $v_i = v_j$, then $f(v_1, \ldots, v_n) = 0$.

Example 2.10. We have seen that $\det : \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$ is a volume form. It is a volume form f with $f(e_1, \ldots, e_n) = 1$ (that is, $\det I = 1$).

Remark. Let's explain the name 'volume form'. Let $\mathbb{F} = \mathbb{R}$, and consider the volume of a rectangular box with a corner at 0 and sides defined by $v_1, \ldots, v_n$ in $\mathbb{R}^n$. The volume of this box is a function of $v_1, \ldots, v_n$ that almost satisfies the properties above. It doesnt quite satisfy linearity, as the volume of a box with sides defined by $-v_1, v_2, \ldots, v_n$ is the same as that of the box with sides defined by $v_1, \ldots, v_n$, but this is the only problem. (Exercise: check that the other properties of a volume form are immediate for voluems of rectangular boxes.) You should think of this as saying that a volume form gives a *signed* version of the volume of a rectangular box (and the actual volume is the absoulute value). In any case, this explains the name. You've also seen this in multi-variable calculus, in the way that the determinant enters into the formula for what happens to integrals when you change coordinates.

Theorem 2.11: Precise form

The set of volume forms forms a vector space of dimension 1. This line is called the *determinant line.*

24 Oct *Proof.* It is immediate from the definition that volume forms are a vector space. Let $e_1, \ldots, e_n$ be a basis of V with $n = \dim V$. Every element of V^n is of the form

$$\left(\sum a_{i1} e_i, \sum a_{i2} e_i, \ldots, \sum a_{in} e_i\right),$$

with $a_{ij} \in \mathbb{F}$ (that is, we have an isomorphism of sets $V^n \xrightarrow{\sim} \mathrm{Mat}_n(\mathbb{F})$). So if f is a volume form, then

$$f\left(\sum_{i_1=1}^{n} a_{i_1 1}\, e_{i_1}, \dots, \sum_{i_n=1}^{n} a_{i_n n}\, e_{i_n}\right) = \sum_{i_1=1}^{n} a_{i_1 1}\, f\left(e_{i_1}, \sum_{i_2=1}^{n} a_{i_2 1}\, e_{i_2}, \dots, \sum_{i_n=1}^{n} a_{i_n n}\, e_{i_n}\right)$$
$$= \cdots = \sum_{1 \le i_1, \dots, i_n \le n} a_{i_1 1} \dots a_{i_n n}\, f(e_{i_1}, \dots, e_{i_n}),$$

by linearity in each variable. But as f is alternating, $f(e_{i_1}, \dots, e_{i_n}) = 0$ unless $i_1, \dots, i_n$ is $1, \dots, n$ in some order; that is,

$$(i_1, \dots, i_n) = \big(\sigma(1), \dots, \sigma(n)\big)$$

for some $\sigma \in S_n$.

Claim. $f(e_{\sigma(1)}, \dots, e_{\sigma(n)}) = \epsilon(\sigma)\, f(e_1, \dots, e_n)$.

Given the claim, we get that the sum above simplifies to

$$\sum_{\sigma \in S_n} a_{\sigma(1),1} \dots a_{\sigma(n),n}\, \epsilon(w)\, f(e_1, \dots, e_n),$$

and so the volume form is determined by $f(e_1, \dots, e_n)$; that is, $\dim(\{\text{vol forms}\}) \le 1$. But $\det : \mathrm{Mat}_n(\mathbb{F}) \to \mathbb{F}$ is a well-defined non-zero volume form, so we must have $\dim(\{\text{vol forms}\}) = 1$.

Note that we have just shown that for any volume form

$$f(v_1, \dots, v_n) = \det(v_1, \dots, v_n)\, f(e_1, \dots, e_n).$$

So to finish our proof, we just have to prove our claim.

Proof of claim. First, for any $v_1, \dots, v_n \in V$, we show that

$$f(\dots, v_i, \dots, v_j, \dots) = -f(\dots, v_j, \dots, v_i, \dots),$$

that is, swapping the ith and jth entries changes the sign. Applying multilinearity is enough to see this:

$$\underbrace{f(\dots, v_i + v_j, \dots, v_i + v_j, \dots)}_{=0 \text{ as alternating}} = \underbrace{f(\dots, v_i, \dots, v_i, \dots)}_{=0 \text{ as alternating}} + \underbrace{f(\dots, v_j, \dots, v_j, \dots)}_{=0 \text{ as alternating}}$$
$$+ f(\dots, v_i, \dots, v_j, \dots) + f(\dots, v_j, \dots, v_i, \dots).$$

Now the claim follows, as an arbitrary permutation can be written as a product of transpositions, and $\epsilon(w) = (-1)^{\#\text{ of transpositions}}$. ☐

Remark. Notice that if $\mathbb{Z}/2 \not\subset \mathbb{F}$ is not a subfield (that is, if $1 + 1 \neq 0$), then for a multilinear form $f(x, y)$ to be alternating, it suffices that $f(x, y) = -f(y, x)$. This is because we have $f(x, x) = -f(x, x)$, so $2f(x, x) = 0$, but $2 \neq 0$ and so 2^{-1} exists, giving $f(x, x) = 0$. If $2 = 0$, then $f(x, y) = -f(y, x)$ for any f and the correct definition of alternating is $f(x, x) = 0$.

If that didn't make too much sense, don't worry: this is included for mathematical interest, and isn't essential to understand anything else in the course.

Remark. If $\sigma \in S_n$, then we can attach to it a matrix $P(\sigma) \in \mathrm{GL}_n$ by

$$P(\sigma)_{ij} = \begin{cases} 1 & \text{if } \sigma^{-1} i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Exercises 2.12. Show that:

(i) $P(w)$ has exactly one non-zero entry in each row and column, and that entry is a 1. Such a matrix is called a *permutation matrix.*
(ii) $P(w)\, e_i = e_j$, hence
(iii) $P : S_n \to \mathrm{GL}_n$ is a group homomorphism;
(iv) $\epsilon(w) = \det P(w)$.

Theorem 2.13

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

Slick proof. Fix $A \in \mathrm{Mat}_n(\mathbb{F})$, and consider $f : \mathrm{Mat}_n(\mathbb{F}) \to \mathbb{F}$ taking $f(B) = \det AB$. We observe that f is a volume form. (Exercise: check this!!) But then

$$f(B) = \det B \cdot f(e_1, \dots, e_n).$$

But by the definition,

$$f(e_1, \dots, e_n) = f(I) = \det A.$$

□

Corollary 2.14. *If* $A \in \mathrm{Mat}_n(\mathbb{F})$ *is invertible, then* $\det A^{-1} = 1/\det A$.

Proof. Since $AA^{-1} = I$, we have

$$\det A \det A^{-1} = \det AA^{-1} = \det I = 1,$$

by the theorem, and rearranging gives the result. □

Corollary 2.15. *If* $P \in \mathrm{GL}_n$, *then*

$$\det(PAP^{-1}) = \det P \det A \det P^{-1} = \det A.$$

Definition. Let $\alpha : V \to V$ be a linear map. Define $\det \alpha \in \mathbb{F}$ as follows: choose *any* basis $b_1, \dots, b_n$ of V, and let A be the matrix of α with respect to the basis. Set $\det \alpha = \det A$, which is well-defined by the corollary.

Remark. Here is a coordinate free definition of $\det \alpha$.

Pick f any volume form for V, $f \neq 0$. Then

$$(x_1, \dots, x_n) \mapsto f(\alpha x_1, \dots, \alpha x_n) = (f\alpha)(x_1, \dots, x_n)$$

is also a volume form. But the space of volume forms is one-dimensional, so there is some $\lambda \in \mathbb{F}$ with $f\alpha = \lambda f$, and we define

$$\lambda = \det \alpha$$

(Though this definition is independent of a basis, we haven't gained much, as we needed to choose a basis to say anything about it.)

Proof 2 of $\det AB = \det A \det B$. We first observe that it's true if B is an elementary column operation; that is, $B = I + \alpha E_{ij}$. Then $\det B = 1$. But

$$\det AB = \det A + \det A',$$

where A' is A except that the ith and jth column of A' are the same as the jth column of A. But then $\det A' = 0$ as two columns are the same.

Next, if B is the permutation matrix $P((i\ j)) = s_{ij}$, that is, the matrix obtained from the identity matrix by swapping the ith and jth columns, then $\det B = -1$, but $A\, s_{ij}$ is A with its ith and jth columns swapped, so $\det AB = \det A \det B$.

Finally, if B is a matrix of zeroes with r ones along the leading diagonal, then if $r = n$, then $B = I$ and $\det B = 1$. If $r < n$, then $\det B = 0$. But then if $r < n$, AB has some columns which are zero, so $\det AB = 0$, and so the theorem is true for these B also.

Now any $B \in \mathrm{Mat}_n(\mathbb{F})$ can be written as a product of these three types of matrices. So if $B = X_1 \cdots X_r$ is a product of these three types of matrices, then

$$\begin{aligned} \det AB &= \det\big((AX_1 \cdots X_{m-1})\, X_m\big) \\ &= \det(AX_1 \cdots X_{m-1}) \det X_m \\ &= \cdots = \det A \det X_1 \cdots \det X_m \\ &= \cdots = \det A \det (X_1 \cdots X_m) \\ &= \det A \det B. \end{aligned}$$

□

Remark. That determinants behave well with respect to row and column operations is also a useful way for humans (as opposed to machines!) to compute determinants.

Proposition 2.16. *Let $A \in \mathrm{Mat}_n(\mathbb{F})$. Then the following are equivalent:*

(i) *A is invertible;*
(ii) $\det A \neq 0$;
(iii) $r(A) = n$.

Proof. (i) $\implies$ (ii). Follows since $\det A^{-1} = 1/\det A$.

(iii) $\implies$ (i). From the rank-nullity theorem, we have

$$r(A) = n \iff \ker \alpha = \{0\} \iff A \text{ invertible}.$$

Finally we must show (ii) $\implies$ (iii). If $r(A) < n$, then $\ker \alpha \neq \{0\}$, so there is some $\Lambda = (\lambda_1, \ldots, \lambda_n)^\mathsf{T} \in \mathbb{F}^n$ such that $A\Lambda = 0$, and $\lambda_k \neq 0$ for some k. Now put

$$B = \begin{pmatrix} 1 & & & \lambda_1 & & \\ & 1 & & \lambda_2 & & \\ & & \ddots & \vdots & & \\ & & & \lambda_k & & \\ & & & \vdots & \ddots & \\ & & & \lambda_n & & 1 \end{pmatrix}$$

Then $\det B = \lambda_k \neq 0$, but AB is a matrix whose kth column is 0, so $\det AB = 0$; that is, $\det A = 0$, since $\lambda_k \neq 0$. □

This is a horrible and unenlightening proof that $\det A \neq 0$ implies the existence of A^{-1}. A good proof would write the matrix coefficients of A^{-1} in terms of $(\det A)^{-1}$ and the matrix coefficients of A. We will now do this, after some showing some further properties of the determinant.

We can compute $\det A$ by expanding along any column or row.

Definition. Let A^{ij} be the matrix obtained from A by deleting the ith row and the jth column.

Theorem 2.17

(i) Expand along the jth column:

$$\begin{aligned}\det A &= (-1)^{j+1}\, a_{1j} \det A^{1j} + (-1)^{j+2}\, a_{2j} \det A^{2j} + \cdots + (-1)^{j+n}\, a_{nj} \det A^{nj} \\ &= (-1)^{j+1} \left[a_{1j} \det A^{1j} - a_{2j} \det A^{2j} + a_{3j} \det A^{3j} - \cdots + (-1)^{n+1}\, a_{nj} \det A^{nj}\right]\end{aligned}$$

(the thing to observe here is that the signs alternate!)

(ii) Expanding along the ith row:

$$\det A = \sum_{j=1}^{n} (-1)^{i+j}\, a_{ij} \det A^{ij}.$$

The proof is boring book keeping.

Proof. Put in the definition of A^{ij} as a sum over $w \in S_{n-1}$, and expand. We can tidy this up slightly, by writing it as follows: write $A = (v_1 \cdots v_n)$, so $v_j = \sum_i a_{ij}\, e_i$. Then

$$\begin{aligned}\det A = \det(v_1, \ldots, v_n) &= \sum_{i=1}^{n} a_{ij} \det(v_1, \ldots, v_{j-1}, e_i, v_{j+1}, \ldots, v_n) \\ &= \sum_{i=1}^{n} (-1)^{j-1} \det\left(e_i, v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n\right).\end{aligned}$$

as $\epsilon(1\,2\,\ldots\,j) = (-1)^{j-1}$ (in class we drew a picture of this symmetric group element, and observed it had $j-1$ crossings.) Now $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^{\mathsf{T}}$, so we pick up $(-1)^{i-1}$ as the sign of the permutation $(1\,2\,\ldots\,i)$ that rotates the 1st through ith rows, and so get

$$\sum_i (-1)^{i+j-2} \det \begin{pmatrix} 1 & * \\ 0 & A^{ij} \end{pmatrix} = \sum_i (-1)^{i+j} \det A^{ij}. \qquad \square$$

Definition. For $A \in \mathrm{Mat}_n(\mathbb{F})$, the *adjugate matrix*, denoted by $\operatorname{adj} A$, is the matrix with

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

Example 2.18.

$$\operatorname{adj}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \qquad \operatorname{adj}\begin{pmatrix} 1 & 1 & 2 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 4 & -2 & -3 \\ 1 & 0 & -1 \\ -2 & 1 & 2 \end{pmatrix}.$$

Theorem 2.19: Cramer's rule

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

Proof. We have

$$\big((\operatorname{adj} A)\, A\big)_{jk} = \sum_{i=1}^{n} (\operatorname{adj} A)_{ji}\, a_{ik} = \sum_{i=1}^{n} (-1)^{i+j} \det A^{ij} a_{ik}$$

Now, if we have a diagonal entry $j = k$ then this is exactly the formula for $\det A$ in (i) above. If $j \neq k$, then by the same formula, this is $\det A'$, where A' is obtained from A by replacing its jth column with the kth column of A; that is A' has the j and kth columns the *same*, so $\det A' = 0$, and so this term is zero. □

Corollary 2.20. $A^{-1} = \dfrac{1}{\det A} \operatorname{adj} A$ *if* $\det A \neq 0$.

The proof of Cramer's rule only involved multiplying and adding, and the fact that they satisfy the usual distributive rules and that multiplication and addition are commutative. A set in which you can do this is called a *commutative ring*. Examples include the integers $\mathbb{Z}$, or polynomials $\mathbb{F}[x]$.

So we've shown that if $A \in \mathrm{Mat}_n(R)$, where R is any commutative ring, then there exists an inverse $A^{-1} \in \mathrm{Mat}_n(R)$ if and only if $\det A$ has an inverse in R: $(\det A)^{-1} \in R$. For example, an integer matrix $A \in \mathrm{Mat}_n(\mathbb{Z})$ has an inverse with integer coefficients if and only if $\det A = \pm 1$.

Moreover, the matrix coefficients of $\operatorname{adj} A$ are polynomials in the matrix coefficients of A, so the matrix coefficients of A^{-1} are polynomials in the matrix coefficients of A and the inverse of the function $\det A$ (which is itself a polynomial function of the matrix coefficients of A).

That's very nice to know.

3 Jordan normal form

In this chapter, unless stated otherwise, we take V to be a finite dimensional vector space over a field $\mathbb{F}$, and $\alpha : V \to V$ to be a linear map. We're going to look at what matrices look like up to conjugacy; that is, what the map α looks like, given the freedom to choose a basis for V.

3.1 Eigenvectors and eigenvalues

Definition. A non-zero vector $v \in V$ is an *eigenvector* for $\alpha : V \to V$ if $\alpha(v) = \lambda v$, for some $\lambda \in \mathbb{F}$. Then λ is called the *eigenvalue* associated with v, and the set

$$V_\lambda = \left\{ v \in V : \alpha(v) = \lambda v \right\}$$

is called the *eigenspace of* λ *for* α, which is a subspace of V.

We observe that if $I : V \to V$ is the identity map, then

$$V_\lambda = \ker(\lambda I - \alpha : V \to V).$$

So if $v \in V_\lambda$, then v is a non-zero vector if and only if $\ker(\lambda I - \alpha) \neq \{0\}$, which is equivalent saying that $\lambda I - \alpha$ is *not* invertible. Thus

$$\det(\lambda I - \alpha) = 0,$$

by the results of the previous chapter.

Definition. If $b_1, \dots, b_n$ is a basis of V, and $A \in \mathrm{Mat}_n(\mathbb{F})$ is a matrix of α, then

$$\mathrm{ch}_\alpha(x) = \det(xI - \alpha) = \det(xI - A)$$

is the *characteristic polynomial* of α.

The following properties follow from the definition:

(i) The general form is

$$\mathrm{ch}_\alpha(x) = \mathrm{ch}_A(x) = \det \begin{pmatrix} x - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & x - a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ -a_{n1} & \cdots & \cdots & x - a_{nn} \end{pmatrix} \in F[x].$$

Observe that $\mathrm{ch}_A(x) \in F[x]$ is a polynomial in x, equal to x^n plus terms of smaller degree, and the coefficients are polynomials in the matrix coefficients a_{ij}.

For example, if $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ then

$$\begin{aligned} \mathrm{ch}_A(x) &= x^2 - x\,(a_{11} + a_{22}) + (a_{11}a_{22} - a_{12}a_{21}) \\ &= x^2 - x.\,\mathrm{tr}\,A + \det A. \end{aligned}$$

(ii) Conjugate matrices have the same characteristic polynomials. Explicitly:

$$\begin{aligned}\operatorname{ch}_{PAP^{-1}}(x) &= \det\left(xI - PAP^{-1}\right) \\ &= \det\left(P\left(xI - A\right)P^{-1}\right) \\ &= \det(xI - A) \\ &= \operatorname{ch}_A(x).\end{aligned}$$

(iii) For $\lambda \in \mathbb{F}$, $\operatorname{ch}_\alpha(\lambda) = 0$ if and only if $V_\lambda = \left\{v \in V : \alpha(v) = \lambda v\right\} \neq \{0\}$; that is, if λ is an eigenvalue of α. This gives us a way to find the eigenvalues of a linear map.

Example 3.1. If A is upper-triangular with a_{ii} in the ith diagonal entry, then

$$\operatorname{ch}_A(x) = (x - a_{11}) \cdots (x - a_{nn}).$$

It follows that the diagonal terms of an upper triangular matrix are its eigenvalues.

Definition. We say that $\operatorname{ch}_A(x)$ *factors* if it factors into linear factors; that is, if

$$\operatorname{ch}_A(x) = \textstyle\prod_{i=1}^{r} (x - \lambda_i)^{n_i},$$

for some $n_i \in \mathbb{N}$, $\lambda_i \in \mathbb{F}$ and $\lambda_i \neq \lambda_j$ for $i \neq j$.

Examples 3.2. If we take $\mathbb{F} = \mathbb{C}$, then the fundamental theorem of algebra says that every polynomial $f \in \mathbb{C}[x]$ factors into linear terms.

In $\mathbb{R}$, consider the rotation matrix

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

then we have characteristic polynomial

$$\operatorname{ch}_A(x) = x^2 - 2x\cos\theta + 1,$$

which factors if and only if $A = \pm I$ and $\theta = 0$ or π.

Definition. If $\mathbb{F}$ is any field, then there is some bigger field $\overline{\mathbb{F}}$, the *algebraic closure of* $\mathbb{F}$, such that $\mathbb{F} \subseteq \overline{\mathbb{F}}$ and every polynomial in $\overline{F}[x]$ factors into linear factors. This is proved next year in the *Galois theory* course.

Theorem 3.3

If A is an $n \times n$ matrix over $\mathbb{F}$, then $\operatorname{ch}_A(x)$ factors if and only if A is conjugate to an upper triangular matrix.

In particular, this means that if $\mathbb{F} = \overline{\mathbb{F}}$, such as $\mathbb{F} = \mathbb{C}$, then every matrix is conjugate to an upper triangular matrix.

We can give a coordinate free formulation of the theorem: if $\alpha : V \to V$ is a linear map, then $\operatorname{ch}_\alpha(x)$ factors if and only if there is some basis $b_1, \ldots, b_n$ of V such that the matrix of α with respect to the basis is upper triangular.

Proof. ($\Leftarrow$) If A is upper triangular, then $\mathrm{ch}_A(x) = \prod(x - a_{ii})$, so done.

($\Rightarrow$) Otherwise, set $V = \mathbb{F}^n$, and $\alpha(x) = Ax$. We induct on $\dim V$. If $\dim V = n = 1$, then we have nothing to prove.

As $\mathrm{ch}_\alpha(x)$ factors, there is some $\lambda \in \mathbb{F}$ such that $\mathrm{ch}_\alpha(\lambda) = 0$, so there is some $\lambda \in \mathbb{F}$ with a non-zero eigenvector b_1. Extend this to a basis $b_1, \dots, b_n$ of V.

Now conjugate A by the change of basis matrix. (In other words, write the linear map α, $x \mapsto Ax$ with respect to this basis b_i rather than the standard basis e_i).

We get a new matrix

$$\widetilde{A} = \begin{pmatrix} \lambda & * \\ 0 & A' \end{pmatrix},$$

and it has characteristic polynomial

$$\mathrm{ch}_{\widetilde{A}}(x) = (x - \lambda)\,\mathrm{ch}_{A'}(x).$$

So $\mathrm{ch}_\alpha(x)$ factors implies that $\mathrm{ch}_{A'}(x)$ factors.

Now, by induction, there is some matrix $P \in \mathrm{GL}_{n-1}(\mathbb{F})$ such that $PA'P^{-1}$ is upper triangular. But now

$$\begin{pmatrix} 1 & \\ & P \end{pmatrix} \widetilde{A} \begin{pmatrix} 1 & \\ & P^{-1} \end{pmatrix} = \begin{pmatrix} \lambda & \\ & PA'P^{-1} \end{pmatrix},$$

proving the theorem. □

Aside: what is the meaning of the matrix A'? We can ask this question more generally. Let $\alpha : V \to V$ be linear, and $W \leq V$ a subspace. Choose a basis $b_1, \dots, b_r$ of W, extend it to be a basis of V (add $b_{r+1}, \dots, b_n$).

Then $\alpha(W) \subseteq W$ if and only if the matrix of α with respect to this basis looks like

$$\begin{pmatrix} X & Z \\ 0 & Y \end{pmatrix},$$

where X is $r \times r$ and Y is $(n-r) \times (n-r)$, and it is clear that $\alpha\big|_W : W \to W$ has matrix X, with respect to a basis $b_1, \dots, b_r$ of W.

Then our question is: What is the meaning of the matrix Y?

The answer requires a new concept, the *quotient vector space.*

> **Exercise 3.4.** Consider V as an abelian group, and consider the coset group $V/W = \{v + W : v \in V\}$. Show that this is a vector space, that $b_{r+1}+W, \dots, b_n+W$ is a basis for it, and $\alpha : V \to V$ induces a linear map $\widetilde{\alpha} : V/W \to V/W$ by $\widetilde{\alpha}(v + W) = \alpha(v) + W$ (you need to check this is well-defined and linear), and that with respect to this basis, Y is the matrix of $\widetilde{\alpha}$.

Remark. Let $V = W' \oplus W''$; that is, $W' \cap W'' = \{0\}$, $W' + W'' = V$, and suppose that $\alpha(W') \subseteq W'$ and $\alpha(W'') \subseteq W''$. We write this as $\alpha = \alpha' \oplus \alpha''$, where $\alpha' : W' \to W'$, $\alpha'' : W'' \to W''$ are the restrictions of α.

In this special case the matrix of α looks even more special then the above for any basis $b_1, \dots, b_r$ of W' and $b_{r+1}, \dots, b_n$ of W'' – we have $Z = 0$ also.

Definition. The *trace* of a matrix $A = (a_{ij})$, denoted $\mathrm{tr}(A)$, is given by

$$\mathrm{tr}(A) = \sum_i a_{ii},$$

Lemma 3.5. $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

Proof. $\mathrm{tr}(AB) = \sum_i (AB)_{ii} = \sum_{i,j} a_{ij}b_{ji} = \sum_j (BA)_{jj} = \mathrm{tr}(BA)$. □

Corollary 3.6. $\mathrm{tr}(PAP^{-1}) = \mathrm{tr}(P^{-1}PA) = \mathrm{tr}(A)$.

So we define, if $\alpha : V \to V$ is linear, $\mathrm{tr}(\alpha) = \mathrm{tr}(A)$, where A is a matrix of α with respect to some basis $b_1, \dots, b_n$, and this doesn't depend on the choice of a basis.

Proposition 3.7. *If* $\mathrm{ch}_\alpha(x)$ *factors as* $(x - \lambda_1) \cdots (x - \lambda_n)$ *(repetition allowed), then*

(i) $\mathrm{tr}\,\alpha = \sum_i \lambda_i$;
(ii) $\det \alpha = \prod_i \lambda_i$.

Proof. As ch_α factors, there is some basis $b_1, \dots, b_n$ of V such that the matrix of A is upper triangular, the diagonal entries are $\lambda_1, \dots, \lambda_n$, and we're done. □

Remark. This is true whatever $\mathbb{F}$ is. Embed $\mathbb{F} \subseteq \overline{\mathbb{F}}$ (for example, $\mathbb{R} \subseteq \mathbb{C}$), and ch_A factors as $(x - \lambda_1) \cdots (x - \lambda_n)$. Now $\lambda_1, \dots, \lambda_n \in \overline{\mathbb{F}}$, not necessarily in $\mathbb{F}$.

Regard $A \in \mathrm{Mat}_n(\overline{\mathbb{F}})$, which doesn't change $\mathrm{tr}\,A$ or $\det A$, and we get the same result. Note that $\sum_i \lambda_i$ and $\prod_i \lambda_i$ are in $\mathbb{F}$ even though $\lambda_i \notin \mathbb{F}$.

Example 3.8. Take $A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$. Eigenvalues are $e^{i\theta}$, $e^{-i\theta}$, so

$$\mathrm{tr}\,A = e^{i\theta} + e^{-i\theta} = 2\cos\theta \qquad \det A = e^{i\theta} \cdot e^{-i\theta} = 1.$$

Note that

$$\begin{aligned} \mathrm{ch}_A(x) &= (x - \lambda_1) \cdots (x - \lambda_n) \\ &= x^n - \left(\sum_i \lambda_i \right) x^{n-1} + \left(\sum_{i<j} \lambda_i \lambda_j \right) x^{n-2} - \cdots + (-1)^n (\lambda_1 \dots \lambda_n) \\ &= x^n - \mathrm{tr}(A) x^{n-1} + e_2 x^{n-2} - \cdots + (-1)^{n-1} e_{n-1} + (-1)^n \det A, \end{aligned}$$

where the coefficients $e_1 = \mathrm{tr}\,A$, $e_2, \dots, e_{n-1}$, $e_n = \det A$ are functions of $\lambda_1, \dots, \lambda_n$ called *elementary symmetric functions* (which we see more of in *Galois theory*).

Each of these e_i depend only on the conjugacy class of A, as

$$\mathrm{ch}_{PAP^{-1}}(x) = \mathrm{ch}_A(x).$$

Note that A and B conjugate implies $\mathrm{ch}_A(x) = \mathrm{ch}_B(x)$, but the converse is false. Consider, for example, the matrices

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

and $\mathrm{ch}_A(x) = \mathrm{ch}_B(x) = x^2$, but A and B are not conjugate.

For an upper triangular matrix, the diagonal entries are the eigenvalues. What is the meaning of the upper triangular coefficients?

This example shows there is *some* information in the upper triangular entries of an upper-triangular matrix, but the question is how much? We would like to always diagonalise A, but this example shows that it isn't always possible. Let's understand when it is possible. 31 Oct

Proposition 3.9. *If $v_1, \dots, v_k$ are eigenvectors with eigenvalues $\lambda_1, \dots, \lambda_k$, and $\lambda_i \neq \lambda_j$ if $i \neq j$, then $v_1, \dots, v_k$ are linearly independent.*

Proof 1. Induct on k. This is clearly true when $k = 1$. Now if the result is false, then there are $a_i \in \mathbb{F}$ s.t. $\sum_{i=1}^k a_i\, v_i = 0$, with some $a_i \neq 0$, and without loss of generality, $a_1 \neq 0$. (In fact, all $a_i \neq 0$, as if not, we have a relation of linearly dependence among $(k-1)$ eigenvectors, contradicting our inductive assumption.)

Apply α to $\sum_{i=1}^k a_i\, v_i = 0$, to get

$$\sum_{i=1}^k \lambda_i\, a_i\, v_i = 0.$$

Now multiply $\sum_{i=1}^k a_i\, v_i$ by λ_1, and we get

$$\sum_{i=1}^k \lambda_1\, a_i\, v_i = 0.$$

Subtract these two, and we get

$$\sum_{i=2}^k \underbrace{(\lambda_i - \lambda_1)}_{\neq 0}\, a_i\, v_i = 0,$$

a relation of linear dependence among $v_2, \dots, v_k$, so $a_i = 0$ for all i, by induction. □

Proof 2. Suppose $\sum a_i\, v_i = 0$. Apply α, we get $\sum \lambda_i\, a_i\, v_i = 0$; apply α^2, we get $\sum \lambda_i^2\, a_i\, v_i = 0$, and so on, so $\sum_{i=1}^k \lambda_i^r\, a_i\, v_i = 0$ for all $r \geq 0$. In particular,

$$\begin{pmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_k \\ \vdots & & \vdots \\ \lambda_1^{k-1} & \cdots & \lambda_k^{k-1} \end{pmatrix} \begin{pmatrix} a_1\, v_1 \\ \vdots \\ \vdots \\ a_k\, v_k \end{pmatrix} = 0.$$

Lemma 3.10 (The Vandermonde determinant)**.** *The determinant of the above matrix is $\prod_{i<j} \left(\lambda_j - \lambda_i\right)$.*

Proof. Exercise! □

By the lemma, if $\lambda_i \neq \lambda_j$, this matrix is invertible, and so $\left(a_1\, v_1, \dots, a_k\, v_k\right)^\mathsf{T} = 0$; that is, $a_i = 0$ for all i. □

Note these two proofs are the same: the first version of the proof was surreptitiously showing that the Vandermonde determinant was non-zero. It looks like the first proof is easier to understand, but the second proof makes clear what's actually going on.

Definition. A map α is *diagonalisable* if there is some basis for V such that the matrix of $\alpha : V \to V$ is diagonal.

Corollary 3.11. *The map α is diagonalisable if and only if* $\mathrm{ch}_\alpha(x)$ *factors into* $\prod_{i=1}^{r}(x-\lambda_i)^{n_i}$, *and* $\dim V_{\lambda_i} = n_i$ *for all i.*

Proof. ($\Rightarrow$) α is diagonalisable means that in some basis it is diagonal, with n_i copies of λ_i in the diagonal entries, hence the characteristic polynomial is as claimed.

($\Leftarrow$) $\sum_i V_{\lambda_i}$ is a direct sum, by the proposition, so

$$\dim\left(\textstyle\sum_i V_{\lambda_i}\right) = \sum \dim V_{\lambda_i} = \sum n_i,$$

and by our assumption, $n = \dim V$. Now in any basis which is the union of basis for the V_{λ_i}, the matrix of α is diagonal. $\square$

Corollary 3.12. *If A is conjugate to an upper triangular matrix with λ_i as the diagonal entries, and the λ_i are distinct, then A is conjugate to the diagonal matrix with λ_i in the entries.*

Example 3.13. $\begin{pmatrix} 1 & 7 \\ 0 & 2 \end{pmatrix}$ is conjugate to $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$.

The upper triangular entries "contain no information". That is, they are an artefact of the choice of basis.

Remark. If $\mathbb{F} = \mathbb{C}$, then the diagonalisable A are dense in $\mathrm{Mat}_n(\mathbb{C}) = \mathbb{C}^{n^2}$ (exercise). In general, if $\mathbb{F} = \overline{\mathbb{F}}$, then diagonalisable A are dense in $\mathrm{Mat}_n(\mathbb{F}) = \mathbb{F}^{n^2}$, in the sense of algebraic geometry.

Exercise 3.14. If $A = \begin{pmatrix} \lambda_1 & \cdots & a_n \\ & \ddots & \vdots \\ 0 & & \lambda_n \end{pmatrix}$, then $Ae_i = \lambda e_i + \sum_{j<i} a_{ji}\, e_j$.

Show that if $\lambda_1, \ldots, \lambda_n$ are distinct, then you can "correct" each e_i to an eigenvalue v_i just by adding smaller terms; that is, there are $p_{ji} \in \mathbb{F}$ such that

$$v_i = e_i + \sum_{j<i} p_{ji} e_j \text{ has } Av_i = \lambda_i v_i,$$

which gives yet another proof of our proposition.

3.2 Cayley-Hamilton theorem

Let $\alpha : V \to V$ be a linear map, and V a finite dimensional vector space over $\mathbb{F}$.

Theorem 3.15: Cayley-Hamilton theorem

Every square matrix over a commutative ring (such as $\mathbb{R}$ or $\mathbb{C}$) satisfies $\mathrm{ch}_A(A) = 0$.

Example 3.16. $A = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$ and $\mathrm{ch}_A(x) = x^2 - 4x + 1$, so $\mathrm{ch}_A(A) = A^2 - 4A + I$. Then

$$A^2 = \begin{pmatrix} 7 & 12 \\ 4 & 7 \end{pmatrix},$$

which does equal $4A - I$.

Remark. We have

$$\mathrm{ch}_A(x) = \det(xI - A) = \det \begin{pmatrix} x - a_{11} & \cdots & -a_{1n} \\ \vdots & \ddots & \vdots \\ -a_{n1} & \cdots & x - a_{nn} \end{pmatrix} = x^n - e_1 x^{n-1} + \cdots \pm e_n,$$

so we don't get a proof by saying $\mathrm{ch}_\alpha(\alpha) = \det(\alpha - \alpha) = 0$. This just doesn't make sense. However, you can make it make sense, and our second proof will do this.

Proof 1. If $A = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$, then $\mathrm{ch}_A(x) = (x - \lambda_1) \cdots (x - \lambda_n)$.

Now if A were in fact diagonal, then

$$\mathrm{ch}_A(A) = (A - \lambda_1 I) \cdots (A - \lambda_n I) = \begin{pmatrix} 0 & & & 0 \\ & * & & \\ & & * & \\ 0 & & & * \end{pmatrix} \begin{pmatrix} * & & & 0 \\ & 0 & & \\ & & * & \\ 0 & & & * \end{pmatrix} \cdots \begin{pmatrix} * & & & 0 \\ & * & & \\ & & * & \\ 0 & & & 0 \end{pmatrix} = 0$$

But even when A is upper triangular, this is zero.

Example 3.17. For example, $\begin{pmatrix} 0 & * & * \\ & * & * \\ & & * \end{pmatrix} \begin{pmatrix} * & * & * \\ & 0 & * \\ & & * \end{pmatrix} \begin{pmatrix} * & * & * \\ & * & * \\ & & 0 \end{pmatrix}$ is still zero.

Here is a nice way of writing this:

Let $W_0 = \{0\}$, $W_i = \langle e_1, \ldots, e_i \rangle \leq V$. Then if A is upper trianglar, then $AW_i \subseteq W_i$, and even $(A - \lambda_i I)\, W_i \subseteq W_{i-1}$. So $(A - \lambda_n I)\, W_n \subseteq W_{n-1}$, and so

$$(A - \lambda_{n-1} I)\,(A - \lambda_n I)\, W_n \subseteq (A - \lambda_{n-1} I)\, W_{n-1} \subseteq W_{n-2},$$

and so on, until

$$\prod_{i=1}^{n} (A - \lambda_i I)\, W_n \subseteq W_0 = \{0\};$$

that is, $\mathrm{ch}_A(A) = 0$.

Now if $\mathbb{F} = \overline{\mathbb{F}}$, then we can choose a basis for V such that $\alpha : V \to V$ has an upper-triangular matrix with respect to this basis, and hence the above shows $\mathrm{ch}_A(x) = 0$; that is, $\mathrm{ch}_A(A) = 0$ for all $A \in \mathrm{Mat}_n(\overline{F})$.

Now, if $\mathbb{F} \subseteq \overline{F}$, then as Cayley-Hamilton is true for all $A \in \mathrm{Mat}_n(\overline{\mathbb{F}})$, then it is certainly still true for $A \in \mathrm{Mat}_n(\mathbb{F}) \subseteq \mathrm{Mat}_n(\overline{\mathbb{F}})$. □

Definition. The *generalised eigenspace with eigenvalue* λ is given by

$$V^\lambda = \left\{ v \in V : (\alpha - \lambda I)^{\dim V}(v) = 0 \right\} = \ker (\lambda I - \alpha)^{\dim V} : V \to V.$$

Note that $V_\lambda \subseteq V^\lambda$.

Example 3.18. Let $A = \begin{pmatrix} \lambda & & * \\ & \ddots & \\ 0 & & \lambda \end{pmatrix}$.

Then $(\lambda I - A) e_i$ has stuff involving $e_1, \dots, e_{i-1}$, so $(\lambda I - A)^{\dim V} e_i = 0$ for all i, as in our proof of Cayley-Hamilton (or indeed, by Cayley-Hamilton).

Further, if $\mu \neq \lambda$, then

$$\mu I - A = \begin{pmatrix} \mu - \lambda & & * \\ & \ddots & \\ 0 & & \mu - \lambda \end{pmatrix},$$

and so

$$(\mu I - A)^n = \begin{pmatrix} (\mu - \lambda)^n & & * \\ & \ddots & \\ 0 & & (\mu - \lambda)^n \end{pmatrix}$$

has non-zero diagonal terms, so zero kernel. Thus in this case, $V^\lambda = V$, $V^\mu = 0$ if $\mu \neq \lambda$, and in general $V^\mu = 0$ if $\mathrm{ch}_\alpha(\mu) \neq 0$, that is, $\ker (A - \mu I)^N = \{0\}$ for all $N \geq 0$.

2 Nov

Theorem 3.19

If $\mathrm{ch}_A(x) = \prod_{i=1}^r (x - \lambda_i)^{n_i}$, with the λ_i distinct, then

$$V \cong \bigoplus_{i=1}^r V^{\lambda_i},$$

and $\dim V^{\lambda_i} = n_i$. In other word, choose any basis of V which is the union of the bases of the V^{λ_i}. Then the matrix of α is block diagonal. Moreover, we can choose the basis of each V^λ so that each diagonal block is upper triangular, with only one eigenvalue— λ–on its diagonals.

We say "different eigenvalues don't interact".

Remark. If $n_1 = n_2 = \cdots = n_r = 1$ (and so $r = n$), then this is our previous theorem that matrices with distinct eigenvalues are diagonalisable.

Proof. Consider

$$h_{\lambda_i}(x) = \prod_{j \neq i} (x - \lambda_j)^{n_j} = \frac{\mathrm{ch}_\alpha(x)}{(x - \lambda_i)^{n_i}}.$$

Then define

$$W^{\lambda_i} = \mathrm{Im}\left(h_{\lambda_i}(A) : V \to V\right) \leq V.$$

Now Cayley-Hamilton implies that

$$\underbrace{(A - \lambda_i I)^{n_i}\, h_{\lambda_i}(A)}_{= \mathrm{ch}_A(A)} = 0,$$

that is,

$$W^{\lambda_i} \subseteq \ker\,(A - \lambda_i I)^{n_i} \subseteq \ker\,(A - \lambda_i I)^n = V^{\lambda_i}.$$

We want to show that

(i) $\sum_i W^{\lambda_i} = V$;
(ii) This sum is direct.

Now, the h_{λ_i} are coprime polynomials, so Euclid's algorithm implies that there are polynomials $f_i \in F[x]$ such that

$$\sum_{i=1}^{r} f_i\, h_{\lambda_i} = 1,$$

and so

$$\sum_{i=1}^{r} h_{\lambda_i}(A)\, f_i(A) = I \subset \mathrm{End}(V).$$

Now, if $v \in V$, then this gives

$$v = \sum_{i=1}^{r} \underbrace{h_{\lambda_i}(A)\, f_i(A)\, v}_{\in W^{\lambda_i}},$$

that is, $\sum_{i=1}^{r} W^{\lambda_i} = V$. This is (i).

To see the sum is direct: if $0 = \sum_{i=1}^{r} w_i$, $w_i \in W^{\lambda_i}$, then we want to show that each $w_i = 0$. But $h_{\lambda_j}(w_i) = 0$, $i \neq j$ as $w_i \in \ker\,(A - \lambda_i I)^{n_i}$, so (i) gives

$$w_i = \sum_{i=1}^{r} h_{\lambda_i}(A)\, f_j(A)\, w_i = f_i(A)\, h_{\lambda_i}(A)\,(w_i),$$

so apply $f_i(A)\, h_{\lambda_i}(A)$ to $\sum_{i=1}^{r} w_i = 0$ and get $w_i = 0$. $\square$

Define

$$\pi_i = f_i(A)\, h_{\lambda_i}(A) = h_{\lambda_i}(A)\, f_i(A).$$

We showed that $\pi_i : V \to V$ has

$$\mathrm{Im}\,\pi_i = W^{\lambda_i} \subseteq V^{\lambda_i} \text{ and } \pi_i | W^{\lambda_i} = \text{identity, and so } \pi_i^2 = \pi_i,$$

that is, π_i is the projection to W^{λ_i}. Compare with $h_{\lambda_i}(A) = h_i$, which has $h_i(V) = W^{\lambda_i} \subseteq V^{\lambda_i}$, $h_i \mid V^{\lambda_i}$ an isomorphism, but not the identity; that is, $f_i \mid V^{\lambda_i} = h_i^{-1} \mid V^{\lambda_i}$.

This tells us to understand what matrices look like up to conjugacy, it is enough to understand matrices with a single eigenvalue λ, and by subtracting λI from our matrix we may as well assume that eigenvalue is zero.

Before we continue investigating this, we digress and give another proof of Cayley-Hamilton.

Proof 2 of Cayley-Hamilton. Let $\varphi : V \to V$ be linear, V finite dimensional over $\mathbb{F}$. Pick a basis $e_1, \dots, e_n$ of V, so $\varphi(e_i) = \sum_j a_{ji}\, e_j$, and we have the matrix $A = (a_{ij})$. Consider

$$\varphi I - A^{\mathsf{T}} = \begin{pmatrix} \varphi - a_{11} & \cdots & -a_{n1} \\ \vdots & \ddots & \vdots \\ -a_{1n} & \cdots & \varphi - a_{nn} \end{pmatrix} \in \mathrm{Mat}_n(\mathrm{End}(V)),$$

where $a_{ij} \in \mathbb{F} \hookrightarrow \mathrm{End}(V)$ by regarding an element λ as the operation of scalar multiplication $V \to V$, $v \mapsto \lambda v$. The elements of $\mathrm{Mat}_n(\mathrm{End}(V))$ act on V^n by the usual formulas. So

$$\left(\varphi I - A^{\mathsf{T}}\right) \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

by the definition of A.

The problem is, it isn't clear how to define $\det : \mathrm{Mat}_n(\mathrm{End}(V)) \to \mathrm{End}(V)$, as the matrix coefficients (that is, elements of $\mathrm{End}(V)$) do not commute in general. But the matrix elements of the above matrix do commute, so this shouldn't be a problem.

To make it not a problem, consider $\varphi I - A^{\mathsf{T}} \in \mathrm{Mat}_n(F[\varphi])$; that is, $F[\varphi]$ are polynomials in the symbol φ. This is a commutative ring and now det behaves as always:

(i) $\det(\varphi I - A^{\mathsf{T}}) = \mathrm{ch}_A(\varphi) \in F[\varphi]$ (by definition);
(ii) $\mathrm{adj}(\varphi I - A^{\mathsf{T}}) \cdot (\varphi I - A^{\mathsf{T}}) = \det(\varphi I - A^{\mathsf{T}}) \cdot I \in \mathrm{Mat}_n(F[\varphi])$, as we've shown.

This is true for any $B \in \mathrm{Mat}_n(R)$, where R is a commutative ring. Here $R = F[\varphi]$, $B = \varphi I - A^{\mathsf{T}}$.

Make $F[\varphi]$ act on V, by $\sum_i a_i\, \varphi^i : v \mapsto \sum_i a_i\, \varphi^i(v)$, so

$$(\varphi I - A^{\mathsf{T}}) \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Thus

$$0 = \mathrm{adj}(\varphi I - A^{\mathsf{T}}) \underbrace{(\varphi I - A^{\mathsf{T}}) \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}}_{=0} = \det(\varphi I - A^{\mathsf{T}}) \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} \mathrm{ch}_A(\varphi)\, e_1 \\ \vdots \\ \mathrm{ch}_A(\varphi)\, e_n \end{pmatrix}.$$

So this says that $\mathrm{ch}_A(A)\, e_i = \mathrm{ch}_A(\varphi)\, e_i = 0$ for all i, so $\mathrm{ch}_A(A) : V \to V$ is the zero map, as $e_1, \dots, e_n$ is a basis of V; that is, $\mathrm{ch}_A(A) = 0$. $\square$

This correct proof is as close to the nonsense tautological "proof" (just set x equal to A) as you can hope for. You will meet it again several times in later life, where it is called *Nakayama's lemma*.

3.3 Combinatorics of nilpotent matrices

5 Nov

Definition. If $\varphi : V \to V$ can be written in block diagonal form; that is, if there are some $W', W'' \leq V$ such that

$$\varphi(W') \subseteq W' \qquad \varphi(W'') \subset W'' \qquad V = W' \oplus W''$$

then we say that φ is *decomposable* and write

$$\varphi = \varphi' \oplus \varphi'' \qquad \varphi' = \varphi\big|_{W'} : W' \to W' \qquad \varphi'' = \varphi\big|_{W''} : W'' \to W''$$

We say that φ is the *direct sum* of φ' and φ''.

Otherwise, we say that φ is *decomposable.*

Examples 3.20.

(i) $\varphi = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = (0 : \mathbb{F} \to \mathbb{F}) \oplus (0 : \mathbb{F} \to \mathbb{F})$

(ii) $\varphi = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} : \mathbb{F}^2 \to \mathbb{F}^2$ is indecomposable, because there is a unique φ-stable line $\langle e_1 \rangle$.

(iii) If $\varphi : V \to V$, then $\mathrm{ch}_\varphi(x) = \prod_{i=1}^{r} (x - \lambda_i)^{n_i}$, for $\lambda_i \neq \lambda_j$ if $i \neq j$.

Then $V = \bigoplus_{i=1}^{r} V^{\lambda_i}$ decomposes φ into pieces $\varphi_{\lambda_i} = \varphi\big|_{V^{\lambda_i}} : V^{\lambda_i} \to V^{\lambda_i}$ such that each φ_λ has only one eigenvalue, λ.

This decomposition is precisely the amount of information in $\mathrm{ch}_\varphi(x)$. So to further understand what matrices are up to conjugacy, we will need new information.

Observe that φ_λ is decomposable if and only if $\varphi_\lambda - \lambda I$ is, and $\varphi_\lambda - \lambda I$ has zero as its only eigenvalue.

Definition. The map φ is *nilpotent* if $\varphi^{\dim V} = 0$ if and only if $\ker \varphi^{\dim V} = V$ if and only if $V^0 = V$ if and only if $\mathrm{ch}_\varphi(x) = x^{\dim V}$. (The only eigenvalue is zero.)

Theorem 3.21

Let φ be nilpotent. Then φ is indecomposable if and only if there is a basis $v_1, \ldots, v_n$ such that

$$\varphi(v_i) = \begin{cases} 0 & \text{if } i = 1, \\ v_{i-1} & \text{if } i > 1, \end{cases}$$

that is, if the matrix of φ is

$$J_n = \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}$$

This is the *Jordan block of size n with eigenvalue 0.*

Definition. The *Jordan block of size n, eigenvalue* λ, is given by

$$J_n(\lambda) = \lambda I + J_n.$$

Theorem 3.22: Jordan normal form

Every matrix is conjugate to a direct sum of Jordan blocks. Morever, these are unique up to rearranging their order.

Proof. Observe that theorem 3.22 $\implies$ theorem 3.21, if we show that J_n is indecomposable (and theorem 3.21 $\implies$ theorem 3.22, existence).

[Proof of Theorem 3.21, $\Leftarrow$] Put $W_i = \langle v_1, \dots, v_i \rangle$.

Then W_i is the *only* subspace W of $\dim i$ such that $\varphi(W) \subseteq W$, and W_{n-i} is *not* a complement to it, as $W_i \cap W_{n-i} = W_{\min(i,n-i)}$. $\square$

Proof of Theorem 3.22, uniqueness. Suppose $\alpha : V \to V$, α nilpotent and $\alpha = \bigoplus_{i=1}^r J_{k_i}$. Rearrange their order so that $k_i \geq k_j$ for $i \geq j$, and group them together, so $\bigoplus_{i=1}^r m_i J_i$.

There are m_i blocks of size i, and

$$m_i = \#\left\{k_a \mid k_a = i\right\}. \tag{$*$}$$

Example 3.23. If $(k_1, k_2, \dots) = (3, 3, 2, 1, 1, 1)$, then $n = 11$, and $m_1 = 3, m_2 = 1, m_3 = 2$, and $m_a = 0$ for $a > 3$. (It is customary to omit the zero entries when listing these numbers).

Definition. Let $\mathcal{P}_n = \{(k_1, k_2, \dots, k_n) \in \mathbb{N}^n \mid k_1 \geq k_2 \geq \dots \geq k_n \geq 0, \sum k_i = n\}$ be the set of *partitions of* n. This is isomorphic to the set $\{m : \mathbb{N} \to \mathbb{N} \mid \sum_i im(i) = n\}$ as above.

We represent $\mathbf{k} \in \mathcal{P}_n$ by a picture, with a row of length k_j for each j (equivalently, with m_i rows of length i). For example, the above partition $(3, 3, 2, 1, 1, 1)$ has picture

$$\mathbf{k} = \begin{matrix} X & X & X \\ X & X & X \\ X & X & \\ X & & \\ X & & \\ X, & & \end{matrix}$$

Now define $\mathbf{k}^\mathsf{T}$, if $\mathbf{k} \in \mathcal{P}_n$, the *dual partition*, to be the partition attached to the transposed diagram.

In the above example $\mathbf{k}^\mathsf{T} = (6, 3, 2)$.

It is clear that $\mathbf{k}$ determines $\mathbf{k}^\mathsf{T}$. In formulas:

$$\mathbf{k}^\mathsf{T} = (m_1 + m_2 + m_3 + \dots + m_n, m_2 + m_3 + \dots + m_n, \dots, m_n)$$

Now, let $\alpha : V \to V$, and $\alpha = \bigoplus_{i=1}^{r} J_{k_i} = \bigoplus_{i=1}^{r} m_i J_i$ as before. Observe that

$$\begin{aligned}
\dim\ker\alpha &= \#\text{ of Jordan blocks} = r = \sum_{i=1}^{n} m_i = (\mathbf{k}^{\mathsf{T}})_1 \\
\dim\ker\alpha^2 &= \#\text{ of Jordan blocks of sizes 1 and 2} = \sum_{i=1}^{n} m_i + \sum_{i=2}^{n} m_i = (\mathbf{k}^{\mathsf{T}})_1 + (\mathbf{k}^{\mathsf{T}})_2 \\
&\vdots \\
\dim\ker\alpha^n &= \#\text{ of Jordan blocks of sizes } 1,\dots,n = \sum_{k=1}^{n}\sum_{i=k}^{n} m_i = \sum_{i=1}^{n} (\mathbf{k}^{\mathsf{T}})_i.
\end{aligned}$$

That is, $\dim\ker\alpha, \dots, \dim\ker\alpha^n$ determine the dual partition to the partition of n into Jordan blocks, and hence determine it.

It follows that the decomposition $\alpha = \bigoplus_{i=1}^{n} m_i J_i$ is unique. □

Remark. This is a practical way to compute JNF of a matrix A. First computer $\mathrm{ch}_A(x) = \prod_{i=1}^{r} (x - \lambda_i)^{n_i}$, then compute eigenvalues with $\ker(A - \lambda_i I), \ker(A - \lambda_i I)^2, \dots, \ker(A - \lambda_i I)^n$.

Corollary 3.24. *The number of nilpotent conjugacy classes is equal to the size of $\mathcal{P}_n$.*

Exercises 3.25.

(i) List all the partitions of $\{1, 2, 3, 4, 5\}$; show there are 7 of them.

(ii) Show that the size of $\mathcal{P}_n$ is the coefficient of x^n in

$$\begin{aligned}
\prod_{i\geq 1} \frac{1}{1 - x^i} &= (1 + x + x^2 + x^3 + \cdots)(1 + x^2 + x^4 + x^6 + \cdots)(1 + x^3 + x^6 + x^9 + \cdots)\cdots \\
&= \prod_{k\geq 1} \sum_{i=0}^{\infty} x^{ki}
\end{aligned}$$

7 Nov

Theorem 3.26: Jordan Normal Form

Every matrix is conjugate to a direct sum of Jordan blocks

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

Proof. It is enough to show this when $\varphi : V \to V$ has a single generalised eigenspace with eigenvalue λ, and now replacing φ by $\varphi - \lambda I$, we can assume that φ is nilpotent.

Induct on $n = \dim V$. The case $n = 1$ is clear.

Consider $V' = \operatorname{Im}\varphi = \varphi(V)$. Then $V' \neq V$, as φ is nilpotent, and $\varphi(V') \subseteq \varphi(V) = V'$, and $\varphi \mid V' : V' \to V'$ is obviously nilpotent, so induction gives the existence of a basis

$$\underbrace{e_1, \dots, e_{k_1}}_{J_{k_1}} \oplus \underbrace{e_{k_1+1}, \dots, e_{k_1+k_2}}_{J_{k_2}} \oplus \dots \oplus \underbrace{\dots, e_{k_1+\dots+k_r}}_{J_{k_r}}$$

such that $\varphi \mid V'$ is in JNF with respect to this basis.

Because $V' = \operatorname{Im}\varphi$, it must be that the tail end of these strings is in $\operatorname{Im}\varphi$; that is, there exist $b_1, \ldots, b_r \in V \backslash V'$ such that $\varphi(b_i) = e_{k_1+\cdots+k_i}$, as $e_{k_1+\cdots+k_i} \notin \varphi(V')$. Notice these are linearly independent, as if $\sum \lambda_i b_i = 0$, then

$$\sum \lambda_i \varphi(b_i) = \sum \lambda_i e_{k_1+\cdots+k_i} = 0.$$

But $e_{k_1}, \ldots, e_{k_1+\ldots+k_r}$ are linearly independent, hence $\lambda_1 = \cdots = \lambda_r = 0$. Even better: $\{e_j, b_i \mid j \leq k_1 + \ldots + k_r, 1 \leq i \leq r\}$ are linearly independent. (Proof: exercise.)

Finally, extend $\underbrace{e_1, e_{k_1+1}, \ldots, e_{k_1+\ldots+k_{r-1}+1}}_{\text{basis of } \ker\varphi \cap \operatorname{Im}\varphi}$ to a basis of $\ker\varphi$, by adding basis vectors.

Denote these by $q_1, \ldots, q_s$. **Exercise.** Show $\{e_j, b_i, q_k\}$ are linearly independent.

Now, the rank-nullity theorem shows that $\dim \operatorname{Im}\varphi + \dim\ker\varphi = \dim V$. But $\dim \operatorname{Im}\varphi$ is the number of the e_i, that is $k_1 + \cdots + k_r$, and $\dim\ker\varphi$ is the number of Jordan blocks, which is $r + s$, (r is the number of blocks of size greater than one, s the number of size one), which is the number of b_i plus the number of q_k.

So this shows that e_j, b_i, q_k are a basis of V, and hence with respect to this basis,

$$\varphi = J_{k_1+1} \oplus \cdots \oplus J_{k_r+1} \oplus \underbrace{J_1 \oplus \cdots \oplus J_1}_{s \text{ times}} \qquad \square$$

3.4 Applications of JNF

> **Definition.** Suppose $\alpha : V \to V$. The *minimum polynomial of* α is a monic polynomial $p(x)$ of smallest degree such that $p(\alpha) = 0$.

Lemma 3.27. *If $q(x) \in F[x]$ and $q(\alpha) = 0$, then $p \mid q$.*

Proof. Write $q = pa + r$, with $a, r \in F[x]$ and $\deg r < \deg p$. Then

$$0 = q(\alpha) = p(\alpha)\, a(\alpha) + r(\alpha) = r(\alpha) \implies r(\alpha) = 0,$$

which contradicts $\deg p$ as minimal unless $r = 0$. $\square$

As $\operatorname{ch}_\alpha(\alpha) = 0$, $p(x) \mid \operatorname{ch}_\alpha(x)$, and in particular, it exists. (And by our lemma, is unique.) Here is a cheap proof that the minimum polynomial exists, which doesn't use Cayley Hamilton.

Proof. $I, \alpha, \alpha^2, \ldots, \alpha^{n^2}$ are n^2+1 linear functions in $\operatorname{End} V$, a vector space of dimension n^2. Hence there must be a relation of linear dependence, $\sum_0^{n^2} a_i \alpha^i = 0$, so $q(x) = \sum_0^{n^2} a_i x^i$ is a polynomial with $q(\alpha) = 0$. $\square$

Now, lets use JNF to determine the minimial polynomial.

> **Exercise 3.28.** Let $A \in Mat_n(\mathbb{F})$, $\operatorname{ch}_A(x) = (x - \lambda_1)^{n_1} \cdots (x - \lambda_r)^{n_r}$. Suppose that the maximal size of a Jordan block with eigenvalue λ_i is k_i. (So $k_i \leq n_i$ for all i). Show that the minimum polynomial of A is $(x - \lambda_1)^{k_1} \cdots (x - \lambda_r)^{k_r}$.

So the minimum polynomial forgets most of the structure of the Jordan normal form.

Another application of JNF is we can use it to compute powers B^n of a matrix B for any $n \geq 0$. First observe that $(PAP^{-1})^n = PA^nP^{-1}$ Now write $B = PAP^{-1}$ with A

in Jordan normal form. So to finish we must compute what the powers of elements in JNF look like. But

$$J_n = \begin{pmatrix} 0 & 1 & & & & 0 \\ & 0 & 1 & & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & 0 & 1 \\ 0 & & & & & 0 \end{pmatrix}, \qquad J_n^2 = \begin{pmatrix} 0 & 0 & 1 & & & 0 \\ & & 0 & 1 & & \\ & & & \ddots & & \\ & & & & 0 & 1 \\ & & & & & 0 \\ 0 & & & & & 0 \end{pmatrix}, \quad \dots, \qquad J_n^{n-1} = \begin{pmatrix} 0 & & & 0 & 1 \\ & & & & 0 \\ & & & & \\ & & & & \\ 0 & & & & 0 \end{pmatrix}$$

and

$$(\lambda I + J_n)^a = \sum_{k \geq 0} \binom{a}{k} \lambda^{a-k} J_n^k.$$

Now assume $\mathbb{F} = \mathbb{C}$.

Definition. $\exp A = \sum_{n \geq 0} \frac{A^n}{n!}$, $A \in \mathrm{Mat}_n(\mathbb{C})$.

This is an infinite sum, and we must show it converges. This means that each matrix coefficient converges. This is very easy, but we omit here for lack of time.

Example 3.29. For a diagonal matrix:

$$\exp \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = \begin{pmatrix} e^{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n} \end{pmatrix}$$

and convergence is usual convergence of exp.

Exercises 3.30.

(i) If $AB = BA$, then $\exp(A + B) = \exp A \exp B$.

(ii) Hence $\exp(J_n + \lambda I) = e^\lambda \exp(J_n)$

(iii) $P \cdot \exp(A) \cdot P^{-1} = \exp(PAP^{-1})$

So now you know how to compute $\exp(A)$, for $A \in \mathrm{Mat}_n(\mathbb{C})$.

We can use this to solve linear ODEs with constant coefficients:

Consider the linear ODE

$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = A\mathbf{y},$$

for $A \in \mathrm{Mat}_n(\mathbb{C})$, $\mathbf{y} = (y_1(t), y_2(t), \dots, y_n(t))^T$, $y_i(t) \in C^\infty(\mathbb{C})$.

Example 3.31. Consider

$$\frac{\mathrm{d}^n z}{\mathrm{d}t^n} + c_{n-1} \frac{\mathrm{d}^{n-1} z}{\mathrm{d}t^{n-1}} + \cdots + c_0 z = 0, \qquad (**)$$

This is a particular case of the above, where A is the matrix

$$\begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ \dots & & & & \\ & & & 0 & 1 \\ -c_0 & -c_1 & \dots & & -c_{n-1} \end{pmatrix}.$$

To see this, consider what $A\mathbf{y} = \mathbf{y}'$ means. Set $z = y_1$, then $y_2 = y_1' = z'$, $y_3 = y_2' = z''$, ..., $y_n = y_{n-1}' = \frac{\mathrm{d}^{n-1}z}{\mathrm{d}t^{n-1}}$ and $(**)$ is the last equation.

There is a unique solution of $\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}t} = A\mathbf{y}$ with fixed initial conditions $y(0)$, by a theorem of analysis. On the other hand:

Exercise 3.32. $\exp(At)\,y(0)$ is a solution, that is

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\exp(At)\,y(0)\big) = A\exp(At)\,y(0)$$

Hence it is the unique solution with value $y(0)$.

Compute this when $A = \lambda I + J_n$ is a Jordan block of size n.

4 Duals

This chapter really belongs after chapter 1 – it's just definitions and intepretations of row reduction.

9 Nov

Definition. Let V be a vector space over a field $\mathbb{F}$. Then

$$V^* = \mathcal{L}(V, \mathbb{F}) = \{\text{linear functions } V \to \mathbb{F}\}$$

is the *dual space* of V.

Examples 4.1.

(i) Let $V = \mathbb{R}^3$. Then $(x, y, z) \mapsto x - y$ is in V^*.
(ii) If $V = C([0,1]) = \langle \text{continuous functions } [0,1] \to \mathbb{R} \rangle$, then $f \mapsto \int_0^1 f(t)\,\mathrm{d}t$ is in $C([0,1])^*$.

Definition. Let V be a finite dimensional vector space over $\mathbb{F}$, and $v_1, \ldots, v_n$ be a basis of V. Then define $v_i^* \in V^*$ by

$$v_i^*(v_j) = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j, \end{cases}$$

and extend linearly. That is, $v_i^* \left(\sum_j \lambda_j \, v_j \right) = \lambda_i$.

Lemma 4.2. *The set $v_1^*, \ldots, v_n^*$ is a basis for V^*, called the* basis dual to *or* dual basis for *$v_1, \ldots, v_n$. In particular,* $\dim V^* = \dim V$.

Proof. Linear independence: if $\sum \lambda_i \, v_i^* = 0$, then $0 = \left(\sum \lambda_i \, v_i^* \right)(v_j) = \lambda_j$, so $\lambda_j = 0$ for all j. Span: if $\varphi \in V^*$, then we claim

$$\varphi = \sum_{j=1}^{n} \varphi(v_j) \cdot v_j^*.$$

As φ is linear, it is enough to check that the right hand side applied to v_k is $\varphi(v_k)$. But

$$\textstyle\sum_j \varphi(v_j) \, v_j^*(v_k) = \sum_j \varphi(v_j) \, \delta_{jk} = \varphi(v_k).$$ □

Remarks.

(i) We know in general that $\dim \mathcal{L}(V, W) = \dim V \dim W$.

(ii) If V is finite dimensional, then this shows that $V \cong V^*$, as any two vector spaces of dimension n are isomorphic. But they are not *canonically* isomorphic (there is no natural choice of isomorphism).

If the vector space V has more structure (for example, a group G acts upon it), then V and V^* are not usually isomorphic in a way that respects this structure.

(iii) If $V = F[x]$, then $V^* \xrightarrow{\sim} \mathbb{F}^{\mathbb{N}}$ by the isomorphism $\theta \in V^* \mapsto (\theta(1), \theta(x), \theta(x^2), \ldots)$, and conversely, if $\lambda_i \in \mathbb{F}$, $i = 0, 1, 2, \ldots$ is any sequence of elements of $\mathbb{F}$, we get an element of V^* by sending $\sum a_i \, x^i \mapsto \sum a_i \, \lambda_i$ (notice this is a finite sum).

Thus V and V^* are not isomorphic, since $\dim V$ is countable, but $\dim \mathbb{F}^{\mathbb{N}}$ is uncountable.

Definition. Let V and W be vector space over $\mathbb{F}$, and α a linear map $V \to W$, $\alpha \in \mathcal{L}(V, W)$. Then we define $\alpha^* : W^* \to V^* \in \mathcal{L}(W^*, V^*)$, by setting $\alpha^*(\theta) = \theta\alpha : V \to \mathbb{F}$.

(Note: α linear, θ linear implies $\theta\alpha$ linear, and so $\alpha^*\theta \in V^*$ as claimed, if $\theta \in W^*$.)

Lemma 4.3. *Let V, W be finite dimensional vector spaces, with*

$v_1, \dots, v_n$ *a basis of V, and $w_1, \dots, w_m$ a basis for W;*
$v_1^*, \dots, v_n^*$ *the dual basis of V^*, and $w_1^*, \dots, w_m^*$ the dual basis for W^*;*

If α is a linear map $V \to W$, and A is the matrix of α with respect to v_i, w_j, then A^T is the matrix of $\alpha^* : W^* \to V^*$ with respect to w_j^*, v_i^*.

Proof. Write $\alpha^*(w_i^*) = \sum_{j=1}^n c_{ji}\, v_j^*$, so c_{ij} is a matrix of α^*. Apply this to v_k:

$$\begin{aligned}\text{LHS} &= \left(\alpha^*(w_i^*)\right)(v_k) = w_i^*(\alpha(v_k)) = w_i^*\left(\textstyle\sum_\ell a_{\ell k}\, w_\ell\right) = a_{ik}\\ \text{RHS} &= c_{ki},\end{aligned}$$

that is, $c_{ji} = a_{ij}$ for all i, j. □

This was the promised interpretation of A^T.

Corollary 4.4.

(i) $(\alpha\beta)^* = \beta^*\alpha^*$;
(ii) $(\alpha + \beta)^* = \alpha^* + \beta^*$;
(iii) $\det \alpha^* = \det \alpha$

Proof. (i) and (ii) are immediate from the definition, or use the result $(AB)^\mathsf{T} = B^\mathsf{T} A^\mathsf{T}$. (iii) we proved in the section on determinants where we showed that $\det A^\mathsf{T} = \det A$. □

Now observe that $\left(A^\mathsf{T}\right)^\mathsf{T} = A$. What does this mean?

Proposition 4.5.

(i) Consider the map $V \to V^{**} = (V^*)^*$ taking $v \mapsto \hat{v}$, where $\hat{v}(\theta) = \theta(v)$ if $\theta \in V^*$. Then $\hat{\hat{v}} \in V^{**}$, and the map $V \mapsto V^{**}$ is linear and injective.

(ii) Hence if V is a finite dimensional vector space over $\mathbb{F}$, then this map is an isomorphism, so $V \xrightarrow{\sim} V^{**}$ canonically.

Proof.

(i) We first show $\hat{\hat{v}} \in V^{**}$, that is $\hat{\hat{v}} : V^* \to \mathbb{F}$, is linear:

$$\begin{aligned}\hat{\hat{v}}(a_1\theta_1 + a_2\theta_2) &= (a_1\theta_1 + a_2\theta_2)(v) = a_1\, \theta_1(v) + a_2\, \theta_2(v)\\ &= a_1\, \hat{\hat{v}}(\theta_1) + a_2\, \hat{\hat{v}}(\theta_2).\end{aligned}$$

Next, the map $V \to V^{**}$ is linear. This is because

$$\begin{aligned}\widehat{\widehat{(\lambda_1 v_1 + \lambda_2 v_2)}}(\theta) &= \theta\left(\lambda_1 v_1 + \lambda_2 v_2\right)\\ &= \lambda_1\, \theta(v_1) + \lambda_2\, \theta(v_2)\\ &= \left(\lambda_1 \hat{\hat{v}}_1 + \lambda_2 \hat{\hat{v}}_2\right)(\theta).\end{aligned}$$

Finally, if $v \neq 0$, then there exists a linear function $\theta : V \to \mathbb{F}$ such that $\theta(v) \neq 0$.

(Proof: extend v to a basis, and then define θ on this basis. We've only proved that this is okay when V is finite dimensional, but it's always okay.)

Thus $\hat{v}(\theta) \neq 0$, so $\hat{v} \neq 0$, and $V \to V^{**}$ is injective.

(ii) Immediate. □

Definition.

(i) If $U \leq V$, then define

$$U^\circ = \big\{\theta \in V^* \mid \theta(U) = 0\big\} = \big\{\theta \in V^* \mid \theta(u) = 0 \ \forall u \in U\big\} \leq V^*.$$

This is the *annihilator* of U, a subspace of V^*, often denoted $U^\perp$.

(ii) If $W \leq V^*$, then define

$$^\circ W = \big\{v \in V \mid \varphi(v) = 0 \ \forall \varphi \in W\big\} \leq V.$$

This is often denoted $^\perp W$.

Example 4.6. If $V = \mathbb{R}^3$, $U = \big\langle (1,2,1) \big\rangle$, then

$$U^\circ = \left\{ \sum_{i=1}^{3} a_i\, e_i^* \in V^* \mid a_1 + 2a_2 + a_3 = 0 \right\} = \left\langle \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix} \right\rangle.$$

Remark. If V is finite dimensional, and $W \leq V^*$, then under the canonical isomorphism $V \to V^{**}$, we have $^\circ W \mapsto W^\circ$, where $^\circ W \leq V$ and $(W^\circ) \leq (V^*)^*$. Proof is an exercise.

Lemma 4.7. *Let V be a finite dimensional vector space with $U \leq V$. Then*

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

Proof. Consider the restriction map $\mathrm{Res} : V^* \to U^*$ taking $\varphi \mapsto \varphi|_U$. (Note that $\mathrm{Res} = \iota^*$, where $\iota : U \hookrightarrow V$ is the inclusion.)

Then $\ker \mathrm{Res} = U^\circ$, by definition, and Res is surjective (why?).

So the rank-nullity theorem implies the result, as $\dim V^* = \dim V$. □

Proposition 4.8. *Let V, W be a finite dimensional vector space over $\mathbb{F}$, with $\alpha \in \mathcal{L}(V, W)$. Then*

(i) $\ker(\alpha^* : W^* \to V^*) = (\operatorname{Im} \alpha)^\circ$ *($\leq W^*$);*

(ii) $\operatorname{rank}(\alpha^*) = \operatorname{rank}(\alpha)$*; that is,* $\operatorname{rank} A^\mathsf{T} = \operatorname{rank} A$*, as promised;*

(iii) $\operatorname{Im} \alpha^* = (\ker \alpha)^\circ$.

Proof.

(i) Let $\theta \in W^*$. Then $\theta \in \ker \alpha^* \iff \theta\alpha = 0 \iff \theta\alpha(v) = 0 \ \forall v \in V \iff \theta \in (\operatorname{Im} \alpha)^\circ$.

(ii) By rank-nullity, we have

$$\begin{aligned} \operatorname{rank} \alpha^* &= \dim W - \dim \ker \alpha^* \\ &= \dim W - \dim(\operatorname{Im} \alpha)^\circ, \text{by (i)}, \\ &= \dim \operatorname{Im} \alpha, \ \text{by the previous lemma}, \\ &= \operatorname{rank} \alpha, \ \text{by definition}. \end{aligned}$$

(iii) Let $\varphi \in \operatorname{Im}\alpha^*$, and then $\varphi = \theta \circ \alpha$ for some $\theta \in W^*$. Now, let $v \in \ker\alpha$. Then $\varphi(v) = \theta\alpha(v) = 0$, so $\varphi \in (\ker\alpha)^\circ$; that is, $\operatorname{Im}\alpha^* \subseteq (\ker\alpha)^\circ$.
But by (ii),

$$\begin{aligned}\dim \operatorname{Im}\alpha^* = \operatorname{rank}(\alpha^*) &= \operatorname{rank}\alpha = \dim V - \dim\ker\alpha \\ &= \dim(\ker\alpha)^\circ\end{aligned}$$

by the previous lemma; that is, they both have the same dimension, so they are equal. □

Lemma 4.9. *Let $U_1, U_2 \leq V$, and V finite dimensional. Then*

(i) $U_1^{\circ\circ} \xrightarrow{\sim} {}^\circ(U_1^\circ) \xrightarrow{\sim} U_1$ *under the isomorphism* $V \xrightarrow{\sim} V^{**}$.
(ii) $(U_1 + U_2)^\circ = U_1^\circ \cap U_2^\circ$.
(iii) $(U_1 \cap U_2)^\circ = U_1^\circ + U_2^\circ$.

Proof. Exercise! □

5 Bilinear forms

12 Nov

Definition. Let V be a vector space over $\mathbb{F}$. A *bilinear form* on V is a multilinear form $V \times V \to \mathbb{F}$; that is, $\psi : V \times V \to \mathbb{F}$ such that

$$\psi(v, a_1 w_1 + a_2 w_2) = a_1\, \psi(v, w_1) + a_2\, \psi(v, w_2)$$
$$\psi(a_1 v_1 + a_2 v_2, w) = a_1\, \psi(v_1, w) + a_2\, \psi(v_2, w).$$

Examples 5.1.

(i) $V = \mathbb{F}^n$, $\psi\big((x_1, \dots, x_n), (y_1, \dots, y_n)\big) = \sum_{i=1}^n x_i\, y_i$, which is the dot product of $\mathbb{F} = \mathbb{R}^n$.

(ii) $V = \mathbb{F}^n$, $A \in \mathrm{Mat}_n(\mathbb{F})$. Define $\psi(v, w) = v^{\mathsf{T}} A w$. This is bilinear.

(i) is the special case when $A = I$. Another special case is $A = 0$, which is also a bilinear form.

(iii) Take $V = C([0, 1])$, the set of continuous functions on $[0, 1]$. Then

$$(f, g) \mapsto \int_0^1 f(t)\, g(t)\, \mathrm{d}t$$

is bilinear.

Definition. The set of bilinear forms of V is denoted

$$\mathrm{Bil}(V) = \{\psi : V \times V \to \mathbb{F} \text{ bilinear}\}$$

Exercise 5.2. If $g \in \mathrm{GL}(V)$, $\psi \in \mathrm{Bil}(V)$, then $g\psi : (v, w) \mapsto \psi(g^{-1}v, g^{-1}w)$ is a bilinear form. Show this defines a group action of $\mathrm{GL}(V)$ on $\mathrm{Bil}(V)$. *In particular, show that $h(g\psi) = (hg)\psi$, and you'll see why the inverse is in the definition of $g\psi$.*

Definition. We say that $\psi, \varphi \in \mathrm{Bil}(V)$ are *isomorphic* if there is some $g \in \mathrm{GL}(V)$ such that $\varphi = g\psi$; that is, if they are in the same orbit.

Q: What are the orbits of $\mathrm{GL}(V)$ on $\mathrm{Bil}(V)$; that is, what is the isomorphism classes of bilinear forms?

Compare with:

- $\mathcal{L}(V, W)/\,\mathrm{GL}(V) \times \mathrm{GL}(W) \leftrightarrow \big\{i \in \mathbb{N} \mid 0 \le i \le \min(\dim V, \dim W)\big\}$ with $\varphi \mapsto \operatorname{rank} \varphi$. Here $(g, h) \circ \varphi = h\varphi g^{-1}$.
- $\mathcal{L}(V, V)/\,\mathrm{GL}(V) \leftrightarrow$ JNF. Here $g \circ \varphi = g\varphi g^{-1}$ and we require $\mathbb{F}$ algebraically closed.
- $\mathrm{Bil}(V)/\,\mathrm{GL}(V) \leftrightarrow ??$, with $(g \circ \psi)(v, w) = \psi(g^{-1}v, g^{-1}w)$.

First, let's express this in matrix form. Let $v_1, \dots, v_n$ be a basis for V, where V is a finite dimensional vector space over $\mathbb{F}$, and $\psi \in \mathrm{Bil}(V)$. Then

$$\psi\left(\textstyle\sum_i x_i\, v_i, \sum_j y_j\, v_j\right) = \textstyle\sum_{i,j} x_i\, y_j\, \psi(v_i, v_j)$$

So if we define a matrix A by $A = (a_{ij})$, $a_{ij} = \psi(v_i, v_j)$, then we say that A is the matrix of the bilinear form with respect to the basis $v_1, \dots, v_n$.

In other words, the isomorphism $V \underset{\theta}{\xrightarrow{\sim}} \mathbb{F}^n$ induces an isomorphism $\operatorname{Bil}(V) \xrightarrow{\sim} \operatorname{Mat}_n \mathbb{F}$, $\psi \mapsto a_{ij} = \psi(v_i, v_j)$.

Now, let $v_1', \dots, v_n'$ be another basis, with $v_j' = \sum_i p_{ij}\, v_i$. Then

$$\psi(v_a', v_b') = \psi\left(\textstyle\sum_i p_{ia}\, v_i, \sum_j p_{jb}\, v_j\right) = \textstyle\sum_{i,j} p_{ia}\, \psi(v_i, v_j)\, p_{jb} = \left(P^\mathsf{T} A P\right)_{ab}$$

So if P is the matrix of the linear map $g^{-1} : V \to V$, then the matrix of $g\psi = \psi(g^{-1}(\cdot), g^{-1}(\cdot))$ is $P^\mathsf{T} A P$.

So the concrete version of our question "what are the orbits $\operatorname{Bil}(V)/\operatorname{GL}(V)$" is "what are the orbits of GL_n on $\operatorname{Mat}_n(\mathbb{F})$ for this action?"

Definition. Suppose Q acts on A by QAQ^T. We say that A and B are *congruent* if $B = QAQ^\mathsf{T}$ for some $Q \in \operatorname{GL}_n$.

We want to understand when two matrices are congruent.

Recall that if $P, Q \in \operatorname{GL}_n$, then $\operatorname{rank}(PAQ) = \operatorname{rank}(A)$. Hence taking $Q = P^\mathsf{T}$, we get $\operatorname{rank}(PAP^\mathsf{T}) = \operatorname{rank} A$, and so the following definition makes sense:

Definition. If $\psi \in \operatorname{Bil}(V)$, then the *rank* of ψ, denoted $\operatorname{rank} \psi$ or $\operatorname{rk} \psi$ is the rank of the matrix of ψ with respect to some (and hence any) basis of V.

We will see later how to give a basis independent definition of the rank.

Definition. A form $\psi \in \operatorname{Bil}(V)$ is

- *symmetric* if $\psi(v, w) = \psi(w, v)$ for all $v, w \in V$. In terms of the matrix A of ψ, this is requiring $A^\mathsf{T} = A$.
- *anti-symmetric* if $\psi(v, v) = 0$ for all $v \in V$, which implies $\psi(v, w) = -\psi(w, v)$ for all $v, w \in V$. In terms of the matrix, $A^\mathsf{T} = -A$.

From now on, we assume that $\operatorname{char} \mathbb{F} \neq 2$, so $1 + 1 = 2 \neq 0$ and $1/2$ exists.

Given ψ, put

$$\begin{aligned} \psi^+(v, w) &= \tfrac{1}{2}\left[\psi(v, w) + \psi(w, v)\right] \\ \psi^-(v, w) &= \tfrac{1}{2}\left[\psi(v, w) - \psi(w, v)\right], \end{aligned}$$

which splits a form into symmetric and anti-symmetric components, and $\psi = \psi^+ + \psi^-$.

Observe that if ψ is symmetric or anti-symmetric, then so is $g\psi = \psi(g(\cdot), g(\cdot))$, or in matrix form, A is symmetric or anti-symmetric if and only if PAP^T is, since $(PAP^\mathsf{T})^\mathsf{T} = PA^\mathsf{T}P^\mathsf{T}$.

So to understand $\operatorname{Bil}(V)/\operatorname{GL}(V)$, we will first understand the simpler question of classifying symmetric and anti-symmetric forms. Set

$$\operatorname{Bil}^\varepsilon(V) = \left\{\psi \in \operatorname{Bil}(V) \mid \psi(v, w) = \varepsilon\, \psi(v, w)\ \forall v, w \in V\right\} \qquad \varepsilon = \pm 1.$$

So $\operatorname{Bil}^+(V)$ is the symmetric forms, and Bil^- is the antisymmetric forms.

So our simpler question is to ask, "What is $\operatorname{Bil}^\varepsilon(V)/\operatorname{GL}(V)$?"

Hard exercise: Once you've finished revising the course, go and classify $\operatorname{Bil}(V)/GL(V)$.

5.1 Symmetric forms

Let V be a finite dimensional vector space over $\mathbb{F}$ and $\operatorname{char}\mathbb{F} \neq 2$. If $\psi \in \operatorname{Bil}^+(V)$ is a symmetric form, then define $Q : V \to \mathbb{F}$ as

$$Q(v) = Q_\psi(v) = \psi(v, v).$$

We have

$$\begin{aligned}
Q(u+v) &= \psi(u+v, u+v) \\
&= \psi(u,u) + \psi(v,v) + \psi(u,v) + \psi(v,u) \\
&= Q(u) + Q(v) + \psi(u,v) + \psi(v,u) \\
Q(\lambda u) &= \psi(\lambda u, \lambda u) \\
&= \lambda^2\, \psi(u,u) \\
&= \lambda^2\, Q(u).
\end{aligned}$$

Definition. A *quadratic form* on V is a function $Q : V \to \mathbb{F}$ such that

(i) $Q(\lambda v) = \lambda^2 Q(v)$;
(ii) Set $\psi_Q(u,v) = \frac{1}{2}\big[Q(u+v) - Q(u) - Q(v)\big]$; then $\psi_Q : V \times V \to \mathbb{F}$ is bilinear.

Lemma 5.3. *The map* $\operatorname{Bil}^+(V) \to \{$*quadratic forms on* $V\}$, $\psi \mapsto Q_\psi$ *is a bijection;* $Q \mapsto \psi_Q$ *is its inverse.*

Proof. Clear. We just note that

$$\begin{aligned}
\psi_Q(v,v) &= \tfrac{1}{2}\big(Q(2v) - 2\,Q(v)\big) \\
&= \tfrac{1}{2}\big(4\,Q(v) - 2\,Q(v)\big) = Q(v),
\end{aligned}$$

as $Q(\lambda u) = \lambda^2 Q(u)$. $\square$

Remark. If $v_1, \ldots, v_n$ is a basis of V with $\psi(v_i, v_j) = a_{ij}$, then

$$Q\left(\sum x_i v_i\right) = \sum a_{ij} x_i x_j = x^{\mathsf{T}} A x,$$

that is, a quadratic form is a homogeneous polynomial of degree 2 in the variables $x_1, \ldots, x_n$.

Theorem 5.4

Let V be a finite dimensional vector space over $\mathbb{F}$ and $\psi \in \operatorname{Bil}^+(V)$ a symmetric bilinear form. Then there is some basis $v_1, \ldots, v_n$ of V such that $\psi(v_i, v_j) = 0$ if $i \neq j$. That is, we can choose a basis so that the matrix of ψ is diagonal.

Proof. Induct on $\dim V$. Now $\dim V = 1$ is clear. It is also clear if $\psi(v,w) = 0$ for all $v, w \in V$.

So assume otherwise. Then there exists a $w \in V$ such that $\psi(w,w) \neq 0$. (As if $\psi(w,w) = 0$ for all $w \in V$; that is, $Q(w) = 0$ for all $w \in V$, then by the lemma, $\psi(v,w) = 0$ for all $v, w \in V$.)

To continue, we need some notation. For an arbitrary $\psi \in \operatorname{Bil}(V)$, $U \leq V$, define

$$U^\perp = \big\{v \in V : \psi(u,v) = 0 \text{ for all } u \in U\big\}.$$

Claim. *$\langle w\rangle \oplus \langle w\rangle^{\perp} = V$ is a direct sum.*

[Proof of claim] As $\psi(w,w) \neq 0$, $w \notin \langle w\rangle^{\perp}$, so $\langle w\rangle \cap \langle w\rangle^{\perp} = 0$, and the sum is direct.

Now we must show $\langle w\rangle + \langle w\rangle^{\perp} = V$.

Let $v \in V$. Consider $v - \lambda w$. We want to find a λ such that $v - \lambda w \in \langle w\rangle^{\perp}$, as then $v = \lambda w + (v - \lambda w)$ shows $v \in \langle w\rangle + \langle w\rangle^{\perp}$.

But $v - \lambda w \in \langle w\rangle^{\perp} \iff \psi(w, v - \lambda w) = 0 \iff \psi(w,v) = \lambda\, \psi(w,w)$; that is, set

14 Nov

$$\lambda = \frac{\psi(v,w)}{\psi(w,w)}.$$

Now let $W = \langle w\rangle^{\perp}$, and $\psi' = \psi\,|_W \colon W \times W \to \mathbb{F}$ the restriction of ψ. This is symmetric bilinear, so by induction there is some basis $v_2, \dots, v_n$ of W such that $\psi(v_i, v_j) = \lambda_i\, \delta_{ij}$ for $\lambda_i \in \mathbb{F}$.

Hence, as $\psi(w, v_i) = \psi(v_i, w) = 0$ if $i \geq 2$, put $v_1 = w$ and we get that with respect to the basis $v_1, \dots, v_n$, the matrix of ψ is

$$\begin{pmatrix} \psi(w,w) & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}. \qquad \square$$

Warning. The diagonal entries are not determined by ψ, for example, consider

$$\begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} a_1^2\,\lambda_1 & & \\ & \ddots & \\ & & a_n^2\,\lambda_n \end{pmatrix},$$

that is, rescaling the basis element v_i to av_i changes $Q(a_i v_i) = a^2\, Q(v_i)$.

Also, we can reorder our basis – equivalently, take $P = P(w)$, the permutation matrix of $w \in S_n$, and note $P^{\mathsf{T}} = P(w^{-1})$, so

$$P(w)\, A\, P(w)^{\mathsf{T}} = P(w)\, A\, P(w)^{-1}.$$

Furthermore, it's not obvious that more complicated things can't happen, for example,

$$P \begin{pmatrix} 2 & \\ & 3 \end{pmatrix} P^{\mathsf{T}} = \begin{pmatrix} 5 & \\ & 30 \end{pmatrix} \text{ if } P = \begin{pmatrix} 1 & -3 \\ 1 & 2 \end{pmatrix}.$$

Corollary 5.5. *Let V be a finite dimensional vector space over $\mathbb{F}$, and suppose $\mathbb{F}$ is algebraically closed (such as $\mathbb{F} = \mathbb{C}$). Then*

$$\mathrm{Bil}^{+}(V) \xrightarrow{\sim} \{i : 0 \leq i \leq \dim V\},$$

under the isomorphism taking $\psi \mapsto \operatorname{rank} \psi$.

Proof. By the above, we can reorder and rescale so the matrix looks like

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

as $\sqrt{\lambda_i}$ is always in $\mathbb{F}$.

That is, there exists a basis of Q such that

$$Q\left(\sum_{i=1}^{n} x_i\, v_i\right) = \sum_{i=1}^{r} x_i^2,$$

where $r = \operatorname{rank} Q \leq n$.

Now let $\mathbb{F} = \mathbb{R}$, and $\psi : V \times V \to \mathbb{R}$ be bilinear symmetric.

By the theorem, there is some basis $v_1, \dots, v_n$ such that $\phi(v_i, v_j) = \lambda_i\, \delta_{ij}$. Replace v_i by $v_i/\sqrt{|\lambda_i|}$ if $\lambda_i \neq 0$ and reorder the baasis, we get ψ is represented by the matrix

$$\begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix},$$

for $p, q \geq 0$, that is, with respect to this basis

$$Q\left(\sum_{i=1}^{n} x_i\, v_i\right) = \sum_{i=1}^{p} x_i^2 - \sum_{i=p+1}^{p+q} x_i^2.$$

□

Note that $\operatorname{rank} \psi = p + q$.

Definition. The *signature* of ψ is $\operatorname{sign} \psi = p - q$.

We need to show this is well defined, and not an artefact of the basis chosen.

Theorem 5.6: Sylvester's law of inertia

The signature does not depend on the choice of basis; that is, if ψ is represented by

$$\begin{pmatrix} I_p & & \\ & I_q & \\ & & 0 \end{pmatrix} \text{ wrt } v_1, \dots, v_n \text{ and by } \begin{pmatrix} I_p' & & \\ & I_q' & \\ & & 0 \end{pmatrix} \text{ wrt } w_1, \dots, w_n,$$

then $p = p'$ and $q = q'$.

Warning: $\operatorname{tr}(P^\mathsf{T} A P) \neq \operatorname{tr}(A)$, so we can't prove it that way.

Definition. Let $Q : V \to \mathbb{R}$ be a quadratic form on V, where V is a vector space over $\mathbb{R}$, and $U \leq V$.

We say Q is *positive semi-definite on U* if for all $u \in U$, $Q(u) \geq 0$. Further, if $Q(u) = 0 \iff u = 0$, then we say that Q is *positive definite* on U.

If $U = V$, then we just say that Q is positive (semi) definite.

We define negative (semi) definite to mean $-Q$ is positive (semi) definite.

Proof of theorem. Let $P = \langle v_1, \dots, v_p \rangle$. So if $v = \sum_{i=1}^{p} \lambda_i v_i \in P$, $Q(v) = \sum_i \lambda_i^2 \geq 0$, and $Q(v) = 0 \iff v = 0$, so Q is positive definite on P.

Let $U = \langle v_{p+1}, \dots, v_{p+q}, \dots, v_n \rangle$, so Q is negative semi-definite on U. And now let P' be any positive definite subspace.

Claim. $P' \cap U = \{0\}$.

Proof of claim. If $v \in P'$, then $Q(v) \geq 0$; if $v \in U$, $Q(u) \leq 0$. so if $v \in P' \cap U$, $Q(v) = 0$. But if P' is positive definite, so $v = 0$. Hence

$$\dim P' + \dim U = \dim(P' + U) \leq \dim V = n,$$

and so

$$\dim P' \leq \dim V - \dim U = \dim P,$$

that is, p is the maximum dimension of any positive definite subspace, and hence $p' = p$. Similarly, q is the maximum dimension of any negative definite subspace, so $q' = q$. $\square$

Note that (p, q) determine (rank, sign), and conversely, $p = \frac{1}{2}(\text{rank} + \text{sign})$ and $q = \frac{1}{2}(\text{rank} - \text{sign})$. So we now have

$$\mathrm{Bil}^+(\mathbb{R}^n)/\operatorname{GL}_n(\mathbb{R}) \to \big\{(p,q) : p, q \geq 0, p + q \leq n\big\} \xrightarrow{\sim} \{(\text{rank}, \text{sgn})\}.$$

Example 5.7. Let $V = \mathbb{R}^2$, and $Q\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 - x_2^2$.

Consider the line $L_\lambda = \langle e_1 + \lambda e_2 \rangle$, $Q\binom{1}{\lambda} = 1 - \lambda^2$, so this is positive definite if $|\lambda| < 1$, and negative definite if $|\lambda| > 1$.

In particular, $p = q = 1$, but there are many choices of positive and negative definite subspaces of maximal dimension. (Recall that lines in $\mathbb{R}^2$ are parameterised by points on the circle $\mathbb{R} \cup \{\infty\}$).

Example 5.8. Compute the rank and signature of

$$Q(x, y, z) = x^2 + y^2 + 2z^2 + 2xy + 2xz - 2yz.$$

Note the matrix A of Q is

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 2 \end{pmatrix}, \text{ that is } Q \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x & y & z \end{pmatrix} A \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

(Recall that we for an arbitrary quadratic form Q, its matrix A is given by

$$Q\left(\textstyle\sum_i x_i v_i\right) = \textstyle\sum_{i,j} a_{ij}\, x_i\, x_j = \sum_i a_{ii}\, x_i^2 + \sum_{i<j} 2a_{ij}\, x_i\, x_j.$$

which is why the off-diagonal terms halved!)

We could apply the method in the proof of the theorem: begin by finding $w \in \mathbb{R}^3$ such that $Q(w) \neq 0$. Take $w = e_1 = (1, 0, 0)$. Now find $\langle w \rangle^\perp$. To do this, we seek λ such that $e_2 + \lambda e_1 \in \langle e_1 \rangle^\perp$. But $Q(e_1, e_2 + \lambda e_1) = 0$ implies $\lambda = -1$. Similarly we find $e_3 - e_1 \in \langle e_1 \rangle^\perp$, so $\langle e_1 \rangle^\perp = \langle e_2 - e_1, e_3 - e_1 \rangle$. Now continue with $Q \mid \langle e_1 \rangle^\perp$, and so on.

Here is a nicer way of writing this same compuation: row and column reduce A:

First, $R2 \mapsto R2 - R1$ and $C2 \mapsto C2 - C1$. In matrix form:

$$(I - E_{21})\, A\, (I - E_{12}) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & -2 \\ 1 & -2 & 2 \end{pmatrix}$$

Next $R3 \mapsto R3 - R1$ and $C3 \mapsto C3 - C1$, giving

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

Then swap $R2$, $R3$, and $C2$, $C3$, giving

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

Then $R3 \mapsto R3 + 2\,R2$ and $C3 \mapsto C3 + 2\,C2$, giving

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

Finally, rescale the last basis vector, giving

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

That is, if we put

$$P = (I - E_{12})\,(I - E_{13})\,P((2\ 3))\,(I - 2\,E_{23}) \begin{pmatrix} 1 & & \\ & 1 & \\ & & \frac{1}{2} \end{pmatrix},$$

then

$$P^{\mathsf{T}} A P = \begin{pmatrix} 1 & & \\ & 1 & \\ & & -1 \end{pmatrix}.$$

Method 2: we could just try to complete the square

$$Q(x, y, z) = (x + y + z)^2 + z^2 - 4yz = (x + y + z)^2 + (z - 2y)^2 - 4y^2$$

Remark. We will see in Chapter 6 that $\operatorname{sign}(A)$ is the number of positive eigenvalues minus the number of negative eigenvalues, so we could also compute it by computing the characteristic polynomial of A.

5.2 Anti-symmetric forms

We begin with a basis independant meaning of the rank of an arbitrary bilinear form.

Proposition 5.9. *Let V be a finite dimensional vector space over $\mathbb{F}$. Then*

$$\operatorname{rank} \psi = \dim V - \dim V^{\perp} = \dim V - \dim {}^{\perp}V,$$

where $V^{\perp} = \{v \in V : \psi(V, v) = 0\}$ *and* ${}^{\perp}V = \{v \in V : \psi(v, V) = 0\}$.

16 Nov *Proof.* Define a linear map $\operatorname{Bil}(V) \to \mathcal{L}(V, V^*)$, $\psi \mapsto \psi_L$ with $\psi_L(v)(w) = \psi(v, w)$. First we check that this is well-defined: $\psi(v, \cdot)$ linear implies $\psi_L(v) \in V^*$, and $\psi(\cdot, w)$ linear implies $\psi_L(\lambda v + \lambda' v') = \lambda\,\psi_L(v) + \lambda'\,\psi_L(v')$; that is, ψ_L is linear, and $\psi_L \in \mathcal{L}(V, V^*)$.

It is clear that the map is injective (as $\psi \not\equiv 0$ implies there are some v, w such that $\psi(v,w) \neq 0$, and so $\psi_L(v)(w) \neq 0$) and hence an isomorphism, as $\mathrm{Bil}(V)$ and $\mathcal{L}(V, V^*)$ are both vector spaces of dimension $(\dim V)^2$.

Let $v_1, \ldots, v_n$ be a basis of V, and $v_1^*, \ldots, v_n^*$ be the dual basis of V^*; that is, $v_i^*(v_j) = \delta_{ij}$. Let $A = (a_{ij})$ be the matrix of ψ_L with respect to these bases; that is,

$$\psi_L(v_j) = \sum_i a_{ij}\, v_i^*. \tag{$*$}$$

Apply both sides of $(*)$, to v_i, and we have

$$\psi(v_j, v_i) = \psi_L(v_j, v_i) = a_{ij}.$$

So the matrix of ψ with respect to the basis v_i is just A^T.

> **Exercise 5.10.** Define $\psi_R \in \mathcal{L}(V, V^*)$ by $\psi_R(v)(w) = \psi(w, v)$. Show the matrix of ψ_R is the matrix of ψ (which we've just seen is the transpose of the matrix of ψ_L).

Now we have

$$\operatorname{rank} A = \dim \operatorname{Im}(\psi_L : V \to V^*) = \dim V - \dim \ker \psi_L$$

and

$$\ker \psi_L = \big\{ v \in V : \psi(v, V) = 0 \big\} = {}^\perp V.$$

But also

$$\operatorname{rank} A = \operatorname{rank} A^\mathsf{T} = \dim \operatorname{Im}(\psi_R : V \to V^*) = \dim V - \dim \ker \psi_R,$$

and $\ker \psi_R = V^\perp$. $\square$

> **Definition.** A form $\psi \in \mathrm{Bil}(V)$ is *non-degenerate* if any of the following equivalent conditions hold:
>
> - $V^\perp = {}^\perp V = \{0\}$;
> - $\operatorname{rank} \psi = \dim V$;
> - $\psi_L : V \to V^*$ taking $v \mapsto \psi(v, \cdot)$ is an isomorphism;
> - $\psi_R : V \to V^*$ taking $v \mapsto \psi(\cdot, v)$ is an isomorphism;
> - for all $v \in V \backslash \{0\}$, there is some $w \in V$ such that $\psi(v, w) \neq 0$; that is, a non-degenerate bilinear form gives an isomorphism between V and V^*.

Proposition 5.11. *Let $W \leq V$ and $\psi \in \mathrm{Bil}(V)$. Then*

$$\dim W + \dim W^\perp - \dim(W \cap {}^\perp V) = \dim V.$$

Proof. Consider the map $V \to W^*$ taking $v \mapsto \psi(\cdot, v)$. (When we write $\psi(\cdot, v) : W \to \mathbb{F}$, we mean the map $w \mapsto \psi(w, v)$.) The kernel is

$$\ker = \big\{ v \in V : \psi(v, w) = 0 \ \forall w \in W \big\} = W^\perp,$$

so rank-nullity gives

$$\dim V = \dim W^\perp + \dim \operatorname{Im}.$$

So what is the image? Recall that

$$\dim \operatorname{Im}(\theta : V \to W^*) = \dim \operatorname{Im}(\theta^* : W = W^{**} \to V^*) = \dim W - \dim \ker \theta^*.$$

But
$$\theta^*(w) = \psi(w, \cdot) : V \to \mathbb{F},$$
and so $\theta^*(w) = 0$ if and only if $w \in {}^\perp V$; that is, $\ker \theta^* = W \cap {}^\perp V$, proving the proposition. $\square$

Remark. If you are comfortable with the notion of a quotient vector space, consider instead the map $V \to (W/W \cap {}^\perp V)^*$, $v \mapsto \psi(\cdot, v)$ and show it is well-defined, surjective and has $W^\perp$ as the kernel.

Example 5.12. If $V = \mathbb{R}^2$, and $Q\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 - x_2^2$, then $A = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$.

Then if $W = \left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle$, $W^\perp = W$ and the proposition says $1 + 1 - 0 = 2$.

Or if we let $V = \mathbb{C}^2$, $Q\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + x_2^2$, so $A = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$, and set $W = \left\langle \begin{pmatrix} 1 \\ i \end{pmatrix} \right\rangle$ then $W = W^\perp$.

Corollary 5.13. $\psi\big|_W : W \times W \to \mathbb{F}$ *is non-degenerate if and only if* $V = W \oplus W^\perp$.

Proof. ($\Leftarrow$) $\psi\big|_W$ is non-degenerate means that for all $w \in W \backslash \{0\}$, there is some $w' \in W$ such that $\psi(w, w') \neq 0$, so if $w \in W^\perp \cap W$, $w \neq 0$, then for all $w' \in W$, $\psi(w, w') = 0$, a contradiction, and so $W \cap W^\perp = \{0\}$. Now
$$\dim(W + W^\perp) = \dim W + \dim W^\perp \geq \dim V,$$
by the proposition, so $W + W^\perp = V$ (and also ψ is non-degenerate on all of V clearly).

($\Rightarrow$) Clear by our earlier remarks that $W \cap W^\perp = 0$ if and only if $\psi\big|_W$ is non-degenerate. $\square$

Theorem 5.14

Let $\psi \in \mathrm{Bil}^-(V)$ be an *anti*-symmetric bilinear form. Then there is some basis $v_1, \ldots, v_n$ of V such that the matrix of ψ is
$$\begin{pmatrix} \boxed{\begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix}} & & & & & 0 \\ & \boxed{\begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix}} & & & & \\ & & \ddots & & & \\ & & & \boxed{\begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix}} & & \\ & & & & 0 & \\ & & & & & \ddots \\ 0 & & & & & 0 \end{pmatrix}$$
In particular, $\operatorname{rank} \psi$ is *even*! ($\mathbb{F}$ is arbitrary.)

Remark. If $\psi \in \mathrm{Bil}^\pm(V)$, then $W^\perp = {}^\perp W$ for all $W \leq V$.

Proof. We induct on $\operatorname{rank}\psi$, if $\operatorname{rank}\psi = 0$, then $\psi = 0$ and we're done.

Otherwise, there are some $v_1, v_2 \in V$ such that $\psi(v_1, v_2) \neq 0$. If $v_2 = \lambda v_1$ then $\psi(v_1, \lambda v_2) = \lambda\, \psi(v_1, v_1) = 0$, as ψ is anti-symmetric; so v_1, v_2 are linearly independent. Change v_2 to $v_2/\psi(v_1, v_2)$.

So now $\psi(v_1, v_2) = 1$. Put $W = \langle v_1, v_2 \rangle$, then $\psi\big|_W$ has matrix $\boxed{\begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix}}$, which is non-degenerate, so the corollary gives $V = W \oplus W^\perp$. And now induction gives the basis of $W^\perp$, $v_3, \dots, v_n$, of the correct form, and $v_1, \dots, v_n$ is our basis. □

So we've shown that there is an isomorphism

$$\operatorname{Bil}^-(V)/\operatorname{GL}(V) \xrightarrow{\sim} \left\{2i : 0 \le i \le \tfrac{1}{2}\dim V\right\}.$$

taking $\psi \mapsto \operatorname{rank}\psi$.

Remark. A non-degenerate anti-symmetric form ψ is usually called a *symplectic form*.

Let $\psi \in \operatorname{Bil}^-(V)$ be non-degenerate, $\operatorname{rank}\psi = n = \dim V$ (even!). Put $L = \langle v_1, v_3, v_5, \dots \rangle$, with $v_1, \dots, v_n$ as above, and then $L^\perp = L$. Such a subspace is called *Lagrangian*.

If $U \le L$, then $U^\perp \ge L^\perp = L$, and so $U \subseteq U^\perp$. Such a subspace is called *isotropic*.

Definition. If $\psi \in \operatorname{Bil}(V)$, the *isometries* of ψ are

$$\begin{aligned} \operatorname{Isom}\psi &= \{g \in \operatorname{GL}(V) : g\psi = \psi\} \\ &= \left\{g \in \operatorname{GL}(V) : \psi(g^{-1}v, g^{-1}w) = \psi(v, w)\ \forall v, w \in V\right\} \\ &= \left\{X \in \operatorname{GL}_n(\mathbb{F}) : XAX^\mathsf{T} = A\right\} \text{ if } A \text{ is a matrix of } \psi. \end{aligned}$$

This is a group.

Exercise 5.15. Show that $\operatorname{Isom}(g\psi) = g\operatorname{Isom}(\psi)g^{-1}$, and so isomorphism bilinear forms have isomorphic isometry groups.

If $\psi \in \operatorname{Bil}^+(V)$, ψ is non-degenerate, we often write $O(\psi)$, the *orthogonal group of* ψ for the isometry group of ψ.

Example 5.16. Suppose $\mathbb{F} = \mathbb{C}$. If $\psi \in \operatorname{Bil}^+(V)$, and ψ is non-degenerate, then ψ is isomorphic to the standard quadratic form, whose matrix $A = I$, and so $\operatorname{Isom}\psi$ is conjugate to the group

$$\operatorname{Isom}(A = I) = \left\{X \in \operatorname{GL}_n(\mathbb{C}) : XX^\mathsf{T} = I\right\} = O_n(\mathbb{C}),$$

which is what we usually call the orthogonal group.

If $\mathbb{F} = \mathbb{R}$, then

$$O_{p,q}(\mathbb{R}) = \left\{X \mid X \begin{pmatrix} I_p & \\ & -I_q \end{pmatrix} X^\mathsf{T} = \begin{pmatrix} I_p & \\ & -I_q \end{pmatrix}\right\}.$$

are the possible isometry groups of non-degenerate symmetric forms.

For any field $\mathbb{F}$, if $\psi \in \mathrm{Bil}^-(\mathbb{F})$ is non-degenerate, then $\mathrm{Isom}\,\psi$ is called the *symplectic group*, and it is conjugate to the group

$$\mathrm{Sp}_{2n}(\mathbb{F}) = \left\{X : XJX^\mathsf{T} = J\right\},$$

where J is the matrix given by

$$J = \begin{pmatrix} \boxed{\begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix}} & & & 0 \\ & \boxed{\begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix}} & & \\ & & \ddots & \\ 0 & & & \boxed{\begin{matrix} 0 & 1 \\ -1 & 0 \end{matrix}} \end{pmatrix}$$

6 Hermitian forms

19 Nov

A non-degenerate quadratic form on a vector space over $\mathbb{C}$ doesn't behave like an inner product on $\mathbb{R}^2$. For example,

$$\text{if } Q\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + x_2^2 \qquad \text{then we have} \qquad Q\begin{pmatrix} 1 \\ i \end{pmatrix} = 1 + i^2 = 0.$$

We don't have a notion of positive definite, but there is a modification of a notion of a bilinear form which does.

Definition. Let V be a vector space over $\mathbb{C}$; then a function $\psi : V \times V \to \mathbb{C}$ is called *sesquilinear* if

(i) For all $v \in V$, $\psi(\cdot, v)$, $u \mapsto \psi(u, v)$ is linear; that is

$$\psi(\lambda_1 u_1 + \lambda_2 u_2, v) = \lambda_1 \psi(u_1, v) + \lambda_2 \psi(u_2, v)$$

(ii) For all $u, v_1, v_2 \in V$, $\lambda_1, \lambda_2 \in \mathbb{C}$,

$$\psi(u, \lambda_1 v_1 + \lambda_2 v_2) = \overline{\lambda_1}\, \psi(u, v_1) + \overline{\lambda_2}\, \psi(u, v_2),$$

where $\overline{z}$ is the complex conjugate of z.

It is called *Hermitian* if it also satisfies

(iii) $\psi(v, w) = \overline{\psi(w, v)}$ for all $v, w \in V$.

Note that (i) and (iii) imply (ii).

Let V be a vector space over $\mathbb{C}$, and $\psi : V \times V \to \mathbb{C}$ a Hermitian form. Define

$$Q(v) = \psi(v, v) = \overline{\psi(v, v)}$$

by (iii), so $Q : V \to \mathbb{R}$.

Lemma 6.1. *We have $Q(v) = 0$ for all $v \in V$ if and only if $\psi(v, w) = 0$ for all $v, w \in V$.*

Proof. We have.

$$\begin{aligned} Q(u \pm v) = \psi(u \pm v, u \pm v) &= \psi(u, u) + \psi(v, v) \pm \psi(u, v) \pm \psi(v, u) \\ &= Q(u) + Q(v) \pm 2\,\Re\,\psi(u, v), \end{aligned}$$

as $z + \overline{z} = 2\,\Re(z)$. Thus

$$\begin{aligned} Q(u + v) - Q(u - v) &= 4\,\Re\,\psi(u, v), \\ Q(u + iv) - Q(u - iv) &= 4\,\Im\,\psi(u, v), \end{aligned}$$

that is, $Q : V \to \mathbb{R}$ determines $\psi : V \times V \to \mathbb{C}$ if Q is Hermitian:

$$\psi(u, v) = \tfrac{1}{4}\left[Q(u + v) + i\,Q(u + iv) - Q(u - v) - i\,Q(u - iv)\right]. \qquad \square$$

Note that

$$Q(\lambda v) = \psi(\lambda v, \lambda v) = \lambda \overline{\lambda}\, \psi(v, v) = |\lambda|^2\, Q(v).$$

If $\psi : V \times V \to \mathbb{C}$ is Hermitian, and $v_1, \ldots, v_n$ is a basis of V, then we write $A = (a_{ij})$, $a_{ij} = \psi(v_i, v_j)$, and we call this the *matrix of ψ with respect to $v_1, \ldots, v_n$*.

Observe that $A^{\mathsf{T}} = \overline{A}$; that is, A is a Hermitian matrix.

Exercise 6.2. Show that if we change basis, $v_j' = \sum_i p_{ij}\, v_i$, $P = (p_{ij})$, then $A \mapsto P^\mathsf{T} A \overline{P}$.

Theorem 6.3

If V is a finite dimensional vector space over $\mathbb{C}$, and $\psi : V \times V \to \mathbb{C}$ is Hermitian, then there is a basis for V such that the matrix A of ψ is

$$\begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix}$$

for some $p, q \geq 0$. Moreover, p and q are uniquely determined by ψ: p is the maximum dimension of a positive definite subspace P, and q is the maximal dimension of a negative definite subspace.

Here $P \leq V$ is positive definite if $Q(v) \geq 0$ for all $v \in P$, and $Q(v) = 0$ when $v \in P$ implies $v = 0$; P is negative definite $-Q$ is positive definite on P.

Proof. Exactly as for real symmetric forms, using the Hermitian ingredients before the theorem instead of their bilinear counterparts.

Definition. If $W \leq V$, then the *orthogonal complement to* W is given by

$$W^\perp = \left\{ v \in V \mid \psi(W, v) = \psi(v, W) = 0 \right\} = {}^\perp W.$$

We say that ψ is *non-degenerate* if $V^\perp = 0$, equivalenyly if $p + q = \dim V$. We also define the *unitary group*

$$\begin{aligned} U(p,q) &= \mathrm{Isom}\begin{pmatrix} I_p & \\ & -I_q \end{pmatrix} \\ &= \left\{ X \in GL_n(\mathbb{C}) \mid X^\mathsf{T} \begin{pmatrix} I_p & \\ & -I_q \end{pmatrix} \overline{X} = \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix} \right\} \\ &= \left\{ \text{stabilizers of the form } \begin{pmatrix} I_p & \\ & -I_q \end{pmatrix} \text{ with respect to } \mathrm{GL}_n(\mathbb{C}) \text{ action} \right\}, \end{aligned}$$

where the action takes $\psi \mapsto g\psi$, with $(g\psi)(x, y) = \psi(g^{-1}x, g^{-1}y)$. Again, note g^{-1} here so that $(gh)\psi = g(h\psi)$.

In the special case where the form ψ is positive definite, that is, conjugate to I_n, we call this the *unitary group*

$$U(n) = U(n, 0) = \left\{ X \in \mathrm{GL}_n(\mathbb{C}) \mid X^\mathsf{T}\overline{X} = I \right\}.$$

Proposition 6.4. *Let V be a vector space over $\mathbb{C}$ (or $\mathbb{R}$), and $\psi : V \times V \to \mathbb{C}$ (or $\mathbb{R}$) a Hermitian (respectively, symmetric) form, so $Q : V \to \mathbb{R}$.*

Let $v_1, \ldots, v_n$ be a basis of V, and $A \in \mathrm{Mat}_n(\mathbb{C})$ the matrix of ψ, so $A^\mathsf{T} = \overline{A}$. Then $Q : V \to \mathbb{R}$ is positive definite if and only if, for all k, $1 \leq k \leq n$, the top left $k \times k$ submatrix of A (called A_k) has $\det A_k \in \mathbb{R}$ and $\det A_k > 0$.

Proof. ($\Rightarrow$) If Q is positive definite, then $A = P^\mathsf{T} I \overline{P} = P^\mathsf{T}\overline{P}$ for some $P \in \mathrm{GL}_n(\mathbb{C})$, and so

$$\det A = \det P^\mathsf{T} \det \overline{P} = \left|\det P\right|^2 > 0. \tag{$*$}$$

But if $U \le V$, as Q is positive definite on V, it is positive definite on U. Take $U = \langle v_1, \dots, v_k\rangle$, then $Q \mid U$ is positive definite, A_k is the matrix of $Q \mid U$, and by ($*$), $\det A_k > 0$.

($\Leftarrow$) Induct on $n = \dim V$. The case $n = 1$ is clear. Now the induction hypothesis tells us that $\psi \mid \langle v_1, \dots, v_{n-1}\rangle$ is positive definite, and hence the dimension of a maximum positive definite subspace is $p \ge n - 1$.

So by classification of Hermitian forms, there is some $P \in \mathrm{GL}_n(\mathbb{C})$ such that

$$A = P^\mathsf{T} \begin{pmatrix} I_{n-1} & 0 \\ 0 & c \end{pmatrix} \overline{P},$$

where $c = 0, 1$ or -1. But $\det A = \left|\det P\right|^2 c > 0$ by assumption, so $c = 1$, and $A = P^\mathsf{T}\overline{P}$; that is, Q is positive definite. □

Definition. If V is a vector space over $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, then an *inner product on* V is a positive definite symmetric bilinear/Hermitian form $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{F}$, and we say that V is an *inner product space.*

Example 6.5. Consider $\mathbb{R}^n$ or $\mathbb{C}^n$, and the dot product $\langle \mathbf{x}, \mathbf{y}\rangle = \sum x_i \overline{y}_i$. These forms behave exactly as our intuition tells us in $\mathbb{R}^2$.

6.1 Inner product spaces

Definition. Let V be an inner product space over $\mathbb{F}$ with $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$. Then $Q(v) \in \mathbb{R}_{\ge 0}$, and so we can define

$$|v| = \sqrt{Q(v)}$$

to be the *length* or *norm* of v. Note that $|v| = 0$ if and only if $v = 0$.

Lemma 6.6 (Cauchy-Schwarz inequality)**.** $\left|\langle v, w\rangle\right| \le |v|\,|w|$.

Proof. As you've seen many times before:

$$\begin{aligned} 0 &\le \langle -\lambda v + w, -\lambda v + w\rangle \\ &= |\lambda|^2 \langle v, v\rangle + \langle w, w\rangle - \lambda \langle v, w\rangle - \overline{\lambda \langle v, w\rangle}. \end{aligned}$$

The result is clear if $v = 0$, otherwise suppose $|v| \ne 0$, and put $\lambda = \overline{\langle v, w\rangle} / \langle v, w\rangle$. We get

$$0 \le \frac{\left|\langle v, w\rangle\right|^2}{\langle v, v\rangle} - \frac{2\left|\langle v, w\rangle\right|^2}{\langle v, v\rangle} + \langle w, w\rangle,$$

that is, $\left|\langle v, w\rangle\right|^2 \le \langle v, v\rangle \langle w, w\rangle$. □

Note that if $\mathbb{F} = \mathbb{R}$, then $\langle v, w\rangle / |w|\,|v| \in [-1, 1]$ so there is some $\theta \in [0, \pi)$ such that $\cos\theta = \langle v, w\rangle / |v|\,|w|$. We call θ the *angle between* v *and* w.

Corollary 6.7 (Triangle inequality)**.** *For all $v, w \in V$, $|v+w| \leq |v| + |w|$.*

Proof. As you've seen many times before:

$$\begin{aligned} |v+w|^2 &= \langle v+w, v+w \rangle \\ &= |v|^2 + 2\Re \langle v, w \rangle + |w|^2 \\ &\leq \left|v^2\right| + 2\,|v|\,|w| + \left|w^2\right| \quad \text{(by lemma)} \\ &= \left(|v| + |w|\right)^2. \end{aligned}$$

□

21 Nov

Given $v_1, \dots, v_n$ with $\left\langle v_i, v_j \right\rangle = 0$ if $i \neq j$, we say that $v_1, \dots, v_n$ are *orthogonal.* If $\left\langle v_i, v_j \right\rangle = \delta_{ij}$, then we say that $v_1, \dots, v_n$ are *orthonormal.*

So $v_1, \dots, v_n$ orthogonal and $v_i \neq 0$ for all i implies that $\hat{v}_1, \dots, \hat{v}_n$ are orthonormal, where $\hat{v}_i = v_i / |v_i|$.

Lemma 6.8. *If $v_1, \dots, v_n$ are non-zero and orthogonal, and if $v = \sum_{i=1}^n \lambda_i v_i$, then $\lambda_i = \langle v, v_i \rangle / |v_i|^2$.*

Proof. $\langle v, v_k \rangle = \sum_{i=1}^n \lambda_i \langle v_i, v_k \rangle = \lambda_k \langle v_k, v_k \rangle$, hence the result. □

In particular, distinct orthonormal vectors $v_1, \dots, v_n$ are linearly independent, since $\sum_i \lambda_i v_i = 0$ implies $\lambda_i = 0$.

As $\langle \cdot, \cdot \rangle$ is Hermitian, we know there is a basis $v_1, \dots, v_n$ such that the matrix of $\langle \cdot, \cdot \rangle$ is

$$\begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix}.$$

As $\langle \cdot, \cdot \rangle$ is positive definite, we know that $p = n$, $q = 0$, rank $= \dim V$; that is, this matrix is I_n. So we know there exists an orthonormal basis $v_1, \dots, v_n$; that is $V \cong \mathbb{R}^n$, with $\langle x, y \rangle = \sum_i x_i y_i$, or $V \cong \mathbb{C}^n$, with $\langle x, y \rangle = \sum_i x_i \overline{y_i}$.

Here is another constructive proof that orthonormal bases exist.

Theorem 6.9: Gram-Schmidt orthogonalisation

Let V have a basis $v_1, \dots, v_n$. Then there exists an orthonormal basis $e_1, \dots, e_n$ such that $\langle v_1, \dots, v_k \rangle = \langle e_1, \dots, e_k \rangle$ for all $1 \leq k \leq n$.

Proof. Induct on k. For $k = 1$, set $e_1 = v_1 / |v_1|$.

Suppose we've found $e_1, \dots, e_k$ such that $\langle e_1, \dots, e_k \rangle = \langle v_1, \dots, v_k \rangle$. Define

$$\tilde{e}_{k+1} = v_{k+1} - \sum_{1 \leq i \leq k} \langle v_{k+1}, e_i \rangle\, e_i.$$

Thus

$$\langle \tilde{e}_{k+1}, e_i \rangle = \langle v_{k+1}, e_i \rangle - \langle v_{k+1}, e_i \rangle = 0 \text{ if } i \leq k.$$

Also $\tilde{e}_{k+1} \neq 0$, as if $\tilde{e}_{k+1} = 0$, then $v_{k+1} \in \langle e_1, \dots, e_k \rangle = \langle v_1, \dots, v_k \rangle$ which contradicts $v_1, \dots, v_{k+1}$ linearly independent.

So put $e_{k+1} = \tilde{e}_{k+1} / |\tilde{e}_{k+1}|$, and then $e_1, \dots, e_{k+1}$ are orthonormal, and $\langle e_1, \dots, e_{k+1} \rangle = \langle v_1, \dots, v_{k+1} \rangle$. □

Corollary 6.10. *Any orthonormal set can be extended to an orthonormal basis.*

Proof. Extend the orthonormal set to a basis; now the Gram-Schmidt algorithm doesn't change $v_1, \dots, v_k$ if they are already orthonormal. $\square$

Recall that if $W \leq V$, $W^\perp = {}^\perp W = \{v \in V \mid \langle v, w \rangle = 0 \; \forall w \in W\}$.

Proposition 6.11. *If $W \leq V$, V an inner product space, then $W \oplus W^\perp = V$.*

Proof 1. If $\langle \cdot, \cdot \rangle$ is positive definite on V, then it is also positive definite on W, and thus $\langle \cdot, \cdot \rangle|_W$ is non-degenerate. If $\mathbb{F} = \mathbb{R}$, then $\langle \cdot, \cdot \rangle$ is bilinear, and we've shown that $W \oplus W^\perp = V$ when the form $\langle \cdot, \cdot \rangle \,|_W$ is non-degenerate. If $\mathbb{F} = \mathbb{C}$, then exactly the same proof for sesquilinear forms shows the result. $\square$

Proof 2. Pick an orthonormal basis $w_1, \dots, w_r$ for W, and extend it to an orthonormal basis for V, $w_1, \dots, w_n$.

Now observe that $\langle w_{r+1}, \dots, w_n \rangle = W^\perp$. Proof ($\subseteq$) is done. For ($\supseteq$): if $\sum_{i=1}^n \lambda_i w_i \in W^\perp$, then take $\langle \cdot, w_i \rangle$, $i \leq r$, and we get $\lambda_i = 0$ for $i \leq r$. So $V = W \oplus W^\perp$. $\square$

Geometric interpretation of the key step in the Gram-Schmidt algorithm

Let V be an inner product space, with $W \leq V$ and $V = W \oplus W^\perp$. Define a map $\pi : V \to W$, the *orthogonal projection onto* W, defined as follows: if $v \in V$, then write $v = w + w'$, where $w \in W$ and $w' \in W^\perp$ uniquely, and set $\pi(v) = w$.

This satisfies $\pi|_W = \mathrm{id} : W \to W$, $\pi^2 = \pi$ and π linear.

Proposition 6.12. *If W has an orthonormal basis $e_1, \dots, e_k$ and $\pi : V \to W$ as above, then*

(i) $\pi(v) = \sum_{i=1}^k \langle v, e_i \rangle e_i$;

(ii) *$\pi(v)$ is the vector in W closest to v; that is, $\left|v - \pi(v)\right| \leq |v - w|$ for all $w \in W$, with equality if and only if $w = \pi(v)$.*

Proof.

(i) If $v \in V$, then put $w = \sum_{i=1}^k \langle v, e_i \rangle e_i$, and $w' = v - w$. So $w \in W$, and we want $w' \in W^\perp$. But

$$\langle w', e_i \rangle = \langle v, e_i \rangle - \langle v, e_i \rangle = 0 \text{ for all i, } 1 \leq i \leq k,$$

so indeed we have $w' \in W^\perp$, and $\pi(v) = w$ by definition.

(ii) We have $v - \pi(v) \in W^\perp$, and if $w \in W$, $\pi(v) - w \in W$, then

$$\begin{aligned} |v - w|^2 &= \left|\big(v - \pi(v)\big) + \big(\pi(v) - w\big)\right|^2 \\ &= \left|v - \pi(v)\right|^2 + \left|\pi(v) - w\right|^2 + 2\Re \underbrace{\langle \underset{\in W^\perp}{v - \pi(v)}, \underset{\in W}{\pi(v) - w} \rangle}_{=0}, \end{aligned}$$

and so $|v - w|^2 \geq \left|v - \pi(v)\right|^2$, with equality if and only if $\left|\pi(v) - w\right| = 0$; that is, if $\pi(v) = w$. $\square$

6.2 Hermitian adjoints for inner products

Let V and W be inner product spaces over F and $\alpha : V \to W$ a linear map.

Proposition 6.13. *There is a unique* linear *map* $\alpha^* : W \to V$ *such that for all* $v \in V$, $w \in W$, $\langle \alpha(v), w \rangle = \langle v, \alpha^*(w) \rangle$. *This map is called the* Hermitian adjoint*.*

Moreover, if $e_1, \ldots, e_n$ *is an orthonormal basis of* V*, and* $f_1, \ldots, f_m$ *is an orthonormal basis for* W*, and* $A = (a_{ij})$ *is the matrix of* α *with respect to these bases, then* $\overline{A^\mathsf{T}}$ *is the matrix of* α^**.*

Proof. If $\beta : W \to V$ is a linear map with matrix $B = (b_{ij})$, then

$$\langle \alpha(v), w \rangle = \langle v, \beta(w) \rangle \text{ for all } v, w$$

if and only if

$$\langle \alpha(e_j), f_k \rangle = \langle e_j, \beta(f_k) \rangle \text{ for all } 0 \le j \le n, 0 \le k \le m.$$

But we have

$$a_{kj} = \left\langle \sum a_{ij}\, f_i, f_k \right\rangle = \langle \alpha(e_j), f_k \rangle = \langle e_j, \beta(f_k) \rangle = \left\langle e_j, \sum b_{ik}\, e_i \right\rangle = \bar{b}_{jk},$$

that is, $B = \overline{A^\mathsf{T}}$. Now define α^* to be the map with matrix $\overline{A^\mathsf{T}}$. □

Exercise 6.14. If $\mathbb{F} = \mathbb{R}$, identify $V \xrightarrow{\sim} V^*$ by $v \mapsto \langle v, \cdot \rangle$, $W \xrightarrow{\sim} W^*$ by $w \mapsto \langle w, \cdot \rangle$, and then show that α^* is just the dual map.

More generally, if $\alpha : V \to W$ defines a linear map over $\mathbb{F}$, $\psi \in \mathrm{Bil}(V)$, $\psi' \in \mathrm{Bil}(W)$, both non-degenerate, then you can define the adjoint by $\psi'(\alpha(v), w) = \psi(v, \alpha^*(w))$

23 Nov for all $v \in V$, $w \in W$, and show that it is the dual map.

Lemma 6.15.

(i) If $\alpha, \beta : V \to W$*, then* $(\alpha + \beta)^* = \alpha^* + \beta^*$*.*

(ii) $(\lambda\alpha)^* = \overline{\lambda}\alpha^*$*.*

(iii) $\alpha^{**} = \alpha$*.*

Proof. Immediate from the properties of $A \to \overline{A^\mathsf{T}}$.

Definition. A map $\alpha : V \to V$ is self-adjoint if $\alpha = \alpha^*$.

If $v_1, \ldots, v_n$ is an orthonormal basis for V, and A is the matrix of α, then α is self-adjoint if and only if $A = \overline{A}^\mathsf{T}$.

In short, if $\mathbb{F} = \mathbb{R}$, then A is symmetric, and if $\mathbb{F} = \mathbb{C}$, then A is Hermitian.

Theorem 6.16

Let $\alpha : V \to V$ be self-adjoint. Then

(i) All the eigenvalues of α are real.

(ii) Eigenvectors with distinct eigenvalues are orthogonal.

(iii) There exists an orthogonal basis of eigenvectors for α. In particular, α is diagonalisable.

Proof.

(i) First assume $\mathbb{F} = \mathbb{C}$. If $\alpha v = \lambda v$ for a non-zero vector v and $\lambda \in \mathbb{C}$, then

$$\lambda \langle v, v \rangle = \langle \lambda v, v \rangle = \big\langle v, \alpha^* v \big\rangle = \langle v, \alpha v \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \langle v, v \rangle ,$$

as α is self-adjoint. Since $v \neq 0$, we have $\langle v, v \rangle \neq 0$ and thus $\lambda = \overline{\lambda}$.

If $\mathbb{F} = \mathbb{R}$, then let $A = A^{\mathsf{T}}$ be the matrix of α; regard it as a matrix over $\mathbb{C}$, which is obviously Hermitian, and then the above shows that the eigenvalue for A is real.

Remark. This shows that we should introduce some notation so that we can phrase this argument without choosing a basis. Here is one way: let V be a vector space over $\mathbb{R}$. Define a new vector space, $V_{\mathbb{C}} = V \oplus iV$, a new vector space over $\mathbb{R}$ of twice the dimension, and make it a complex vector space by saying that $i\,(v + iw) = (-w + iv)$, so $\dim_{\mathbb{R}} V = \dim_{\mathbb{C}} V_{\mathbb{C}}$. Now suppose the matrix of $\alpha : V \to V$ is A. Then show the matrix of $\alpha_{\mathbb{C}} : V_{\mathbb{C}} \to V_{\mathbb{C}}$ is also A, where $\alpha_{\mathbb{C}}(v + iw) = \alpha(v) + i\,\alpha(w)$.

Now we can phrase (i) of the proof using $V_{\mathbb{C}}$: show $\lambda \in \mathbb{R}$ implies that we can choose a λ-eigenvector $v \in V_{\mathbb{C}}$ to be in $V \subseteq V_{\mathbb{C}}$.

(ii) If $\alpha(v_i) = \lambda_i v_i$, $i = 1, 2$, where $v_i \neq 0$ and $\lambda_1 \neq \lambda_2$, then

$$\lambda_1 \langle v_1, v_2 \rangle = \langle \alpha v_1, v_2 \rangle = \langle v_1, \alpha v_2 \rangle = \overline{\lambda}_2 \langle v_1, v_2 \rangle ,$$

as $\alpha = \alpha^*$, so if $\langle v_1, v_2 \rangle \neq 0$, then $\lambda_1 = \overline{\lambda}_2 = \lambda_2$, a contradiction.

(iii) Induct on $\dim V$. The case $\dim V = 1$ is clear, so assume $n = \dim V > 1$. By (i), there is a real eigenvalue λ, and an eigenvector $v_1 \in V$ such that $\alpha(v_1) = \lambda v_1$. Thus $V = \langle v_1 \rangle \oplus \langle v_1 \rangle^{\perp}$ as V is an inner product space. Now put $W = \langle v_1 \rangle^{\perp}$.

Claim. $\alpha(W) \subseteq W$; that is, if $\langle x, v_1 \rangle = 0$, then $\big\langle \alpha(x), v_1 \big\rangle = 0$.

Proof. We have

$$\big\langle \alpha(x), v_1 \big\rangle = \big\langle x, \alpha^*(v_1) \big\rangle = \big\langle x, \alpha(v_1) \big\rangle = \overline{\lambda} \langle x, v_1 \rangle = 0.$$

Also, $\alpha|_W : W \to W$ is self-adjoint, as $\big\langle \alpha(v), w \big\rangle = \big\langle v, \alpha(w) \big\rangle$ for all $v, w \in V$, and so this is also true for all $v, w \in W$. Hence by induction W has an orthonormal basis $v_2, \dots, v_n$, and so $\hat{v}_1, v_2, \dots, v_n$ is an orthonormal basis for V. □

Definition. Let V be an inner product space over $\mathbb{C}$. Then the *group of isometries* of the form $\langle \cdot, \cdot \rangle$, denoted $U(V)$, is defined to be

$$\begin{aligned} U(V) = \operatorname{Isom}(V) &= \Big\{ \alpha : V \to V \mid \big\langle \alpha(v), \alpha(w) \big\rangle = \langle v, w \rangle \ \forall v, w \in V \Big\} \\ &= \Big\{ \alpha \in \mathrm{GL}(V) \mid \big\langle \alpha(v), w' \big\rangle = \Big\langle v, \alpha^{-1} w' \Big\rangle \ \forall v, w' \in V \Big\}, \end{aligned}$$

putting $w' = \alpha(w)$. Now we note that $\alpha : V \to V$ an isometry implies that α is an isomorphism. This is because $v \neq 0$ if and only if $|v| \neq 0$, and α is an isometry, so we have $|\alpha v| = |v| \neq 0$, and so α is injective.

$$= \Big\{ \alpha \in \mathrm{GL}(V) \mid \alpha^{-1} = \alpha^* \Big\} .$$

This is called the *unitary group*.

If $V = \mathbb{C}^n$, and $\langle \cdot, \cdot \rangle$ is the standard inner product $\langle x, y \rangle = \sum_i x_i \overline{y}_i$, then we write

$$U_n = U(n) = U(\mathbb{C}^n) = \left\{ X \in GL_n(\mathbb{C}) \mid \overline{X}^{\mathsf{T}} \cdot X = I \right\}.$$

So an orthonormal basis (that is, a choice of isomorphism $V \xrightarrow{\sim} \mathbb{C}^n$) gives us an isomorphism $U(V) \xrightarrow{\sim} U_n$.

Theorem 6.17

Let V be an inner product space over $\mathbb{C}$, and $\alpha : V \to V$ an isometry; that is, $\alpha^* = \alpha^{-1}$, and $\alpha \in U(V)$. Then

(i) All eigenvalues λ of α have $|\lambda| = 1$; that is, they lie on the unit circle.

(ii) Eigenvectors with distinct eigenvalues are orthogonal.

(iii) There exists an orthonormal basis of eigenvectors for α; in particular α is diagonalisable.

Remark. If V is an inner product space over $\mathbb{R}$, then $\operatorname{Isom} \langle \cdot, \cdot \rangle = O(V)$, the usual orthogonal group, also denoted $O_n(\mathbb{R})$. If we choose an orthonormal basis for V, then $\alpha \in O(V)$ if A, the matrix of α, has $A^{\mathsf{T}} A = I$.

Then this theorem applied to A considered as a complex matrix shows that A is diagonalisable over $\mathbb{C}$, but as all the eigenvalues of A have $|\lambda| = 1$, it is not diagonalisable over $\mathbb{R}$ unles the only eigenvalues are ± 1.

Example 6.18. The matrix

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = A \in O(2)$$

is diagonalisable over $\mathbb{C}$, and conjugate to

$$\begin{pmatrix} e^{i\theta} & \\ & e^{-i\theta} \end{pmatrix},$$

but not over $\mathbb{R}$, unless $\sin\theta = 0$.

Proof.

(i) If $\alpha(v) = \lambda v$, for v non-zero, then

$$\lambda \langle v, v \rangle = \langle \lambda v, v \rangle = \langle \alpha(v), v \rangle = \langle v, \alpha^*(v) \rangle = \left\langle v, \alpha^{-1}(v) \right\rangle = \left\langle v, \lambda^{-1} v \right\rangle = \overline{\lambda}^{-1} \langle v, v \rangle,$$

and so $\lambda = \overline{\lambda}^{-1}$ and $\lambda\overline{\lambda} = 1$.

(ii) If $\alpha(v_i) = \lambda_i v_i$, for v non-zero and $\lambda_i \neq \lambda_j$:

$$\lambda_i \langle v_i, v_j \rangle = \langle \alpha(v_i), v_j \rangle = \left\langle v_i, \alpha^{-1}(v_j) \right\rangle = \overline{\lambda}_j^{-1} \langle v_i, v_j \rangle = \lambda_j \langle v_i, v_j \rangle,$$

and so $\lambda_i \neq \lambda_j$ implies $\langle v_i, v_j \rangle = 0$.

(iii) Induct on $n = \dim V$. If V is a vector space over $\mathbb{C}$, then a non-zero eigenvector v_1 exists with some eigenvalue λ, so $\alpha(v_1) = \lambda v_1$.

Put $W = \langle v_1 \rangle^{\perp}$, so $V = \langle v_1 \rangle \oplus W$, as V is an inner product space.

Claim. $\alpha(W) \subseteq W$; that is, $\langle x, v_1 \rangle$ implies $\big\langle \alpha(x), v_1 \big\rangle = 0$.

Proof. We have

$$\big\langle \alpha(x), v_1 \big\rangle = \Big\langle x, \alpha^{-1}(v_1) \Big\rangle = \Big\langle x, \lambda^{-1}(v_1) \Big\rangle = \overline{\lambda^{-1}} \langle x, v_1 \rangle = 0.$$

Also, $\big\langle \alpha(v), \alpha(w) \big\rangle = \langle v, w \rangle$ for all $v, w \in V$ implies that this is true for all $v, w \in W$, so $\alpha|_W$ is unitary; that is $(\alpha|_W)^* = (\alpha|_W)^{-1}$, so induction gives an orthonormal basis of W, namely $v_2, \dots, v_n$ of eigenvectors for α, and so $\hat{v}_1, v_2, \dots, v_n$ is an orthonormal basis for V. □

Remark. The previous two theorems admit the following generalisation: define $\alpha : V \to V$ to be *normal* if $\alpha\alpha^* = \alpha^*\alpha$; that is, if α and α^* commute.

Theorem 6.19

If α is normal, then there is an orthonormal basis consisting of eigenvalues for α.

Proof. Exercise! □

26 Nov

Recall that

(i) $\mathrm{GL}_n(\mathbb{C})$ acts on $\mathrm{Mat}_n(\mathbb{C})$ taking $(P, A) \mapsto PAP^{-1}$.

Interpretation: a choice of basis of a vector space V identifies $\mathrm{Mat}_n(\mathbb{C}) \cong \mathcal{L}(V, V)$, and a change of basis changes A to PAP^{-1}.

(ii) $\mathrm{GL}_n(\mathbb{C})$ acts on $\mathrm{Mat}_n(\mathbb{C})$ taking $(P, A) \mapsto PA\overline{P}^{\mathsf{T}}$.

Interpretation: a choice of basis of a vector space V identifies $\mathrm{Mat}_n(\mathbb{C})$ with sesquilinear forms.

A change of basis changes A to $PA\overline{P}^{\mathsf{T}}$ (where P is $\overline{Q}^{-1}$, if Q is the change of basis matrix).

These are genuinely different; that is, the theory of linear maps and sesquilinear forms are different.

But we have $P \in U_n$ if and only if $\overline{P}^{\mathsf{T}} P = I$, and $P^{-1} = \overline{P}^{\mathsf{T}}$, and then these two actions coincide! This occurs if and only if the columns of P are an orthonormal basis with respect to usual inner product on $\mathbb{C}^n$.

Proposition 6.20.

(i) Let $A \in \mathrm{Mat}_n(\mathbb{C})$ be Hermitian, so $\overline{A}^{\mathsf{T}} = A$. Then there exists a $P \in U_n$ such that $PAP^{-1} = PA\overline{P}^{\mathsf{T}}$ is real and diagonal.

(ii) Let $A \in \mathrm{Mat}_n(\mathbb{R})$ be symmetric, with $A^{\mathsf{T}} = A$. Then there exists a $P \in O_n(\mathbb{R})$ such that $PAP^{-1} = PAP^{\mathsf{T}}$ is real and diagonal.

Proof. Given $A \in \mathrm{Mat}_n(\mathbb{F})$ (for $\mathbb{F} = \mathbb{C}$ or $\mathbb{R}$), the map $\alpha : \mathbb{F}^n \to \mathbb{F}^n$ taking $x \mapsto Ax$ is self-adjoint with respect to the standard inner product. By theorem 6.17, there is an orthonormal basis of eigenvectors for $\alpha : \mathbb{F}^n \to \mathbb{F}^n$, that is, there are some $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ such that $Av_i = \lambda_i v_i$. Then

$$A \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} \lambda_1 v_1 \cdots \lambda_n v_n \end{pmatrix} = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

If we set $Q = \big(v_1 \cdots v_n\big) \in \mathrm{Mat}_n(\mathbb{F})$, then

$$AQ = Q \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},$$

and $v_1, \dots, v_n$ are an orthonormal basis if and only if $\overline{Q}^\mathsf{T} = Q^{-1}$, so we put $P = Q^{-1}$ and get the result. $\square$

Corollary 6.21. *If ψ is a Hermitian form on V with matrix A, then the signature* $\mathrm{sign}(\psi)$ *is the number of positive eigenvalues of A less the number of negative eigenvalues.*

Proof. If the matrix A is diagonal, then this is clear: rescale the basis vectors $v_i \mapsto v_i / |v_i|$, and the signature is the number of original diagonal entries which are positive, less the number which are negative.

Now for general A, the proposition shows that we can choose $P \in U_n$ such that $PAP^{-1} = PA\overline{P}^\mathsf{T}$ is diagonal, and this represents the same form with respect to the new basis, but also has the same eigenvalues. $\square$

Corollary 6.22. *Both* $\mathrm{rank}(\psi)$ *and* $\mathrm{sign}(\psi)$ *can be read off the characteristic polynomialomial of any matrix A for ψ.*

Exercise 6.23. Let $\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be $\psi(x, y) = x^\mathsf{T} A y$, where

$$A = \begin{pmatrix} 0 & 1 & \cdots & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & 1 & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix}.$$

Show that $\mathrm{ch}_A(x) = (x+1)^{n-1}\big(x - (n-1)\big)$, so the signature is $2-n$ and the rank is n.

Another consequence of the proposition is the simultaneous diagonalisation of some bilinear forms.

Theorem 6.24

Let V be a finite dimensional vector space over $\mathbb{C}$ (or $\mathbb{R}$), and $\varphi, \psi : V \times V \to \mathbb{F}$ be two Hermitian (symmetric) bilinear forms.

If φ is positive definite, then there is some basis $v_1, \dots, v_n$ of V such that with respect to this basis, both forms φ and ψ are diagonal; that is, $\psi(v_i, v_j) = \varphi(v_i, v_j) = 0$ if $i \neq j$.

Proof. As φ is positive definite, there exists an orthonormal basis for φ; that is, some $w_1, \dots, w_n$ such that $\varphi(w_i, w_j) = \delta_{ij}$.

Now let B be the matrix of ψ with respect to this basis; that is, $b_{ij} = \psi(w_i, w_j) = \overline{b}_{ji} = \psi(w_j, w_i)$, as ψ is Hermitian.

By the proposition, there is some $P \in U_n$ (or $O_n(\mathbb{R})$, if V is over $\mathbb{R}$) such that

$$\overline{P}^{\mathsf{T}} BP = D = \begin{pmatrix} \lambda_1 & & 0 \\ & \cdots & \\ 0 & & \lambda_n \end{pmatrix}$$

is diagonal, for $\lambda_i \in \mathbb{R}$, and now the matrix of φ with respect to our new basis, is $\overline{P}^{\mathsf{T}} IP = I$, also diagonal. □

Now we ask what is the "meaning" of the diagonal entries $\lambda_1, \dots, \lambda_n$?

If $\varphi, \psi : V \times V \to \mathbb{F}$ are any two bilinear/sesquilinear forms, then they determine (anti)-linear maps $V \to V^*$ taking $v \mapsto \varphi(\cdot, v)$ and $v \mapsto \psi(\cdot, v)$, and if φ is a non-degenerate form, then the map $V \to V^*$ taking $v \mapsto \varphi(\cdot, v)$ is an (anti)-linear *isomorphism*. So we can take its inverse, and compose with the map $V \to V^*$, $v \mapsto \psi(\cdot, v)$ to get a (linear!) map $V \to V$. Then $\lambda_1, \dots, \lambda_n$ are the eigenvalues of this map.

> **Exercise 6.25.** If φ, ψ are both *not* positive definite, then need they be simultaneously diagonalisable?

Remark. In coordinates: if we choose any basis for V, let the matrix of φ be A and that for ψ be B, with respect to this basis. Then $A = Q^{\mathsf{T}} Q$ for some $Q \in \mathrm{GL}_n(\mathbb{C})$ as φ is positive definite, and then the above proof shows that

$$B = \overline{Q}^{-\mathsf{T}} \overline{P}^{-\mathsf{T}} DP^{-1} Q^{-1},$$

since $P^{-1} = \overline{P}^{\mathsf{T}}$. Then

$$\begin{aligned} \det(D - xI) &= \det(Q^{-\mathsf{T}} \left(P^{-\mathsf{T}} DP - xQ^{\mathsf{T}} Q \right) Q^{-1}) \\ &= \det A \det(B - xA), \end{aligned}$$

and the diagonal entries are the roots of the polynomial $\det(B - xA)$; that is, the roots of $\det(BA^{-1} - xI)$, as claimed.

28 Nov

Consider the relationship between $O_n(\mathbb{R}) \hookrightarrow \mathrm{GL}_n(\mathbb{R})$, $U_n \hookrightarrow \mathrm{GL}_n(\mathbb{C})$.

> **Example 6.26.** Take $n = 1$. We have
>
> $$\mathrm{GL}_1(\mathbb{C}) = \mathbb{C}^* \qquad \text{and} \qquad U_1 = \left\{ \lambda \in \mathbb{C} : |\lambda| = 1 \right\} = S^1$$
>
> We have $\mathbb{C}^* = S^1 \times \mathbb{R}_{>0}$, with $\lambda r \leftarrow\!\shortmid (\lambda, r)$.
>
> In $\mathbb{R}$, we have $\mathrm{GL}_1(\mathbb{R}) = \mathbb{R}^*$, $O_1(\mathbb{R}) = \{\pm 1\}$ and $\mathbb{R}^* = \{\pm 1\} \times \mathbb{R}_{>0}$.
>
> For $n > 1$, Gram-Schmidt orthonormalisation tells us the relation: define
>
> $$\mathcal{A} = \left\{ \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \middle| \lambda_i \in \mathbb{R}_{>0} \right\}, \qquad N(\mathbb{F}) = \left\{ \begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \middle| * \in \mathbb{F} \right\},$$
>
> where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Then $\mathcal{A}$ as a set, is homeomorphic to $\mathbb{R}^n$, and $N(\mathbb{F})$ as a set (not a group) is isomorphic to $\mathbb{F}^{\frac{1}{2}(n-1)n}$, so $\mathbb{R}^{n(n-1)/2}$ or $\mathbb{C}^{n(n-1)/2}$.

Exercise 6.27. Show that

$$\mathcal{A} \cdot N(\mathbb{F}) = \left\{ \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \mid \lambda_i \in \mathbb{R}_{>0}, * \in \mathbb{F} \right\}$$

is a group, $N(\mathbb{F})$ is a normal subgroup and $A \cap N(\mathbb{F}) = \{I\}$.

Theorem 6.28

Any $A \in \mathrm{GL}_n(\mathbb{C})$ can be written uniquely as $A = QR$, with $Q \in U_n$, $R \in \mathcal{A} \cdot N(\mathbb{C})$.

Similarly, any $A \in \mathrm{GL}_n(\mathbb{R})$ can be written uniquely as $A = QR$, with $Q \in O_n(\mathbb{R})$, $R \in \mathcal{A} \cdot N(\mathbb{R})$.

Example 6.29. $n = 1$ is above.

Proof. This is just Gram-Schmidt.

Write $A = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$, $v_i \in \mathbb{F}^n$ so $v_1, \ldots, v_n$ is a basis for $\mathbb{F}^n$. Now the Gram-Schmidt algorithm gives an orthonormal basis $e_1, \ldots, e_n$. Recall how it went: set

$$\begin{aligned}
\tilde{e}_1 &= v_1, \\
\tilde{e}_2 &= v_2 - \frac{\langle v_2, \tilde{e}_1 \rangle}{\langle \tilde{e}_1, \tilde{e}_1 \rangle} \cdot \tilde{e}_1, \\
\tilde{e}_3 &= v_3 - \frac{\langle v_3, \tilde{e}_2 \rangle}{\langle \tilde{e}_2, \tilde{e}_2 \rangle} \cdot \tilde{e}_2 - \frac{\langle v_3, \tilde{e}_1 \rangle}{\langle \tilde{e}_1, \tilde{e}_1 \rangle} \cdot \tilde{e}_1 \\
&\vdots \\
\tilde{e}_n &= v_n - \sum_{i=1}^{n-1} \frac{\langle v_n, \tilde{e}_i \rangle}{\tilde{e}_i, \tilde{e}_i} \cdot \tilde{e}_i,
\end{aligned}$$

so that $\tilde{e}_1, \ldots, \tilde{e}_n$ are orthogonal, and if we set $e_i = \tilde{e}_i / |\tilde{e}_i|$, then $e_1, \ldots, e_n$ are an orthonormal basis. So

$$\begin{aligned}
\tilde{e}_i &= v_i + \text{correction terms} \\
&= v_i + \langle \tilde{e}_1, \ldots, \tilde{e}_{i-1} \rangle \\
&= v_i + \langle v_1, \ldots, v_{i-1} \rangle,
\end{aligned}$$

so we can write

$$\begin{aligned}
\tilde{e}_1 &= v_1, \\
\tilde{e}_2 &= v_2 + (*)\, v_1, \\
\tilde{e}_3 &= v_3 + (*)\, v_2 + (*)\, v_1,
\end{aligned}$$

that is,

$$\begin{pmatrix} \tilde{e}_1 & \cdots & \tilde{e}_n \end{pmatrix} = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix}, \text{ with } * \in \mathbb{F},$$

and

$$\begin{pmatrix}\tilde{e}_1 & \cdots & \tilde{e}_n\end{pmatrix}\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix} = \begin{pmatrix}e_1 & \cdots & e_n\end{pmatrix},$$

with $\lambda_i = 1/|\tilde{e}_i|$. So if $Q = \begin{pmatrix}e_1 & \cdots & e_n\end{pmatrix}$, this is

$$Q = A\underbrace{\begin{pmatrix}1 & & *\\ & \ddots & \\ & & 1\end{pmatrix}\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix}}_{\text{call this } R^{-1}}.$$

Thus $QR = A$, with $R \in \mathcal{A}\cdot N(\mathbb{F})$, and $e_1, \ldots, e_n$ is an orthonormal basis if and only if $Q \in U_n$; that is, if $\overline{Q}^{\mathsf{T}}Q = I$.

For uniqueness: if $QR = Q'R'$, then

$$\underbrace{(Q)^{-1}Q}_{\in U_n} = \underbrace{R'R^{-1}}_{\in \mathcal{A}\cdot N(\mathbb{F})}.$$

So it is enough to show that if $X = \begin{pmatrix}x_{ij}\end{pmatrix} \in \mathcal{A}\cdot N(\mathbb{F}) \cap U_n$, then $X = I$. But

$$X = \begin{pmatrix}x_{11} & & *\\ & \ddots & \\ 0 & & x_{nn}\end{pmatrix},$$

and both the columns and the rows are orthonormal bases since $X \in U_n$. Since the columns are an orthonormal basis, $|x_{11}| = 1$ implies $x_{12} = x_{13} = \cdots = x_{1n} = 0$, as $\sum_{i=1}^n |x_{1i}|^2 = 1$.

Then $x_{11} \in \mathbb{R}_{>0} \cap \{\lambda \in \mathbb{C} \mid |\lambda| = 1\}$ implies $x_{11} = 1$, so

$$X = \begin{pmatrix}1 & 0\\ 0 & \boxed{X'}\end{pmatrix},$$

with $X' \in U_{n-1} \cap \mathcal{A}\cdot N(\mathbb{F})$, so induction gives $X' = I$.

Warning. Notice that U_n is a group, $\mathcal{A}\cdot N(\mathbb{C})$ is a group, and if you want you can make $U_n \times \mathcal{A}\cdot N(\mathbb{C})$ into a group by the direct product. But if you do this, then the map in the theorem is *not* a group homomorphism.

The theorem says the map

$$\begin{array}{rrcl}\phi & : & U_n \times \mathcal{A}\cdot N(\mathbb{C}) & \longrightarrow & \mathrm{GL}_n(\mathbb{C})\\ & & (Q,R) & \longmapsto & QR\end{array}$$

is a *bijection* of sets, not an isomorphism of groups.

This theorem tells us that the 'shape' of the group $\mathrm{GL}_n(\mathbb{C})$ and the shape of the group U_n are the "same" – one differs from another by the product of a space of the form $\mathbb{C}^k$, a vector space. You will learn in topology the precise words for this – these two groups are *homotopic* – and you will learn later on that this means that many of their essential features are the same.

Finally (!!!), let's give another proof that every element of the unitary group is diagonalisable. We already know a very strong form of this. The following proof gives a weaker result, but gives it for a wider class of groups. It uses the same ideas as in in the above (probably cryptic) remark.

Consider the map

$$\begin{array}{cccc} \theta & : & \underset{=\mathbb{C}^{n^2}=\mathbb{R}^{2n^2}}{\mathrm{Mat}_n(\mathbb{C})} & \longrightarrow & \underset{=\mathbb{C}^{n^2}=\mathbb{R}^{2n^2}}{\mathrm{Mat}_n(\mathbb{C})} \\ & & A & \longmapsto & \overline{A}^{\mathsf{T}} A \end{array}.$$

This is a continuous map, and $\theta^{-1}(\{I\}) = U_n$, so as this is the inverse image of a closed set, it is a closed subset of $\mathbb{C}^{n^2}$. We also observe that $\sum_j \left|a_{ij}\right|^2 = 1$ implies $U_n \subseteq \left\{(a_{ij}) \mid \left|a_{ij}\right| \leq 1\right\}$ is a bounded set, so U_n is a closed bounded subset of $\mathbb{C}^{n^2}$. Thus U_n is a *compact topological space*, and a group (a *compact group*). $\square$

Proposition 6.30. *Let $G \leq \mathrm{GL}_n(\mathbb{C})$ be a subgroup such that G is also a closed bounded subset, that is, a* compact subgroup of $\mathrm{GL}_n(\mathbb{C})$. *Then if $g \in G$, then g is diagonalisable as an element of* $\mathrm{GL}_n(\mathbb{C})$. *That is, there is some $P \in \mathrm{GL}_n(\mathbb{C})$ such that PgP^{-1} is diagonal.*

> **Example 6.31.** Any $g \in U_n$ is diagonalisable.

Proof. Consider the sequence of elements $1, g, g^2, g^3, \ldots$ in G. As G is a closed bounded subset, it must have a convergent subsequence.

Let $P \in \mathrm{GL}_n(\mathbb{C})$ such that PgP^{-1} is in JNF.

Claim. The sequence $a_1, a_2, \ldots, a_n$ in GL_n converges if and only if $Pa_1P^{-1}, Pa_2P^{-1}, \ldots$ converges.

Proof of claim. For fixed P, the map $A \mapsto PAP^{-1}$ is a continuous map on $\mathbb{C}^{n^2}$. This implies the claim, as the matrix coefficients are linear functions of the matrix coefficients on A.

If PgP^{-1} has a Jordan block of size $a > 1$,

$$\begin{pmatrix} \lambda & 1 & 0 \\ & \ddots & 1 \\ 0 & & \lambda \end{pmatrix} = (\lambda I + J_a)\,,\ \lambda \neq 0,$$

then

$$\begin{aligned} (\lambda I + J_a)^N &= \lambda^n I + N\lambda^{N-1} J_a + \binom{N}{2} \lambda^{N-2} J_a^2 + \cdots \\ &= \begin{pmatrix} \lambda^N & N\lambda^{N-1} & & \\ & \ddots & & \\ & & \ddots & N\lambda^{N-1} \\ & & & \lambda^N \end{pmatrix}. \end{aligned}$$

If $|\lambda| > 1$, this has unbounded coefficients on the diagonal as $N \to \infty$; if $|\lambda| < 1$, this has unbounded coefficients on the diagonal as $N \to -\infty$, contradicting the existance of a convergent subsequence.

So it must be that $|\lambda| = 1$. But now examine the entries just above the diagonal, and observe these are unbounded as $N \to \infty$, contradicting the existance of a convergent subsequence. $\square$