Part IB — Linear Algebra

Theorems with proof

Based on lectures by S. J. Wadsley

Notes taken by Dexter Chua

Michaelmas 2015

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

0 Introduction

1 Vector spaces

1.1 Definitions and examples

Proposition. In any vector space V, $0\mathbf{v} = \mathbf{0}$ for all $\mathbf{v} \in V$, and $(-1)\mathbf{v} = -\mathbf{v}$, where $-\mathbf{v}$ is the additive inverse of $\mathbf{v}$.

Proposition. Let U, W be subspaces of V. Then $U + W$ and $U \cap W$ are subspaces.

Proof. Let $\mathbf{u}_i + \mathbf{w}_i \in U + W$, $\lambda, \mu \in \mathbb{F}$. Then

$$\lambda(\mathbf{u}_1 + \mathbf{w}_1) + \mu(\mathbf{u}_2 + \mathbf{w}_2) = (\lambda\mathbf{u}_1 + \mu\mathbf{u}_2) + (\lambda\mathbf{w}_1 + \mu\mathbf{w}_2) \in U + W.$$

Similarly, if $\mathbf{v}_i \in U \cap W$, then $\lambda\mathbf{v}_1 + \mu\mathbf{v}_2 \in U$ and $\lambda\mathbf{v}_1 + \mu\mathbf{v}_2 \in W$. So $\lambda\mathbf{v}_1 + \mu\mathbf{v}_2 \in U \cap W$.

Both $U \cap W$ and $U + W$ contain $\mathbf{0}$, and are non-empty. So done. □

1.2 Linear independence, bases and the Steinitz exchange lemma

Lemma. $S \subseteq V$ is linearly dependent if and only if there are distinct $\mathbf{s}_0, \cdots, \mathbf{s}_n \in S$ and $\lambda_1, \cdots, \lambda_n \in \mathbb{F}$ such that

$$\sum_{i=1}^{n} \lambda_i \mathbf{s}_i = \mathbf{s}_0.$$

Proof. If S is linearly dependent, then there are some $\lambda_1, \cdots, \lambda_n \in \mathbb{F}$ not all zero and $\mathbf{s}_1, \cdots, \mathbf{s}_n \in S$ such that $\sum \lambda_i \mathbf{s}_i = \mathbf{0}$. Wlog, let $\lambda_1 \neq 0$. Then

$$\mathbf{s}_1 = \sum_{i=2}^{n} -\frac{\lambda_i}{\lambda_1}\mathbf{s}_i.$$

Conversely, if $\mathbf{s}_0 = \sum_{i=1}^{n} \lambda_i \mathbf{s}_i$, then

$$(-1)\mathbf{s}_0 + \sum_{i=1}^{n} \lambda_i \mathbf{s}_i = \mathbf{0}.$$

So S is linearly dependent. □

Proposition. If $S = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is a subset of V over $\mathbb{F}$, then it is a basis if and only if every $\mathbf{v} \in V$ can be written uniquely as a finite linear combination of elements in S, i.e. as

$$\mathbf{v} = \sum_{i=1}^{n} \lambda_i \mathbf{e}_i.$$

Proof. We can view this as a combination of two statements: it can be spanned in at least one way, and it can be spanned in at most one way. We will see that the first part corresponds to S spanning V, and the second part corresponds to S being linearly independent.

In fact, S spanning V is defined exactly to mean that every item $\mathbf{v} \in V$ can be written as a finite linear combination in at least one way.

Now suppose that S is linearly independent, and we have

$$\mathbf{v} = \sum_{i=1}^{n} \lambda_i \mathbf{e}_i = \sum_{i=1}^{n} \mu_i \mathbf{e}_i.$$

Then we have

$$\mathbf{0} = \mathbf{v} - \mathbf{v} = \sum_{i=1}^{n} (\lambda_i - \mu_i)\mathbf{e}_i.$$

Linear independence implies that $\lambda_i - \mu_i = 0$ for all i. Hence $\lambda_i = \mu_i$. So $\mathbf{v}$ can be expressed in a unique way.

On the other hand, if S is not linearly independent, then we have

$$\mathbf{0} = \sum_{i=1}^{n} \lambda_i \mathbf{e}_i$$

where $\lambda_i \neq 0$ for some i. But we also know that

$$\mathbf{0} = \sum_{i=1}^{n} 0 \cdot \mathbf{e}_i.$$

So there are two ways to write $\mathbf{0}$ as a linear combination. So done. □

Theorem (Steinitz exchange lemma)**.** Let V be a vector space over $\mathbb{F}$, and $S = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ a finite linearly independent subset of V, and T a spanning subset of V. Then there is some $T' \subseteq T$ of order n such that $(T \setminus T') \cup S$ still spans V. In particular, $|T| \geq n$.

Corollary. Let $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ be a linearly independent subset of V, and suppose $\{\mathbf{f}_1, \cdots, \mathbf{f}_m\}$ spans V. Then there is a re-ordering of the $\{\mathbf{f}_i\}$ such that $\{\mathbf{e}_1, \cdots, \mathbf{e}_n, \mathbf{f}_{n+1}, \cdots, \mathbf{f}_m\}$ spans V.

Proof. Suppose that we have already found $T'_r \subseteq T$ of order $0 \leq r < n$ such that

$$T_r = (T \setminus T'_r) \cup \{\mathbf{e}_1, \cdots, \mathbf{e}_r\}$$

spans V.

(Note that the case $r = 0$ is trivial, since we can take $T'_r = \emptyset$, and the case $r = n$ is the theorem which we want to achieve.)

Suppose we have these. Since T_r spans V, we can write

$$\mathbf{e}_{r+1} = \sum_{i=1}^{k} \lambda_i \mathbf{t}_i, \quad \lambda_i \in \mathbb{F}, \mathbf{t}_i \in T_r.$$

We know that the $\mathbf{e}_i$ are linearly independent, so not all $\mathbf{t}_i$'s are $\mathbf{e}_i$'s. So there is some j such that $\mathbf{t}_j \in (T \setminus T'_r)$. We can write this as

$$\mathbf{t}_j = \frac{1}{\lambda_j}\mathbf{e}_{r+1} + \sum_{i \neq j} -\frac{\lambda_i}{\lambda_j}\mathbf{t}_i.$$

We let $T'_{r+1} = T'_r \cup \{\mathbf{t}_j\}$ of order $r + 1$, and

$$T_{r+1} = (T \setminus T'_{r+1}) \cup \{\mathbf{e}_1, \cdots, \mathbf{e}_{r+1}\} = (T_r \setminus \{\mathbf{t}_j\}) \cup \{\mathbf{e}_{r+1}\}$$

Since $\mathbf{t}_j$ is in the span of $T_r \cup \{\mathbf{e}_{r+1}\}$, we have $\mathbf{t}_j \in \langle T_{r+1}\rangle$. So

$$V \supseteq \langle T_{r+1}\rangle \supseteq \langle T_r\rangle = V.$$

So $\langle T_{r+1}\rangle = V$.
Hence we can inductively find T_n. □

Corollary. Suppose V is a vector space over $\mathbb{F}$ with a basis of order n. Then

(i) Every basis of V has order n.

(ii) Any linearly independent set of order n is a basis.

(iii) Every spanning set of order n is a basis.

(iv) Every finite spanning set contains a basis.

(v) Every linearly independent subset of V can be extended to basis.

Proof. Let $S = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ be the basis for V.

(i) Suppose T is another basis. Since S is independent and T is spanning, $|T| \geq |S|$.

The other direction is less trivial, since T might be infinite, and Steinitz does not immediately apply. Instead, we argue as follows: since T is linearly independent, every finite subset of T is independent. Also, S is spanning. So every finite subset of T has order at most $|S|$. So $|T| \leq |S|$. So $|T| = |S|$.

(ii) Suppose now that T is a linearly independent subset of order n, but $\langle T\rangle \neq V$. Then there is some $\mathbf{v} \in V \setminus \langle T\rangle$. We now show that $T \cup \{\mathbf{v}\}$ is independent. Indeed, if

$$\lambda_0\mathbf{v} + \sum_{i=1}^{m} \lambda_i\mathbf{t}_i = \mathbf{0}$$

with $\lambda_i \in \mathbb{F}$, $\mathbf{t}_1, \cdots, \mathbf{t}_m \in T$ distinct, then

$$\lambda_0\mathbf{v} = \sum_{i=1}^{m} (-\lambda_i)\mathbf{t}_i.$$

Then $\lambda_0\mathbf{v} \in \langle T\rangle$. So $\lambda_0 = 0$. As T is linearly independent, we have $\lambda_0 = \cdots = \lambda_m = 0$. So $T \cup \{\mathbf{v}\}$ is a linearly independent subset of size $> n$. This is a contradiction since S is a spanning set of size n.

(iii) Let T be a spanning set of order n. If T were linearly dependent, then there is some $\mathbf{t}_0, \cdots, \mathbf{t}_m \in T$ distinct and $\lambda_1, \cdots, \lambda_m \in \mathbb{F}$ such that

$$\mathbf{t}_0 = \sum \lambda_i\mathbf{t}_i.$$

So $\mathbf{t}_0 \in \langle T \setminus \{\mathbf{t}_0\}\rangle$, i.e. $\langle T \setminus \{\mathbf{t}_0\}\rangle = V$. So $T \setminus \{\mathbf{t}_0\}$ is a spanning set of order $n - 1$, which is a contradiction.

(iv) Suppose T is any finite spanning set. Let $T' \subseteq T$ be a spanning set of least possible size. This exists because T is finite. If $|T'|$ has size n, then done by (iii). Otherwise by the Steinitz exchange lemma, it has size $|T'| > n$. So T' must be linearly dependent because S is spanning. So there is some $\mathbf{t}_0, \cdots, \mathbf{t}_m \in T$ distinct and $\lambda_1, \cdots, \lambda_m \in \mathbb{F}$ such that $\mathbf{t}_0 = \sum \lambda_i \mathbf{t}_i$. Then $T' \setminus \{\mathbf{t}_0\}$ is a smaller spanning set. Contradiction.

(v) Suppose T is a linearly independent set. Since S spans, there is some $S' \subseteq S$ of order $|T|$ such that $(S \setminus S') \cup T$ spans V by the Steinitz exchange lemma. So by (ii), $(S \setminus S') \cup T$ is a basis of V containing T. □

Lemma. If V is a finite dimensional vector space over $\mathbb{F}$, $U \subseteq V$ is a proper subspace, then U is finite dimensional and $\dim U < \dim V$.

Proof. Every linearly independent subset of V has size at most $\dim V$. So let $S \subseteq U$ be a linearly independent subset of largest size. We want to show that S spans U and $|S| < \dim V$.

If $\mathbf{v} \in V \setminus \langle S \rangle$, then $S \cup \{\mathbf{v}\}$ is linearly independent. So $\mathbf{v} \notin U$ by maximality of S. This means that $\langle S \rangle = U$.

Since $U \neq V$, there is some $\mathbf{v} \in V \setminus U = V \setminus \langle S \rangle$. So $S \cup \{\mathbf{v}\}$ is a linearly independent subset of order $|S| + 1$. So $|S| + 1 \leq \dim V$. In particular, $\dim U = |S| < \dim V$. □

Proposition. If U, W are subspaces of a finite dimensional vector space V, then

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

Proof. Let $R = \{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ be a basis for $U \cap W$. This is a linearly independent subset of U. So we can extend it to be a basis of U by

$$S = \{\mathbf{v}_1, \cdots, \mathbf{v}_r, \mathbf{u}_{r+1}, \cdots, \mathbf{u}_s\}.$$

Similarly, for W, we can obtain a basis

$$T = \{\mathbf{v}_1, \cdots, \mathbf{v}_r, \mathbf{w}_{r+1}, \cdots, \mathbf{w}_t\}.$$

We want to show that $\dim(U + W) = s + t - r$. It is sufficient to prove that $S \cup T$ is a basis for $U + W$.

We first show spanning. Suppose $\mathbf{u} + \mathbf{w} \in U + W$, $\mathbf{u} \in U, \mathbf{w} \in W$. Then $\mathbf{u} \in \langle S \rangle$ and $\mathbf{w} \in \langle T \rangle$. So $\mathbf{u} + \mathbf{w} \in \langle S \cup T \rangle$. So $U + W = \langle S \cup T \rangle$.

To show linear independence, suppose we have a linear relation

$$\sum_{i=1}^{r} \lambda_i \mathbf{v}_i + \sum_{j=r+1}^{s} \mu_j \mathbf{u}_j + \sum_{k=r+1}^{t} \nu_k \mathbf{w}_k = \mathbf{0}.$$

So

$$\sum \lambda_i \mathbf{v}_i + \sum \mu_j \mathbf{u}_j = -\sum \nu_k \mathbf{w}_k.$$

Since the left hand side is something in U, and the right hand side is something in W, they both lie in $U \cap W$.

Since S is a basis of U, there is only one way of writing the left hand vector as a sum of $\mathbf{v}_i$ and $\mathbf{u}_j$. However, since R is a basis of $U \cap W$, we can write the

left hand vector just as a sum of $\mathbf{v}_i$'s. So we must have $\mu_j = 0$ for all j. Then we have

$$\sum \lambda_i \mathbf{v}_i + \sum \nu_k \mathbf{w}_k = \mathbf{0}.$$

Finally, since T is linearly independent, $\lambda_i = \nu_k = 0$ for all i, k. So $S \cup T$ is linearly independent. □

Proposition. If V is a finite dimensional vector space over $\mathbb{F}$ and $U \cup V$ is a subspace, then

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

Proof. Let $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ be a basis for U and extend this to a basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_m, \mathbf{v}_{m+1}, \cdots, \mathbf{v}_n\}$ for V. We want to show that $\{\mathbf{v}_{m+1} + U, \cdots, \mathbf{v}_n + U\}$ is a basis for V/U.

It is easy to see that this spans V/U. If $\mathbf{v} + U \in V/U$, then we can write

$$\mathbf{v} = \sum \lambda_i \mathbf{u}_i + \sum \mu_i \mathbf{v}_i.$$

Then

$$\mathbf{v} + U = \sum \mu_i (\mathbf{v}_i + U) + \sum \lambda_i (\mathbf{u}_i + U) = \sum \mu_i (\mathbf{v}_i + U).$$

So done.

To show that they are linearly independent, suppose that

$$\sum \lambda_i (\mathbf{v}_i + U) = \mathbf{0} + U = U.$$

Then this requires

$$\sum \lambda_i \mathbf{v}_i \in U.$$

Then we can write this as a linear combination of the $\mathbf{u}_i$'s. So

$$\sum \lambda_i \mathbf{v}_i = \sum \mu_j \mathbf{u}_j$$

for some μ_j. Since $\{\mathbf{u}_1, \cdots, \mathbf{u}_m, \mathbf{v}_{n+1}, \cdots, \mathbf{v}_n\}$ is a basis for V, we must have $\lambda_i = \mu_j = 0$ for all i, j. So $\{\mathbf{v}_i + U\}$ is linearly independent. □

1.3 Direct sums

2 Linear maps

2.1 Definitions and examples

Lemma. If U and V are vector spaces over $\mathbb{F}$ and $\alpha : U \to V$, then α is an isomorphism iff α is a bijective linear map.

Proof. If α is an isomorphism, then it is clearly bijective since it has an inverse function.

Suppose α is a linear bijection. Then as a function, it has an inverse $\beta : V \to U$. We want to show that this is linear. Let $\mathbf{v}_1, \mathbf{v}_2 \in V$, $\lambda, \mu \in \mathbb{F}$. We have

$$\alpha\beta(\lambda\mathbf{v}_1 + \mu\mathbf{v}_2) = \lambda\mathbf{v}_1 + \mu\mathbf{v}_2 = \lambda\alpha\beta(\mathbf{v}_1) + \mu\alpha\beta(\mathbf{v}_2) = \alpha(\lambda\beta(\mathbf{v}_1) + \mu\beta(\mathbf{v}_2)).$$

Since α is injective, we have

$$\beta(\lambda\mathbf{v}_1 + \mu\mathbf{v}_2) = \lambda\beta(\mathbf{v}_1) + \mu\beta(\mathbf{v}_2).$$

So β is linear. □

Proposition. Let $\alpha : U \to V$ be an $\mathbb{F}$-linear map. Then

(i) If α is injective and $S \subseteq U$ is linearly independent, then $\alpha(S)$ is linearly independent in V.

(ii) If α is surjective and $S \subseteq U$ spans U, then $\alpha(S)$ spans V.

(iii) If α is an isomorphism and $S \subseteq U$ is a basis, then $\alpha(S)$ is a basis for V.

Proof.

(i) We prove the contrapositive. Suppose that α is injective and $\alpha(S)$ is linearly dependent. So there are $\mathbf{s}_0, \cdots, \mathbf{s}_n \in S$ distinct and $\lambda_1, \cdots, \lambda_n \in \mathbb{F}$ not all zero such that

$$\alpha(\mathbf{s}_0) = \sum_{i=1}^{n} \lambda_i \alpha(\mathbf{s}_i) = \alpha\left(\sum_{i=1}^{n} \lambda_i \mathbf{s}_i\right).$$

Since α is injective, we must have

$$\mathbf{s}_0 = \sum_{i=1}^{n} \lambda_i \mathbf{s}_i.$$

This is a non-trivial relation of the $\mathbf{s}_i$ in U. So S is linearly dependent.

(ii) Suppose α is surjective and S spans U. Pick $\mathbf{v} \in V$. Then there is some $\mathbf{u} \in U$ such that $\alpha(\mathbf{u}) = \mathbf{v}$. Since S spans U, there is some $\mathbf{s}_1, \cdots, \mathbf{s}_n \in S$ and $\lambda_1, \cdots, \lambda_n \in \mathbb{F}$ such that

$$\mathbf{u} = \sum_{I=1}^{n} \lambda_i \mathbf{s}_i.$$

Then

$$\mathbf{v} = \alpha(\mathbf{u}) = \sum_{i=1}^{n} \lambda_i \alpha(\mathbf{s}_i).$$

So $\alpha(S)$ spans V.

(iii) Follows immediately from (i) and (ii). □

Corollary. If U and V are finite-dimensional vector spaces over $\mathbb{F}$ and $\alpha : U \to V$ is an isomorphism, then $\dim U = \dim V$.

Proof. Let S be a basis for U. Then $\alpha(S)$ is a basis for V. Since α is injective, $|S| = |\alpha(S)|$. So done. □

Proposition. Suppose V is a $\mathbb{F}$-vector space of dimension $n < \infty$. Then writing $\mathbf{e}_1, \cdots, \mathbf{e}_n$ for the standard basis of $\mathbb{F}^n$, there is a bijection

$$\Phi : \{\text{isomorphisms } \mathbb{F}^n \to V\} \to \{\text{(ordered) basis}(\mathbf{v}_1, \cdots, \mathbf{v}_n) \text{ for } V\},$$

defined by

$$\alpha \mapsto (\alpha(\mathbf{e}_1), \cdots, \alpha(\mathbf{e}_n)).$$

Proof. We first make sure this is indeed a function — if α is an isomorphism, then from our previous proposition, we know that it sends a basis to a basis. So $(\alpha(\mathbf{e}_1), \cdots, \alpha(\mathbf{e}_n))$ is indeed a basis for V.

We now have to prove surjectivity and injectivity.

Suppose $\alpha, \beta : \mathbb{F}^n \to V$ are isomorphism such that $\Phi(\alpha) = \Phi(\beta)$. In other words, $\alpha(\mathbf{e}_i) = \beta(\mathbf{e}_i)$ for all i. We want to show that $\alpha = \beta$. We have

$$\alpha\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) = \alpha\left(\sum_{i=1}^n x_i \mathbf{e}_i\right) = \sum x_i \alpha(\mathbf{e}_i) = \sum x_i \beta(\mathbf{e}_i) = \beta\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right).$$

Hence $\alpha = \beta$.

Next, suppose that $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ is an ordered basis for V. Then define

$$\alpha\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) = \sum x_i \mathbf{v}_i.$$

It is easy to check that this is well-defined and linear. We also know that α is injective since $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ is linearly independent. So if $\sum x_i \mathbf{v}_i = \sum y_i \mathbf{v}_i$, then $x_i = y_i$. Also, α is surjective since $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ spans V. So α is an isomorphism, and by construction $\Phi(\alpha) = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$. □

2.2 Linear maps and matrices

Proposition. Suppose U, V are vector spaces over $\mathbb{F}$ and $S = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is a basis for U. Then every function $f : S \to V$ extends uniquely to a linear map $U \to V$.

Proof. For uniqueness, first suppose $\alpha, \beta : U \to V$ are linear and extend $f : S \to V$. We have sort-of proved this already just now.

If $\mathbf{u} \in U$, we can write $\mathbf{u} = \sum_{i=1}^n u_i \mathbf{e}_i$ with $u_i \in \mathbb{F}$ since S spans. Then

$$\alpha(\mathbf{u}) = \alpha\left(\sum u_i \mathbf{e}_i\right) = \sum u_i \alpha(\mathbf{e}_i) = \sum u_i f(\mathbf{e}_i).$$

Similarly,

$$\beta(\mathbf{u}) = \sum u_i f(\mathbf{e}_i).$$

So $\alpha(\mathbf{u}) = \beta(\mathbf{u})$ for every $\mathbf{u}$. So $\alpha = \beta$.

For existence, if $\mathbf{u} \in U$, we can write $\mathbf{u} = \sum u_i \mathbf{e}_i$ in a unique way. So defining

$$\alpha(\mathbf{u}) = \sum u_i f(\mathbf{e}_i)$$

is unambiguous. To show linearity, let $\lambda, \mu \in \mathbb{F}$, $\mathbf{u}, \mathbf{v} \in U$. Then

$$\begin{aligned}
\alpha(\lambda\mathbf{u} + \mu\mathbf{v}) &= \alpha\left(\sum(\lambda u_i + \mu v_i)\mathbf{e}_i\right) \\
&= \sum(\lambda u_i + \mu v_i) f(\mathbf{e}_i) \\
&= \lambda\left(\sum u_i f(\mathbf{e}_i)\right) + \mu\left(\sum v_i f(\mathbf{e}_i)\right) \\
&= \lambda\alpha(\mathbf{u}) + \mu\alpha(\mathbf{v}).
\end{aligned}$$

Moreover, α does extend f. □

Corollary. If U and V are finite-dimensional vector spaces over $\mathbb{F}$ with bases $(\mathbf{e}_1, \cdots, \mathbf{e}_m)$ and $(\mathbf{f}_1, \cdots, \mathbf{f}_n)$ respectively, then there is a bijection

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

sending A to the unique linear map α such that $\alpha(\mathbf{e}_i) = \sum a_{ji}\mathbf{f}_j$.

Proof. If α is a linear map $U \to V$, then for each $1 \leq i \leq m$, we can write $\alpha(\mathbf{e}_i)$ uniquely as

$$\alpha(\mathbf{e}_i) = \sum_{j=1}^{n} a_{ji}\mathbf{f}_j$$

for some $a_{ji} \in \mathbb{F}$. This gives a matrix $A = (a_{ij})$. The previous proposition tells us that every matrix A arises in this way, and α is determined by A. □

Proposition. Suppose U, V, W are finite-dimensional vector spaces over $\mathbb{F}$ with bases $R = (\mathbf{u}_1, \cdots, \mathbf{u}_r)$, $S = (\mathbf{v}_1, .., \mathbf{v}_s)$ and $T = (\mathbf{w}_1, \cdots, \mathbf{w}_t)$ respectively.

If $\alpha : U \to V$ and $\beta : V \to W$ are linear maps represented by A and B respectively (with respect to R, S and T), then $\beta\alpha$ is linear and represented by BA with respect to R and T.

Proof. Verifying $\beta\alpha$ is linear is straightforward. Next we write $\beta\alpha(\mathbf{u}_i)$ as a linear combination of $\mathbf{w}_1, \cdots, \mathbf{w}_t$:

$$\begin{aligned}
\beta\alpha(\mathbf{u}_i) &= \beta\left(\sum_k A_{ki}\mathbf{v}_k\right) \\
&= \sum_k A_{ki}\beta(\mathbf{v}_k) \\
&= \sum_k A_{ki}\sum_j B_{jk}\mathbf{w}_j \\
&= \sum_j \left(\sum_k B_{jk}A_{ki}\right)\mathbf{w}_j \\
&= \sum_j (BA)_{ji}\mathbf{w}_j
\end{aligned}$$

□

2.3 The first isomorphism theorem and the rank-nullity theorem

Theorem (First isomorphism theorem)**.** Let $\alpha : U \to V$ be a linear map. Then $\ker \alpha$ and $\operatorname{im} \alpha$ are subspaces of U and V respectively. Moreover, α induces an isomorphism

$$\bar{\alpha} : U/\ker \alpha \to \operatorname{im} \alpha$$
$$(\mathbf{u} + \ker \alpha) \mapsto \alpha(\mathbf{u})$$

Proof. We know that $\mathbf{0} \in \ker \alpha$ and $\mathbf{0} \in \operatorname{im} \alpha$.

Suppose $\mathbf{u}_1, \mathbf{u}_2 \in \ker \alpha$ and $\lambda_1, \lambda_2 \in \mathbb{F}$. Then

$$\alpha(\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2) = \lambda_1 \alpha(\mathbf{u}_1) + \lambda_2 \alpha(\mathbf{u}_2) = \mathbf{0}.$$

So $\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 \in \ker \alpha$. So $\ker \alpha$ is a subspace.

Similarly, if $\alpha(\mathbf{u}_1), \alpha(\mathbf{u}_2) \in \operatorname{im} \alpha$, then $\lambda\alpha(\mathbf{u}_1) + \lambda_2\alpha(\mathbf{u}_2) = \alpha(\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2) \in \operatorname{im} \alpha$. So $\operatorname{im} \alpha$ is a subspace.

Now by the first isomorphism theorem of groups, $\bar{\alpha}$ is a well-defined isomorphism of groups. So it remains to show that $\bar{\alpha}$ is a linear map. Indeed, we have

$$\bar{\alpha}(\lambda(\mathbf{u} + \ker \alpha)) = \alpha(\lambda \mathbf{u}) = \lambda \alpha(\mathbf{u}) = \lambda(\bar{\alpha}(\mathbf{u} + \ker \alpha)).$$

So $\bar{\alpha}$ is a linear map. □

Corollary (Rank-nullity theorem)**.** If $\alpha : U \to V$ is a linear map and U is finite-dimensional, then

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

Proof. By the first isomorphism theorem, we know that $U/\ker \alpha \cong \operatorname{im} \alpha$. So we have

$$\dim \operatorname{im} \alpha = \dim(U/\ker \alpha) = \dim U - \dim \ker \alpha.$$

So the result follows. □

Proposition. If $\alpha : U \to V$ is a linear map between finite-dimensional vector spaces over $\mathbb{F}$, then there are bases $(\mathbf{e}_1, \cdots, \mathbf{e}_m)$ for U and $(\mathbf{f}_1, \cdots, \mathbf{f}_n)$ for V such that α is represented by the matrix

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

where $r = r(\alpha)$ and I_r is the $r \times r$ identity matrix.

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

Proof. Let $\mathbf{e}_{k+1}, \cdots, \mathbf{e}_m$ be a basis for the kernel of α. Then we can extend this to a basis of the $(\mathbf{e}_1, \cdots, \mathbf{e}_m)$.

Let $\mathbf{f}_i = \alpha(\mathbf{e}_i)$ for $1 \leq i \leq k$. We now show that $(\mathbf{f}_1, \cdots, \mathbf{f}_k)$ is a basis for $\operatorname{im} \alpha$ (and thus $k = r$). We first show that it spans. Suppose $\mathbf{v} \in \operatorname{im} \alpha$. Then we have

$$\mathbf{v} = \alpha\left(\sum_{i=1}^{m} \lambda_i \mathbf{e}_i\right)$$

for some $\lambda_i \in \mathbb{F}$. By linearity, we can write this as

$$\mathbf{v} = \sum_{i=1}^{m} \lambda_i \alpha(\mathbf{e}_i) = \sum_{i=1}^{k} \lambda_i \mathbf{f}_i + \mathbf{0}.$$

So $\mathbf{v} \in \langle \mathbf{f}_1, \cdots, \mathbf{f}_k \rangle$.

To show linear dependence, suppose that

$$\sum_{i=1}^{k} \mu_i \mathbf{f}_i = \mathbf{0}.$$

So we have

$$\alpha\left(\sum_{i=1}^{k} \mu_i \mathbf{e}_i\right) = \mathbf{0}.$$

So $\sum_{i=1}^{k} \mu_i \mathbf{e}_i \in \ker \alpha$. Since $(\mathbf{e}_{k+1}, \cdots, \mathbf{e}_m)$ is a basis for $\ker \alpha$, we can write

$$\sum_{i=1}^{k} \mu_i \mathbf{e}_i = \sum_{i=k+1}^{m} \mu_i \mathbf{e}_i$$

for some μ_i $(i = k+1, \cdots, m)$. Since $(\mathbf{e}_1, \cdots, \mathbf{e}_m)$ is a basis, we must have $\mu_i = 0$ for all i. So they are linearly independent.

Now we extend $(\mathbf{f}_1, \cdots, \mathbf{f}_r)$ to a basis for V, and

$$\alpha(\mathbf{e}_i) = \begin{cases} \mathbf{f}_i & 1 \leq i \leq k \\ 0 & k+1 \leq i \leq m \end{cases}.$$ □

Corollary. Suppose $\alpha : U \to V$ is a linear map between vector spaces over $\mathbb{F}$ both of dimension $n < \infty$. Then the following are equivalent

(i) α is injective;

(ii) α is surjective;

(iii) α is an isomorphism.

Proof. It is clear that, (iii) implies (i) and (ii), and (i) and (ii) together implies (iii). So it suffices to show that (i) and (ii) are equivalent.

Note that α is injective iff $n(\alpha) = 0$, and α is surjective iff $r(\alpha) = \dim V = n$. By the rank-nullity theorem, $n(\alpha) + r(\alpha) = n$. So the result follows immediately. □

Lemma. Let $A \in M_{n,n}(\mathbb{F}) = M_n(\mathbb{F})$ be a square matrix. The following are equivalent

(i) There exists $B \in M_n(\mathbb{F})$ such that $BA = I_n$.

(ii) There exists $C \in M_n(\mathbb{F})$ such that $AC = I_n$.

If these hold, then $B = C$. We call A *invertible* or *non-singular*, and write $A^{-1} = B = C$.

Proof. Let $\alpha, \beta, \gamma, \iota : \mathbb{F}^n \to \mathbb{F}^n$ be the linear maps represented by matrices A, B, C, I_n respectively with respect to the standard basis.

We note that (i) is equivalent to saying that there exists β such that $\beta\alpha = \iota$. This is true iff α is injective, which is true iff α is an isomorphism, which is true iff α has an inverse α^{-1}.

Similarly, (ii) is equivalent to saying that there exists γ such that $\alpha\gamma = \iota$. This is true iff α is injective, which is true iff α is isomorphism, which is true iff α has an inverse α^{-1}.

So these are the same things, and we have $\beta = \alpha^{-1} = \gamma$. □

2.4 Change of basis

Theorem. Suppose that $(\mathbf{e}_1, \cdots, \mathbf{e}_m)$ and $(\mathbf{u}_1, \cdots, \mathbf{u}_m)$ are basis for a finite-dimensional vector space U over $\mathbb{F}$, and $(\mathbf{f}_1, \cdots, \mathbf{f}_n)$ and $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ are basis of a finite-dimensional vector space V over $\mathbb{F}$.

Let $\alpha : U \to V$ be a linear map represented by a matrix A with respect to $(\mathbf{e}_i)$ and $(\mathbf{f}_i)$ and by B with respect to $(\mathbf{u}_i)$ and $(\mathbf{v}_i)$. Then

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

where P and Q are given by

$$\mathbf{u}_i = \sum_{k=1}^{m} P_{ki}\mathbf{e}_k, \quad \mathbf{v}_i = \sum_{k=1}^{n} Q_{ki}\mathbf{f}_k.$$

Proof. On the one hand, we have

$$\alpha(\mathbf{u}_i) = \sum_{j=1}^{n} B_{ji}\mathbf{v}_j = \sum_j \sum_\ell B_{ji}Q_{\ell j}\mathbf{f}_\ell = \sum_\ell [QB]_{\ell i}\mathbf{f}_\ell.$$

On the other hand, we can write

$$\alpha(\mathbf{u}_i) = \alpha\left(\sum_{k=1}^{m} P_{ki}\mathbf{e}_k\right) = \sum_{k=1}^{m} P_{ki}\sum_\ell A_{\ell k}\mathbf{f}_\ell = \sum_\ell [AP]_{\ell i} f_\ell.$$

Since the $\mathbf{f}_\ell$ are linearly independent, we conclude that

$$QB = AP.$$

Since Q is invertible, we get $B = Q^{-1}AP$. □

Corollary. If $A \in \mathrm{Mat}_{n,m}(\mathbb{F})$, then there exists invertible matrices $P \in \mathrm{GL}_m(\mathbb{F}), Q \in \mathrm{GL}_n(\mathbb{F})$ so that

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

for some $0 \leq r \leq \min(m, n)$.

Theorem. If $A \in \mathrm{Mat}_{n,m}(\mathbb{F})$, then $r(A) = r(A^T)$, i.e. the row rank is equivalent to the column rank.

Proof. We know that there are some invertible P, Q such that

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

where $r = r(A)$. We can transpose this whole equation to obtain

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

So $r(A^T) = r$. □

2.5 Elementary matrix operations

Proposition. If $A \in \mathrm{Mat}_{n,m}(\mathbb{F})$, then there exists invertible matrices $P \in \mathrm{GL}_m(\mathbb{F}), Q \in \mathrm{GL}_n(\mathbb{F})$ so that

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

for some $0 \leq r \leq \min(m, n)$.

Proof. We claim that there are elementary matrices $E_1^m, \cdots, E_a^m$ and $F_1^n, \cdots, F_b^n$ (these E are not necessarily the shears, but any elementary matrix) such that

$$E_1^m \cdots E_a^m A F_1^n \cdots F_b^n = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

This suffices since the $E_i^m \in \mathrm{GL}_M(\mathbb{F})$ and $F_j^n \in \mathrm{GL}_n(\mathbb{F})$. Moreover, to prove the claim, it suffices to find a sequence of elementary row and column operations reducing A to this form.

If $A = 0$, then done. If not, there is some i, j such that $A_{ij} \neq 0$. By swapping row 1 and row i; and then column 1 and column j, we can assume $A_{11} \neq 0$. By rescaling row 1 by $\frac{1}{A_{11}}$, we can further assume $A_{11} = 1$.

Now we can add $-A_{1j}$ times column 1 to column j for each $j \neq 1$, and then add $-A_{i1}$ times row 1 to row $i \neq 1$. Then we now have

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

Now B is smaller than A. So by induction on the size of A, we can reduce B to a matrix of the required form, so done. □

3 Duality

3.1 Dual space

Lemma. If V is a finite-dimensional vector space over f with basis $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$, then there is a basis $(\varepsilon_1, \cdots, \varepsilon_n)$ for V^* (called the *dual basis* to $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$) such that

$$\varepsilon_i(\mathbf{e}_j) = \delta_{ij}.$$

Proof. Since linear maps are characterized by their values on a basis, there exists unique choices for $\varepsilon_1, \cdots, \varepsilon_n \in V^*$. Now we show that $(\varepsilon_1, \cdots, \varepsilon_n)$ is a basis.

Suppose $\theta \in V^*$. We show that we can write it uniquely as a combination of $\varepsilon_1, \cdots, \varepsilon_n$. We have $\theta = \sum_{i=1}^n \lambda_i \varepsilon_i$ if and only if $\theta(\mathbf{e}_j) = \sum_{i=1}^n \lambda_i \varepsilon_i(\mathbf{e}_j)$ (for all j) if and only if $\lambda_j = \theta(\mathbf{e}_j)$. So we have uniqueness and existence. □

Corollary. If V is finite dimensional, then $\dim V = \dim V^*$.

Proposition. Let V be a finite-dimensional vector space over $\mathbb{F}$ with bases $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ and $(\mathbf{f}_1, \cdots, \mathbf{f}_n)$, and that P is the change of basis matrix so that

$$\mathbf{f}_i = \sum_{k=1}^n P_{ki} \mathbf{e}_k.$$

Let $(\varepsilon_1, \cdots, \varepsilon_n)$ and $(\eta_1, \cdots, \eta_n)$ be the corresponding dual bases so that

$$\varepsilon_i(\mathbf{e}_j) = \delta_{ij} = \eta_i(\mathbf{f}_j).$$

Then the change of basis matrix from $(\varepsilon_1, \cdots, \varepsilon_n)$ to $(\eta_1, \cdots, \eta_n)$ is $(P^{-1})^T$, i.e.

$$\varepsilon_i = \sum_{\ell=1}^n P^T_{\ell i} \eta_\ell.$$

Proof. For convenience, write $Q = P^{-1}$ so that

$$\mathbf{e}_j = \sum_{k=1}^n Q_{kj} \mathbf{f}_k.$$

So we can compute

$$\begin{aligned}
\left(\sum_{\ell=1}^n P_{i\ell}\eta_\ell\right)(\mathbf{e}_j) &= \left(\sum_{\ell=1}^n P_{i\ell}\eta_\ell\right)\left(\sum_{k=1}^n Q_{kj}\mathbf{f}_k\right) \\
&= \sum_{k,\ell} P_{i\ell}\delta_{\ell k} Q_{kj} \\
&= \sum_{k,\ell} P_{i\ell} Q_{\ell j} \\
&= [PQ]_{ij} \\
&= \delta_{ij}.
\end{aligned}$$

So $\varepsilon_i = \sum_{\ell=1}^n P^T_{\ell i}\eta_\ell$. □

Proposition. Let V be a vector space over $\mathbb{F}$ and U a subspace. Then

$$\dim U + \dim U^0 = \dim V.$$

Proof. Let $(\mathbf{e}_1, \cdots, \mathbf{e}_k)$ be a basis for U and extend to $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ a basis for V. Consider the dual basis for V^*, say $(\varepsilon_1, \cdots, \varepsilon_n)$. Then we will show that

$$U^0 = \langle \varepsilon_{k+1}, \cdots, \varepsilon_n \rangle.$$

So $\dim U^0 = n - k$ as required. This is easy to prove — if $j > k$, then $\varepsilon_j(\mathbf{e}_i) = 0$ for all $i \leq k$. So $\varepsilon_{k+1}, \cdots, \varepsilon_n \in U^0$. On the other hand, suppose $\theta \in U^0$. Then we can write

$$\theta = \sum_{j=1}^{n} \lambda_j \varepsilon_j.$$

But then $0 = \theta(\mathbf{e}_i) = \lambda_i$ for $i \leq k$. So done. □

Proof. Consider the restriction map $V^* \to U^*$, given by $\theta \mapsto \theta|_U$. This is obviously linear. Since every linear map $U \to \mathbb{F}$ can be extended to $V \to \mathbb{F}$, this is a surjection. Moreover, the kernel is U^0. So by rank-nullity theorem,

$$\dim V^* = \dim U^0 + \dim U^*.$$

Since $\dim V^* = \dim V$ and $\dim U^* = \dim U$, we're done. □

Proof. We can show that $U^0 \simeq (V/U)^*$, and then deduce the result. Details are left as an exercise. □

3.2 Dual maps

Proposition. Let $\alpha \in \mathcal{L}(V, W)$ be a linear map. Then $\alpha^* \in \mathcal{L}(W^*, V^*)$ is a linear map.

Proof. Let $\lambda, \mu \in \mathbb{F}$ and $\theta_1, \theta_2 \in W^*$. We want to show

$$\alpha^*(\lambda\theta_1 + \mu\theta_2) = \lambda\alpha^*(\theta_1) + \mu\alpha^*(\theta_2).$$

To show this, we show that for every $\mathbf{v} \in V$, the left and right give the same result. We have

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

So $\alpha^* \in \mathcal{L}(W^*, V^*)$. □

Proposition. Let V, W be finite-dimensional vector spaces over $\mathbb{F}$ and $\alpha : V \to W$ be a linear map. Let $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ be a basis for V and $(\mathbf{f}_1, \cdots, \mathbf{f}_m)$ be a basis for W; $(\varepsilon_1, \cdots, \varepsilon_n)$ and $(\eta_1, \cdots, \eta_m)$ the corresponding dual bases.

Suppose α is represented by A with respect to $(\mathbf{e}_i)$ and $(\mathbf{f}_i)$ for V and W. Then α^* is represented by A^T with respect to the corresponding dual bases.

Proof. We are given that

$$\alpha(\mathbf{e}_i) = \sum_{k=1}^{m} A_{ki}\mathbf{f}_k.$$

We must compute $\alpha^*(\eta_i)$. To do so, we evaluate it at $\mathbf{e}_j$. We have

$$\alpha^*(\eta_i)(\mathbf{e}_j) = \eta_i(\alpha(\mathbf{e}_j)) = \eta_i\left(\sum_{k=1}^{m} A_{kj}\mathbf{f}_k\right) = \sum_{k=1}^{m} A_{kj}\delta_{ik} = A_{ij}.$$

We can also write this as

$$\alpha^*(\eta_i)(\mathbf{e}_j) = \sum_{k=1}^{n} A_{ik}\varepsilon_k(\mathbf{e}_j).$$

Since this is true for all j, we have

$$\alpha^*(\eta_i)\sum_{k=1}^{n} A_{ik}\varepsilon_k = \sum_{k=1}^{n} A^T_{ki}\varepsilon_k.$$

So done. □

Lemma. Let $\alpha \in \mathcal{L}(V, W)$ with V, W finite dimensional vector spaces over $\mathbb{F}$. Then

(i) $\ker \alpha^* = (\operatorname{im}\alpha)^0$.

(ii) $r(\alpha) = r(\alpha^*)$ (which is another proof that row rank is equal to column rank).

(iii) $\operatorname{im}\alpha^* = (\ker\alpha)^0$.

Proof.

(i) If $\theta \in W^*$, then

$$\begin{aligned}\theta \in \ker\alpha^* &\Leftrightarrow \alpha^*(\theta) = 0\\ &\Leftrightarrow (\forall \mathbf{v} \in V)\ \theta\alpha(\mathbf{v}) = 0\\ &\Leftrightarrow (\forall \mathbf{w} \in \operatorname{im}\alpha)\ \theta(\mathbf{w}) = 0\\ &\Leftrightarrow \theta \in (\operatorname{im}\alpha)^0.\end{aligned}$$

(ii) As $\operatorname{im}\alpha \leq W$, we've seen that

$$\dim\operatorname{im}\alpha + \dim(\operatorname{im}\alpha)^0 = \dim W.$$

Using (i), we see

$$n(\alpha^*) = \dim(\operatorname{im}\alpha)^0.$$

So

$$r(\alpha) + n(\alpha^*) = \dim W = \dim W^*.$$

By the rank-nullity theorem, we have $r(\alpha) = r(\alpha^*)$.

(iii) The proof in (i) doesn't quite work here. We can only show that one includes the other. To draw the conclusion, we will show that the two spaces have the dimensions, and hence must be equal.

Let $\theta \in \operatorname{im} \alpha^*$. Then $\theta = \phi\alpha$ for some $\phi \in W^*$. If $\mathbf{v} \in \ker \alpha$, then

$$\theta(\mathbf{v}) = \phi(\alpha(\mathbf{v})) = \phi(\mathbf{0}) = \mathbf{0}.$$

So $\operatorname{im} \alpha^* \subseteq (\ker \alpha)^0$.

But we know

$$\dim(\ker \alpha)^0 + \dim \ker \alpha = \dim V,$$

So we have

$$\dim(\ker \alpha)^0 = \dim V - n(\alpha) = r(\alpha) = r(\alpha^*) = \dim \operatorname{im} \alpha^*.$$

Hence we must have $\operatorname{im} \alpha^* = (\ker \alpha)^0$. □

Lemma. Let V be a vector space over $\mathbb{F}$. Then there is a linear map $\mathrm{ev} : V \to (V^*)^*$ given by

$$\mathrm{ev}(\mathbf{v})(\theta) = \theta(\mathbf{v}).$$

We call this the *evaluation* map.

Proof. We first show that $\mathrm{ev}(\mathbf{v}) \in V^{**}$ for all $\mathbf{v} \in V$, i.e. $\mathrm{ev}(\mathbf{v})$ is linear for any $\mathbf{v}$. For any $\lambda, \mu \in \mathbb{F}$, $\theta_1, \theta_2 \in V^*$, then for $\mathbf{v} \in V$, we have

$$\begin{aligned}\mathrm{ev}(\mathbf{v})(\lambda\theta_1 + \mu\theta_2) &= (\lambda\theta_1 + \mu\theta_2)(\mathbf{v})\\ &= \lambda\theta_1(\mathbf{v}) + \mu\theta_2(\mathbf{v})\\ &= \lambda\,\mathrm{ev}(\mathbf{v})(\theta_1) + \mu\,\mathrm{ev}(\mathbf{v})(\theta_2).\end{aligned}$$

So done. Now we show that ev itself is linear. Let $\lambda, \mu \in \mathbb{F}$, $\mathbf{v}_1, \mathbf{v}_2 \in V$. We want to show

$$\mathrm{ev}(\lambda\mathbf{v}_1 + \mu\mathbf{v}_2) = \lambda\,\mathrm{ev}(\mathbf{v}_1) + \mu\,\mathrm{ev}(\mathbf{v}_2).$$

To show these are equal, pick $\theta \in V^*$. Then

$$\begin{aligned}\mathrm{ev}(\lambda\mathbf{v}_1 + \mu\mathbf{v}_2)(\theta) &= \theta(\lambda\mathbf{v}_1 + \mu\mathbf{v}_2)\\ &= \lambda\theta(\mathbf{v}_1) + \mu\theta(\mathbf{v}_2)\\ &= \lambda\,\mathrm{ev}(\mathbf{v}_1)(\theta) + \mu\,\mathrm{ev}(\mathbf{v}_2)(\theta)\\ &= (\lambda\,\mathrm{ev}(\mathbf{v}_1) + \mu\,\mathrm{ev}(\mathbf{v}_2))(\theta).\end{aligned}$$

So done. □

Lemma. If V is finite-dimensional, then $\mathrm{ev} : V \to V^{**}$ is an isomorphism.

Proof. We first show it is injective. Suppose $\mathrm{ev}(\mathbf{v}) = \mathbf{0}$ for some $\mathbf{v} \in V$. Then $\theta(\mathbf{v}) = \mathrm{ev}(\mathbf{v})(\theta) = 0$ for all $\theta \in V^*$. So $\dim\langle\mathbf{v}\rangle^0 = \dim V^* = \dim V$. So $\dim\langle\mathbf{v}\rangle = 0$. So $\mathbf{v} = 0$. So ev is injective. Since V and V^{**} have the same dimension, this is also surjective. So done. □

Lemma. Let V, W be finite-dimensional vector spaces over $\mathbb{F}$ after identifying (V and V^{**}) and (W and W^{**}) by the evaluation map. Then we have

(i) If $U \leq V$, then $U^{00} = U$.

(ii) If $\alpha \in \mathcal{L}(V, W)$, then $\alpha^{**} = \alpha$.

Proof.

(i) Let $\mathbf{u} \in U$. Then $\mathbf{u}(\theta) = \theta(\mathbf{u}) = 0$ for all $\theta \in U^0$. So $\mathbf{u}$ annihilates everything in U^0. So $\mathbf{u} \in U^{00}$. So $U \subseteq U^{00}$. We also know that

$$\dim U = \dim V - \dim U^0 = \dim V - (\dim V - \dim U^{00}) = \dim U^{00}.$$

So we must have $U = U^{00}$.

(ii) The proof of this is basically — the transpose of the transpose is the original matrix. The only work we have to do is to show that the dual of the dual basis is the original basis.

Let $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ be a basis for V and $(\mathbf{f}_1, \cdots, \mathbf{f}_m)$ be a basis for W, and let $(\varepsilon_1, \cdots, \varepsilon_n)$ and $(\eta_1, \cdots, \eta_n)$ be the corresponding dual basis. We know that

$$\mathbf{e}_i(\varepsilon_j) = \delta_{ij} = \varepsilon_j(\mathbf{e}_i), \quad \mathbf{f}_i(\eta_j) = \delta_{ij} = \eta_j(\mathbf{f}_i).$$

So $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ is dual to $(\varepsilon_1, \cdots, \varepsilon_n)$, and similarly for $\mathbf{f}$ and η.

If α is represented by A, then α^* is represented by A^T. So α^{**} is represented by $(A^T)^T = A$. So done. □

Proposition. Let V be a finite-dimensional vector space $\mathbb{F}$ and U_1, U_2 are subspaces of V. Then we have

(i) $(U_1 + U_2)^0 = U_1^0 \cap U_2^0$

(ii) $(U_1 \cap U_2)^0 = U_1^0 + U_2^0$

Proof.

(i) Suppose $\theta \in V^*$. Then

$$\begin{aligned}\theta \in (U_1 + U_2)^0 &\Leftrightarrow \theta(\mathbf{u}_1 + \mathbf{u}_2) = 0 \text{ for all } \mathbf{u}_i \in U_i\\ &\Leftrightarrow \theta(\mathbf{u}) = 0 \text{ for all } \mathbf{u} \in U_1 \cup U_2\\ &\Leftrightarrow \theta \in U_1^0 \cap U_2^0.\end{aligned}$$

(ii) We have

$$(U_1 \cap U_2)^0 = ((U_1^0)^0 \cap (U_2^0)^0)^0 = (U_1^0 + U_2^0)^{00} = U_1^0 + U_2^0.$$

So done. □

4 Bilinear forms I

Proposition. Suppose $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ and $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ are basis for V such that

$$\mathbf{v}_i = \sum P_{ki}\mathbf{e}_k \text{ for all } i = 1, \cdots, n;$$

and $(\mathbf{f}_1, \cdots, \mathbf{f}_m)$ and $(\mathbf{w}_1, \cdots, \mathbf{w}_m)$ are bases for W such that

$$\mathbf{w}_i = \sum Q_{\ell j}\mathbf{f}_\ell \text{ for all } j = 1, \cdots, m.$$

Let $\psi : V \times W \to \mathbb{F}$ be a bilinear form represented by A with respect to $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ and $(\mathbf{f}_1, \cdots, \mathbf{f}_m)$, and by B with respect to the bases $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ and $(\mathbf{w}_1, \cdots, \mathbf{w}_m)$. Then

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

Proof. We have

$$\begin{aligned} B_{ij} &= \phi(\mathbf{v}_i, \mathbf{w}_j) \\ &= \phi\left(\sum P_{ki}\mathbf{e}_k, \sum Q_{\ell j}\mathbf{f}_\ell\right) \\ &= \sum P_{ki}Q_{\ell j}\phi(\mathbf{e}_k, \mathbf{f}_\ell) \\ &= \sum_{k,\ell} P_{ik}^T A_{k\ell} Q_{\ell j} \\ &= (P^T A Q)_{ij}. \end{aligned}$$

□

Lemma. Let $(\varepsilon_1, \cdots, \varepsilon_n)$ be a basis for V^* dual to $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ of V; $(\eta_1, \cdots, \eta_n)$ be a basis for W^* dual to $(\mathbf{f}_1, \cdots, \mathbf{f}_n)$ of W.

If A represents ψ with respect to $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ and $(\mathbf{f}_1, \cdots, \mathbf{f}_m)$, then A also represents ψ_R with respect to $(\mathbf{f}_1, \cdots, \mathbf{f}_m)$ and $(\varepsilon_1, \cdots, \varepsilon_n)$; and A^T represents ψ_L with respect to $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ and $(\eta_1, \cdots, \eta_m)$.

Proof. We just have to compute

$$\psi_L(\mathbf{e}_i)(\mathbf{f}_j) = A_{ij} = \left(\sum A_{i\ell}\eta_\ell\right)(\mathbf{f}_j).$$

So we get

$$\psi_L(\mathbf{e}_i) = \sum A_{\ell i}^T \eta_\ell.$$

So A^T represents ψ_L.

We also have

$$\psi_R(\mathbf{f}_j)(\mathbf{e}_i) = A_{ij}.$$

So

$$\psi_R(\mathbf{f}_j) = \sum A_{kj}\varepsilon_k.$$

□

Lemma. Let V and W be finite-dimensional vector spaces over $\mathbb{F}$ with bases $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ and $(\mathbf{f}_1, \cdots, \mathbf{f}_m)$ be their basis respectively.

Let $\psi : V \times W \to \mathbb{F}$ be a bilinear form represented by A with respect to these bases. Then ϕ is non-degenerate if and only if A is (square and) invertible. In particular, V and W have the same dimension.

Proof. Since ψ_R and ψ_L are represented by A and A^T (in some order), they both have trivial kernel if and only if $n(A) = n(A^T) = 0$. So we need $r(A) = \dim V$ and $r(A^T) = \dim W$. So we need $\dim V = \dim W$ and A have full rank, i.e. the corresponding linear map is bijective. So done. □

5 Determinants of matrices

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

Proof.

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

Since $\varepsilon(\tau) = \varepsilon(\tau^{-1})$, we get

$$\begin{aligned}&= \sum_{\tau\in S_n} \varepsilon(\tau)\prod_{j=1}^n A_{j\tau(j)}\\ &= \det A.\end{aligned}$$

□

Lemma. If A is an upper triangular matrix, i.e.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ 0 & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & a_{nn}\end{pmatrix}$$

Then

$$\det A = \prod_{i=1}^n a_{ii}.$$

Proof. We have

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

But $A_{i\sigma(i)} = 0$ whenever $i > \sigma(i)$. So

$$\prod_{i=1}^n A_{i\sigma(i)} = 0$$

if there is some $i \in \{1, \cdots, n\}$ such that $i > \sigma(i)$.

However, the only permutation in which $i \leq \sigma(i)$ for all i is the identity. So the only thing that contributes in the sum is $\sigma = \mathrm{id}$. So

$$\det A = \prod_{i=1}^n A_{ii}.$$

□

Lemma. $\det A$ is a volume form.

Proof. To see that det is multilinear, it is sufficient to show that each

$$\prod_{i=1}^{n} A_{i\sigma(i)}$$

is multilinear for all $\sigma \in S_n$, since linear combinations of multilinear forms are multilinear. But each such product is contains precisely one entry from each column, and so is multilinear.

To show it is alternating, suppose now there are some k, ℓ distinct such that $A^{(k)} = A^{(\ell)}$. We let τ be the transposition $(k\ \ell)$. By Lagrange's theorem, we can write

$$S_n = A_n \amalg \tau A_n,$$

where $A_n = \ker \varepsilon$ and $\amalg$ is the disjoint union. We also know that

$$\sum_{\sigma \in A_n} \prod_{i=1}^{n} A_{i\sigma(i)} = \sum_{\sigma \in A_n} \prod_{i=1}^{n} A_{i,\tau\sigma(i)},$$

since if $\sigma(i)$ is not k or l, then τ does nothing; if $\sigma(i)$ is k or l, then τ just swaps them around, but $A^{(k)} = A^{(l)}$. So we get

$$\sum_{\sigma \in A_n} \prod_{i=1}^{n} A_{i\sigma(i)} = \sum_{\sigma' \in \tau A_n} \prod_{i=1}^{n} A_{i\sigma'(i)},$$

But we know that

$$\det A = \text{LHS} - \text{RHS} = 0.$$

So done. □

Lemma. Let d be a volume form on $\mathbb{F}^n$. Then swapping two entries changes the sign, i.e.

$$d(\mathbf{v}_1, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_n) = -d(\mathbf{v}_1, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_n).$$

Proof. By linearity, we have

$$\begin{aligned}
0 &= d(\mathbf{v}_1, \cdots, \mathbf{v}_i + \mathbf{v}_j, \cdots, \mathbf{v}_i + \mathbf{v}_j, \cdots, \mathbf{v}_n) \\
&= d(\mathbf{v}_1, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_n) \\
&\quad + d(\mathbf{v}_1, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_n) \\
&\quad + d(\mathbf{v}_1, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_n) \\
&\quad + d(\mathbf{v}_1, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_n) \\
&= d(\mathbf{v}_1, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_n) \\
&\quad + d(\mathbf{v}_1, \cdots, \mathbf{v}_j, \cdots, \mathbf{v}_i, \cdots, \mathbf{v}_n).
\end{aligned}$$

So done. □

Corollary. If $\sigma \in S_n$, then

$$d(\mathbf{v}_{\sigma(1)}, \cdots, \mathbf{v}_{\sigma(n)}) = \varepsilon(\sigma) d(\mathbf{v}_1, \cdots, \mathbf{v}_n)$$

for any $\mathbf{v}_i \in \mathbb{F}^n$.

Theorem. Let d be any volume form on $\mathbb{F}^n$, and let $A = (A^{(1)} \ \cdots \ A^{(n)}) \in \mathrm{Mat}_n(\mathbb{F})$. Then

$$d(A^{(1)}, \cdots, A^{(n)}) = (\det A) d(\mathbf{e}_1, \cdots, \mathbf{e}_n),$$

where $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is the standard basis.

Proof. We can compute

$$\begin{aligned} d(A^{(1)}, \cdots, A^{(n)}) &= d\left(\sum_{i=1}^n A_{i1}\mathbf{e}_i, A^{(2)}, \cdots, A^{(n)}\right) \\ &= \sum_{i=1}^n A_{i1} d(\mathbf{e}_i, A^{(2)}, \cdots, A^{(n)}) \\ &= \sum_{i,j=1}^n A_{i1}A_{j2} d(\mathbf{e}_i, \mathbf{e}_j, A^{(3)}, \cdots, A^{(n)}) \\ &= \sum_{i_1, \cdots, i_n} d(\mathbf{e}_{i_1}, \cdots, \mathbf{e}_{i_n}) \prod_{j=1}^n A_{i_j j}. \end{aligned}$$

We know that lots of these are zero, since if $i_k = i_j$ for some k, j, then the term is zero. So we are just summing over distinct tuples, i.e. when there is some σ such that $i_j = \sigma(j)$. So we get

$$d(A^{(1)}, \cdots, A^{(n)}) = \sum_{\sigma \in S_n} d(\mathbf{e}_{\sigma(1)}, \cdots, \mathbf{e}_{\sigma(n)}) \prod_{j=1}^n A_{\sigma(j)j}.$$

However, by our corollary up there, this is just

$$d(A^{(1)}, \cdots, A^{(n)}) = \sum_{\sigma \in S_n} \varepsilon(\sigma) d(\mathbf{e}_1, \cdots, \mathbf{e}_n) \prod_{j=1}^n A_{\sigma(j)j} = (\det A) d(\mathbf{e}_1, \cdots, \mathbf{e}_n).$$

So done. □

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

Proof. Let d be a non-zero volume form on $\mathbb{F}^n$ (e.g. the "determinant"). Then we can compute

$$d(AB\mathbf{e}_1, \cdots, AB\mathbf{e}_n) = (\det AB) d(\mathbf{e}_1, \cdots, \mathbf{e}_n),$$

but we also have

$$d(AB\mathbf{e}_1, \cdots, AB\mathbf{e}_n) = (\det A) d(B\mathbf{e}_1, \cdots, B\mathbf{e}_n) = (\det A)(\det B) d(\mathbf{e}_1, \cdots, \mathbf{e}_n).$$

Since d is non-zero, we must have $\det AB = \det A \det B$. □

Corollary. If $A \in \mathrm{Mat}_n(\mathbb{F})$ is invertible, then $\det A \neq 0$. In fact, when A is invertible, then $\det(A^{-1}) = (\det A)^{-1}$.

Proof. We have

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

So done. □

Theorem. 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. We have proved that (i) $\Rightarrow$ (ii) above, and the rank-nullity theorem implies (iii) $\Rightarrow$ (i). We will prove (ii) $\Rightarrow$ (iii). In fact we will show the contrapositive. Suppose $r(A) < n$. By rank-nullity theorem, $n(A) > 0$. So there is some $\mathbf{x} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$ such that $A\mathbf{x} = \mathbf{0}$. Suppose $\lambda_k \neq 0$. We define B as follows:

$$B = \begin{pmatrix} 1 & & & \lambda_1 & & & \\ & \ddots & & \vdots & & & \\ & & 1 & \lambda_{k-1} & & & \\ & & & \lambda_k & & & \\ & & & \lambda_{k+1} & 1 & & \\ & & & \vdots & & \ddots & \\ & & & \lambda_n & & & 1 \end{pmatrix}$$

So AB has the kth column identically zero. So $\det(AB) = 0$. So it is sufficient to prove that $\det(B) \neq 0$. But $\det B = \lambda_k \neq 0$. So done. $\square$

Lemma. Let $A \in \mathrm{Mat}_n(\mathbb{F})$. Then

(i) We can expand $\det A$ along the jth column by

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det \hat{A}_{ij}.$$

(ii) We can expand $\det A$ along the ith row by

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det \hat{A}_{ij}.$$

Proof. Since $\det A = \det A^T$, (i) and (ii) are equivalent. So it suffices to prove just one of them. We have

$$\det A = d(A^{(1)}, \cdots, A^{(n)}),$$

where d is the volume form induced by the determinant. Then we can write as

$$\begin{aligned} \det A &= d\left(A^{(1)}, \cdots, \sum_{i=1}^{n} A_{ij}\mathbf{e}_i, \cdots, A^{(n)}\right) \\ &= \sum_{i=1}^{n} A_{ij} d(A^{(1)}, \cdots, \mathbf{e}_i, \cdots, A^{(n)}) \end{aligned}$$

The volume form on the right is the determinant of a matrix with the jth column replaced with $\mathbf{e}_i$. We can move our columns around so that our matrix becomes

$$B = \begin{pmatrix} \hat{A}_{ij} & 0 \\ \text{stuff} & 1 \end{pmatrix}$$

We get that $\det B = \det \hat{A}^{ij}$, since the only permutations that give a non-zero sum are those that send n to n. In the row and column swapping, we have made $n - j$ column transpositions and $n - i$ row transpositions. So we have

$$\begin{aligned} \det A &= \sum_{i=1}^{n} A_{ij}(-1)^{n-j}(-1)^{n-i} \det B \\ &= \sum_{i=1}^{n} A_{ij}(-1)^{i+j} \det \hat{A}_{ij}. \end{aligned}$$

□

Theorem. If $A \in \mathrm{Mat}_n(\mathbb{F})$, then $A(\operatorname{adj} A) = (\det A)I_n = (\operatorname{adj} A)A$. In particular, if $\det A \neq 0$, then

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A.$$

Proof. We compute

$$[(\operatorname{adj} A)A]_{jk} = \sum_{i=1}^{n} (\operatorname{adj} A)_{ji} A_{ik} = \sum_{i=1}^{n} (-1)^{i+j} \det \hat{A}_{ij} A_{ik}. \qquad (*)$$

So if $j = k$, then $[(\operatorname{adj} A)A]_{jk} = \det A$ by the lemma.

Otherwise, if $j \neq k$, consider the matrix B obtained from A by replacing the jth column by the kth column. Then the right hand side of $(*)$ is just $\det B$ by the lemma. But we know that if two columns are the same, the determinant is zero. So the right hand side of $(*)$ is zero. So

$$[(\operatorname{adj} A)A]_{jk} = \det A \delta_{jk}$$

The calculation for $[A \operatorname{adj} A] = (\det A)I_n$ can be done in a similar manner, or by considering $(A \operatorname{adj} A)^T = (\operatorname{adj} A)^T A^T = (\operatorname{adj}(A^T))A^T = (\det A)I_n$. □

Lemma. Let A, B be square matrices. Then for any C, we have

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

Proof. Suppose $A \in \mathrm{Mat}_k(\mathbb{F})$, and $B \in \mathrm{Mat}_\ell(\mathbb{F})$, so $C \in \mathrm{Mat}_{k,\ell}(\mathbb{F})$. Let

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

Then by definition, we have

$$\det X = \sum_{\sigma \in S_{k+\ell}} \varepsilon(\sigma) \prod_{i=1}^{k+\ell} X_{i\sigma(i)}.$$

If $j \leq k$ and $i > k$, then $X_{ij} = 0$. We only want to sum over permutations σ such that $\sigma(i) > k$ if $i > k$. So we are permuting the last j things among themselves, and hence the first k things among themselves. So we can decompose this into $\sigma = \sigma_1\sigma_2$, where σ_1 is a permutation of $\{1, \cdots, k\}$ and fixes the remaining things, while σ_2 fixes $\{1, \cdots, k\}$, and permutes the remaining. Then

$$\begin{aligned}
\det X &= \sum_{\sigma=\sigma_1\sigma_2} \varepsilon(\sigma_1\sigma_2) \prod_{i=1}^{k} X_{i\sigma_1(i)} \prod_{j=1}^{\ell} X_{k+j\ \sigma_2(k+j)} \\
&= \left(\sum_{\sigma_1 \in S_k} \varepsilon(\sigma_1) \prod_{i=1}^{k} A_{i\sigma_1(i)}\right)\left(\sum_{\sigma_2 \in S_\ell} \varepsilon(\sigma_2) \prod_{j=1}^{\ell} B_{j\sigma_2(j)}\right) \\
&= (\det A)(\det B)
\end{aligned}$$

□

Corollary.

$$\det \begin{pmatrix} A_1 & & & \text{stuff} \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_n \end{pmatrix} = \prod_{i=1}^{n} \det A_i$$

6 Endomorphisms

6.1 Invariants

Lemma. Suppose $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ and $(\mathbf{f}_1, \cdots, \mathbf{f}_n)$ are bases for V and $\alpha \in \text{End}(V)$. If A represents α with respect to $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ and B represents α with respect to $(\mathbf{f}_1, \cdots, \mathbf{f}_n)$, then

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

where P is given by

$$\mathbf{f}_i = \sum_{j=1}^{n} P_{ji}\mathbf{e}_j.$$

Proof. This is merely a special case of an earlier more general result for arbitrary maps and spaces. □

Lemma.

(i) If $A \in \text{Mat}_{m,n}(\mathbb{F})$ and $B \in \text{Mat}_{n,m}(\mathbb{F})$, then

$$\operatorname{tr} AB = \operatorname{tr} BA.$$

(ii) If $A, B \in \text{Mat}_n(\mathbb{F})$ are similar, then $\operatorname{tr} A = \operatorname{tr} B$.

(iii) If $A, B \in \text{Mat}_n(\mathbb{F})$ are similar, then $\det A = \det B$.

Proof.

(i) We have

$$\operatorname{tr} AB = \sum_{i=1}^{m}(AB)_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{m} B_{ji}A_{ij} = \operatorname{tr} BA.$$

(ii) Suppose $B = P^{-1}AP$. Then we have

$$\operatorname{tr} B = \operatorname{tr}(P^{-1}(AP)) = \operatorname{tr}((AP)P^{-1}) = \operatorname{tr} A.$$

(iii) We have

$$\det(P^{-1}AP) = \det P^{-1} \det A \det P = (\det P)^{-1} \det A \det P = \det A. \quad \square$$

Lemma. If A and B are similar, then they have the same characteristic polynomial.

Proof.

$$\det(tI - P^{-1}AP) = \det(P^{-1}(tI - A)P) = \det(tI - A). \quad \square$$

Lemma. Let $\alpha \in \text{End}(V)$ and $\lambda_1, \cdots, \lambda_k$ distinct eigenvalues of α. Then

$$E(\lambda_1) + \cdots + E(\lambda_k) = \bigoplus_{i=1}^{k} E(\lambda_i)$$

is a direct sum.

Proof. Suppose

$$\sum_{i=1}^{k} \mathbf{x}_i = \sum_{i=1}^{k} \mathbf{y}_i,$$

with $\mathbf{x}_i, \mathbf{y}_i \in E(\lambda_i)$. We want to show that they are equal. We are going to find some clever map that tells us what $\mathbf{x}_i$ and $\mathbf{y}_i$ are. Consider $\beta_j \in \mathrm{End}(V)$ defined by

$$\beta_j = \prod_{r \neq j} (\alpha - \lambda_r \iota).$$

Then

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

Each summand is zero, unless $i \neq j$. So this is equal to

$$\beta_j \left(\sum_{i=1}^{k} \mathbf{x}_i\right) = \prod_{r \neq j} (\lambda_j - \lambda_r)(\mathbf{x}_j).$$

Similarly, we obtain

$$\beta_j \left(\sum_{i=1}^{k} \mathbf{y}_i\right) = \prod_{r \neq j} (\lambda_j - \lambda_r)(\mathbf{y}_j).$$

Since we know that $\sum \mathbf{x}_i = \sum \mathbf{y}_i$, we must have

$$\prod_{r \neq j} (\lambda_j - \lambda_r)\mathbf{x}_j = \prod_{r \neq j} (\lambda_j - \lambda_r)\mathbf{y}_j.$$

Since we know that $\prod_{r \neq j} (\lambda_r - \lambda_j) \neq 0$, we must have $\mathbf{x}_i = \mathbf{y}_i$ for all i.

So each expression for $\sum \mathbf{x}_i$ is unique. □

Theorem. Let $\alpha \in \mathrm{End}(V)$ and $\lambda_1, \cdots, \lambda_k$ be distinct eigenvalues of α. Write E_i for $E(\lambda_i)$. Then the following are equivalent:

(i) α is diagonalizable.

(ii) V has a basis of eigenvectors for α.

(iii) $V = \bigoplus_{i=1}^{k} E_i$.

(iv) $\dim V = \sum_{i=1}^{k} \dim E_i$.

Proof.

– (i) ⇔ (ii): Suppose $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ is a basis for V. Then

$$\alpha(\mathbf{e}_i) = A_{ji}\mathbf{e}_j,$$

where A represents α. Then A is diagonal iff each $\mathbf{e}_i$ is an eigenvector. So done

– (ii) $\Leftrightarrow$ (iii): It is clear that (ii) is true iff $\sum E_i = V$, but we know that this must be a direct sum. So done.

– (iii) $\Leftrightarrow$ (iv): This follows from example sheet 1 Q10, which says that $V = \bigoplus_{i=1}^{k} E_i$ iff the bases for E_i are disjoint and their union is a basis of V. □

6.2 The minimal polynomial

6.2.1 Aside on polynomials

Lemma (Polynomial division)**.** If $f, g \in \mathbb{F}[t]$ (and $g \neq 0$), then there exists $q, r \in \mathbb{F}[t]$ with $\deg r < \deg g$ such that

$$f = qg + r.$$

Lemma. If $\lambda \in \mathbb{F}$ is a root of f, i.e. $f(\lambda) = 0$, then there is some g such that

$$f(t) = (t - \lambda)g(t).$$

Proof. By polynomial division, let

$$f(t) = (t - \lambda)g(t) + r(t)$$

for some $g(t), r(t) \in \mathbb{F}[t]$ with $\deg r < \deg(t - \lambda) = 1$. So r has to be constant, i.e. $r(t) = a_0$ for some $a_0 \in \mathbb{F}$. Now evaluate this at λ. So

$$0 = f(\lambda) = (\lambda - \lambda)g(\lambda) + r(\lambda) = a_0.$$

So $a_0 = 0$. So $r = 0$. So done. □

Lemma. A non-zero polynomial $f \in \mathbb{F}[t]$ has at most $\deg f$ roots, counted with multiplicity.

Corollary. Let $f, g \in \mathbb{F}[t]$ have degree $< n$. If there are $\lambda_1, \cdots, \lambda_n$ distinct such that $f(\lambda_i) = g(\lambda_i)$ for all i, then $f = g$.

Proof. Given the lemma, consider $f - g$. This has degree less than n, and $(f - g)(\lambda_i) = 0$ for $i = 1, \cdots, n$. Since it has at least $n \geq \deg(f - g)$ roots, we must have $f - g = 0$. So $f = g$. □

Corollary. If $\mathbb{F}$ is infinite, then f and g are equal if and only if they agree on all points.

Theorem (The fundamental theorem of algebra)**.** Every non-constant polynomial over $\mathbb{C}$ has a root in $\mathbb{C}$.

6.2.2 Minimal polynomial

Theorem (Diagonalizability theorem)**.** Suppose $\alpha \in \mathrm{End}(V)$. Then α is diagonalizable if and only if there exists non-zero $p(t) \in \mathbb{F}[t]$ such that $p(\alpha) = 0$, and $p(t)$ can be factored as a product of *distinct* linear factors.

Proof. Suppose α is diagonalizable. Let $\lambda_1, \cdots, \lambda_k$ be the distinct eigenvalues of α. We have

$$V = \bigoplus_{i=1}^{k} E(\lambda_i).$$

So each $\mathbf{v} \in V$ can be written (uniquely) as

$$\mathbf{v} = \sum_{i=1}^{k} \mathbf{v}_i \text{ with } \alpha(\mathbf{v}_i) = \lambda_i \mathbf{v}_i.$$

Now let

$$p(t) = \prod_{i=1}^{k} (t - \lambda_i).$$

Then for any $\mathbf{v}$, we get

$$p(\alpha)(\mathbf{v}) = \sum_{i=1}^{k} p(\alpha)(\mathbf{v}_i) = \sum_{i=1}^{k} p(\lambda_i)\mathbf{v}_i = \mathbf{0}.$$

So $p(\alpha) = 0$. By construction, p has distinct linear factors.

Conversely, suppose we have our polynomial

$$p(t) = \prod_{i=1}^{k} (t - \lambda_i),$$

with $\lambda_1, \cdots, \lambda_k \in \mathbb{F}$ distinct, and $p(\alpha) = 0$ (we can wlog assume p is monic, i.e. the leading coefficient is 1). We will show that

$$V = \sum_{i=1}^{k} E_\alpha(\lambda_i).$$

In other words, we want to show that for all $\mathbf{v} \in V$, there is some $\mathbf{v}_i \in E_\alpha(\lambda_i)$ for $i = 1, \cdots, k$ such that $\mathbf{v} = \sum \mathbf{v}_i$.

To find these $\mathbf{v}_i$ out, we let

$$q_j(t) = \prod_{i \neq j} \frac{t - \lambda_i}{\lambda_j - \lambda_i}.$$

This is a polynomial of degree $k - 1$, and $q_j(\lambda_i) = \delta_{ij}$.

Now consider

$$q(t) = \sum_{i=1}^{k} q_i(t).$$

We still have $\deg q \leq k - 1$, but $q(\lambda_i) = 1$ for any i. Since q and 1 agree on k points, we must have $q = 1$.

Let $\pi_j : V \to V$ be given by $\pi_j = q_j(\alpha)$. Then the above says that

$$\sum_{j=1}^{k} \pi_j = \iota.$$

Hence given $\mathbf{v} \in V$, we know that $\mathbf{v} = \sum \pi_j \mathbf{v}$.

We now check that $\pi_j \mathbf{v} \in E_\alpha(\lambda_j)$. This is true since

$$(\alpha - \lambda_j \iota)\pi_j \mathbf{v} = \frac{1}{\prod_{i\neq j}(\lambda_j - \lambda_i)} \prod_{i=1}^{k} (\alpha - \lambda_\iota)(\mathbf{v}) = \frac{1}{\prod_{i\neq j}(\lambda_j - \lambda_i)} p(\alpha)(\mathbf{v}) = \mathbf{0}.$$

So

$$\alpha \mathbf{v}_j = \lambda_j \mathbf{v}_j.$$

So done. □

Lemma. Let $\alpha \in \mathrm{End}(V)$, and $p \in \mathbb{F}[t]$. Then $p(\alpha) = 0$ if and only if $M_\alpha(t)$ is a factor of $p(t)$. In particular, M_α is unique.

Proof. For all such p, we can write $p(t) = q(t)M_\alpha(t) + r(t)$ for some r of degree less than $\deg M_\alpha$. Then

$$p(\alpha) = q(\alpha)M_\alpha(\alpha) + r(\alpha).$$

So if $r(\alpha) = 0$ iff $p(\alpha) = 0$. But $\deg r < \deg M_\alpha$. By the minimality of M_α, we must have $r(\alpha) = 0$ iff $r = 0$. So $p(\alpha) = 0$ iff $M_\alpha(t) \mid p(t)$.

So if M_1 and M_2 are both minimal polynomials for α, then $M_1 \mid M_2$ and $M_2 \mid M_1$. So M_2 is just a scalar multiple of M_1. But since M_1 and M_2 are monic, they must be equal. □

Theorem (Diagonalizability theorem 2.0)**.** Let $\alpha \in \mathrm{End}(V)$. Then α is diagonalizable if and only if $M_\alpha(t)$ is a product of its distinct linear factors.

Proof. ($\Leftarrow$) This follows directly from the previous diagonalizability theorem.

($\Rightarrow$) Suppose α is diagonalizable. Then there is some $p \in \mathbb{F}[t]$ non-zero such that $p(\alpha) = 0$ and p is a product of distinct linear factors. Since M_α divides p, M_α also has distinct linear factors. □

Theorem. Let $\alpha, \beta \in \mathrm{End}(V)$ be both diagonalizable. Then α and β are simultaneously diagonalizable (i.e. there exists a basis with respect to which both are diagonal) if and only if $\alpha\beta = \beta\alpha$.

Proof. ($\Rightarrow$) If there exists a basis $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ for V such that α and β are represented by A and B respectively, with both diagonal, then by direct computation, $AB = BA$. But AB represents $\alpha\beta$ and BA represents $\beta\alpha$. So $\alpha\beta = \beta\alpha$.

($\Leftarrow$) Suppose $\alpha\beta = \beta\alpha$. The idea is to consider each eigenspace of α individually, and then diagonalize β in each of the eigenspaces. Since α is diagonalizable, we can write

$$V = \bigoplus_{i=1}^{k} E_\alpha(\lambda_i),$$

where λ_i are the eigenvalues of V. We write E_i for $E_\alpha(\lambda_i)$. We want to show that β sends E_i to itself, i.e. $\beta(E_i) \subseteq E_i$. Let $\mathbf{v} \in E_i$. Then we want $\beta(\mathbf{v})$ to be in E_i. This is true since

$$\alpha(\beta(\mathbf{v})) = \beta(\alpha(\mathbf{v})) = \beta(\lambda_i \mathbf{v}) = \lambda_i \beta(\mathbf{v}).$$

So $\beta(\mathbf{v})$ is an eigenvector of α with eigenvalue λ_i.

Now we can view $\beta|_{E_i} \in \mathrm{End}(E_i)$. Note that

$$M_\beta(\beta|_{E_i}) = M_\beta(\beta)|_{E_i} = 0.$$

Since $M_\beta(t)$ is a product of its distinct linear factors, it follows that $\beta|_{E_i}$ is diagonalizable. So we can choose a basis B_i of eigenvectors for $\beta|_{E_i}$. We can do this for *all* i.

Then since V is a direct sum of the E_i's, we know that $B = \bigcup_{i=1}^k B_i$ is a basis for V consisting of eigenvectors for both α and β. So done. □

6.3 The Cayley-Hamilton theorem

Theorem (Cayley-Hamilton theorem)**.** Let V be a finite-dimensional vector space and $\alpha \in \mathrm{End}(V)$. Then $\chi_\alpha(\alpha) = 0$, i.e. $M_\alpha(t) \mid \chi_\alpha(t)$. In particular, $\deg M_\alpha \leq n$.

Lemma. An endomorphism α is triangulable if and only if $\chi_\alpha(t)$ can be written as a product of linear factors, not necessarily distinct. In particular, if $\mathbb{F} = \mathbb{C}$ (or any algebraically closed field), then every endomorphism is triangulable.

Proof. Suppose that α is triangulable and represented by

$$\begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Then

$$\chi_\alpha(t) = \det \begin{pmatrix} t-\lambda_1 & * & \cdots & * \\ 0 & t-\lambda_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t-\lambda_n \end{pmatrix} = \prod_{i=1}^n (t-\lambda_i).$$

So it is a product of linear factors.

We are going to prove the converse by induction on the dimension of our space. The base case $\dim V = 1$ is trivial, since every 1×1 matrix is already upper triangular.

Suppose $\alpha \in \mathrm{End}(V)$ and the result holds for all spaces of dimensions $< \dim V$, and χ_α is a product of linear factors. In particular, $\chi_\alpha(t)$ has a root, say $\lambda \in \mathbb{F}$.

Now let $U = E(\lambda) \neq 0$, and let W be a complementary subspace to U in V, i.e. $V = U \oplus W$. Let $\mathbf{u}_1, \cdots, \mathbf{u}_r$ be a basis for U and $\mathbf{w}_{r+1}, \cdots, \mathbf{w}_n$ be a basis for W so that $\mathbf{u}_1, \cdots, \mathbf{u}_r, \mathbf{w}_{r+1}, \cdots, \mathbf{w}_n$ is a basis for V, and α is represented by

$$\begin{pmatrix} \lambda I_r & \text{stuff} \\ 0 & B \end{pmatrix}$$

We know $\chi_\alpha(t) = (t-\lambda)^r \chi_B(t)$. So $\chi_B(t)$ is also a product of linear factors. We let $\beta : W \to W$ be the map defined by B with respect to $\mathbf{w}_{r+1}, \cdots, \mathbf{w}_n$.

(Note that in general, β is not $\alpha|_W$ in general, since α does not necessarily map W to W. However, we can say that $(\alpha - \beta)(\mathbf{w}) \in U$ for all $\mathbf{w} \in W$. This can

be much more elegantly expressed in terms of quotient spaces, but unfortunately that is not officially part of the course)

Since $\dim W < \dim V$, there is a basis $\mathbf{v}_{r+1}, \cdots, \mathbf{v}_n$ for W such that β is represented by C, which is upper triangular.

For $j = 1, \cdots, n-r$, we have

$$\alpha(\mathbf{v}_{j+r}) = \mathbf{u} + \sum_{k=1}^{n-r} C_{kj}\mathbf{v}_{k+r}$$

for some $\mathbf{u} \in U$. So α is represented by

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

with respect to $(\mathbf{u}_1, \cdots, \mathbf{u}_r, \mathbf{v}_{r+1}, \cdots, \mathbf{v}_n)$, which is upper triangular. □

Theorem (Cayley-Hamilton theorem)**.** Let V be a finite-dimensional vector space and $\alpha \in \mathrm{End}(V)$. Then $\chi_\alpha(\alpha) = 0$, i.e. $M_\alpha(t) \mid \chi_\alpha(t)$. In particular, $\deg M_\alpha \leq n$.

Proof. In this proof, we will work over $\mathbb{C}$. By the lemma, we can choose a basis $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is represented by an upper triangular matrix.

$$A = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

We must prove that

$$\chi_\alpha(\alpha) = \chi_A(\alpha) = \prod_{i=1}^{n}(\alpha - \lambda_i\iota) = 0.$$

Write $V_j = \langle \mathbf{e}_1, \cdots, \mathbf{e}_j \rangle$. So we have the inclusions

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

We also know that $\dim V_j = j$. This increasing sequence is known as a *flag*.

Now note that since A is upper-triangular, we get

$$\alpha(\mathbf{e}_i) = \sum_{k=1}^{i} A_{ki}\mathbf{e}_k \in V_i.$$

So $\alpha(V_j) \subseteq V_j$ for all $j = 0, \cdots, n$.

Moreover, we have

$$(\alpha - \lambda_j\iota)(\mathbf{e}_j) = \sum_{k=1}^{j-1} A_{kj}\mathbf{e}_k \subseteq V_{j-1}$$

for all $j = 1, \cdots, n$. So every time we apply one of these things, we get to a smaller space. Hence by induction on $n - j$, we have

$$\prod_{i=j}^{n}(\alpha - \lambda_i\iota)(V_n) \subseteq V_{j-1}.$$

In particular, when $j = 1$, we get

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

So $\chi_\alpha(\alpha) = 0$ as required.

Note that if our field $\mathbb{F}$ is not $\mathbb{C}$ but just a subfield of $\mathbb{C}$, say $\mathbb{R}$, we can just pretend it is a complex matrix, do the same proof. □

Proof. We'll now prove the theorem again, which is somewhat a formalization of the "nonsense proof" where we just substitute $t = \alpha$ into $\det(\alpha - t\iota)$.

Let α be represented by A, and $B = tI - A$. Then

$$B \operatorname{adj} B = \det B I_n = \chi_\alpha(t) I_n.$$

But we know that $\operatorname{adj} B$ is a matrix with entries in $\mathbb{F}[t]$ of degree at most $n - 1$. So we can write

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

with $B_i \in \operatorname{Mat}_n(\mathbb{F})$. We can also write

$$\chi_\alpha(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_0.$$

Then we get the result

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

We would like to just throw in $t = A$, and get the desired result, but in all these derivations, t is assumed to be a real number, and, $tI_n - A$ is the matrix

$$\begin{pmatrix} t - a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & t - a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & t - a_{nn} \end{pmatrix}$$

It doesn't make sense to put our A in there.

However, what we *can* do is to note that since this is true for all values of t, the coefficients on both sides must be equal. Equating coefficients in t^k, we have

$$\begin{aligned} -AB_0 &= a_0 I_n \\ B_0 - AB_1 &= a_1 I_n \\ &\vdots \\ B_{n-2} - AB_{n-1} &= a_{n-1} I_n \\ AB_{n-1} - 0 &= I_n \end{aligned}$$

We now multiply each row a suitable power of A to obtain

$$\begin{aligned} -AB_0 &= a_0 I_n \\ AB_0 - A^2 B_1 &= a_1 A \\ &\vdots \\ A^{n-1}B_{n-2} - A^n B_{n-1} &= a_{n-1} A^{n-1} \\ A^n B_{n-1} - 0 &= A^n. \end{aligned}$$

Summing this up then gives $\chi_\alpha(A) = 0$. □

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

(i) λ is an eigenvalue of α.

(ii) λ is a root of $\chi_\alpha(t)$.

(iii) λ is a root of $M_\alpha(t)$.

Proof.

- (i) $\Leftrightarrow$ (ii): λ is an eigenvalue of α if and only if $(\alpha - \lambda\iota)(\mathbf{v}) = 0$ has a non-trivial root, iff $\det(\alpha - \lambda\iota) = 0$.
- (iii) $\Rightarrow$ (ii): This follows from Cayley-Hamilton theorem since $M_\alpha \mid \chi_\alpha$.
- (i) $\Rightarrow$ (iii): Let λ be an eigenvalue, and $\mathbf{v}$ be a corresponding eigenvector. Then by definition of M_α, we have

  $$M_\alpha(\alpha)(\mathbf{v}) = 0(\mathbf{v}) = 0.$$

  We also know that

  $$M_\alpha(\alpha)(\mathbf{v}) = M_\alpha(\lambda)\mathbf{v}.$$

  Since $\mathbf{v}$ is non-zero, we must have $M_\alpha(\lambda) = 0$.
- (iii) $\Rightarrow$ (i): This is not necessary since it follows from the above, but we could as well do it explicitly. Suppose λ is a root of $M_\alpha(t)$. Then $M_\alpha(t) = (t - \lambda)g(t)$ for some $g \in \mathbb{F}[t]$. But $\deg g < \deg M_\alpha$. Hence by minimality of M_α, we must have $g(\alpha) \neq 0$. So there is some $\mathbf{v} \in V$ such that $g(\alpha)(\mathbf{v}) \neq 0$. Then

  $$(\alpha - \lambda\iota)g(\alpha)(\mathbf{v}) = M_\alpha(\alpha)\mathbf{v} = 0.$$

  So we must have $\alpha(g(\alpha)(\mathbf{v})) = \lambda g(\alpha)(\mathbf{v})$. So $g(\alpha)(\mathbf{v}) \in E_\alpha(\lambda) \setminus \{0\}$. So (i) holds. □

6.4 Multiplicities of eigenvalues and Jordan normal form

Lemma. If λ is an eigenvalue of α, then

(i) $1 \leq g_\lambda \leq a_\lambda$

(ii) $1 \leq c_\lambda \leq a_\lambda$.

Proof.

(i) The first inequality is easy. If λ is an eigenvalue, then $E(\lambda) \neq 0$. So $g_\lambda = \dim E(\lambda) \geq 1$. To prove the other inequality, if $\mathbf{v}_1, \cdots, \mathbf{v}_g$ is a basis for $E(\lambda)$, then we can extend it to a basis for V, and then α is represented by

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

So $\chi_\alpha(t) = (t - \lambda)^g \chi_B(t)$. So $a_\lambda > g = g_\lambda$.

(ii) This is straightforward since $M_\alpha(\lambda) = 0$ implies $1 \leq c_\lambda$, and since $M_\alpha(t) \mid \chi_\alpha(t)$, we know that $c_\lambda \leq \alpha_\lambda$. □

Lemma. Suppose $\mathbb{F} = \mathbb{C}$ and $\alpha \in \mathrm{End}(V)$. Then the following are equivalent:

(i) α is diagonalizable.

(ii) $g_\lambda = a_\lambda$ for all eigenvalues of α.

(iii) $c_\lambda = 1$ for all λ.

Proof.

- (i) $\Leftrightarrow$ (ii): α is diagonalizable iff $\dim V = \sum \dim E_\alpha(\lambda_i)$. But this is equivalent to
$$\dim V = \sum g_{\lambda_i} \leq \sum a_{\lambda_i} = \deg \chi_\alpha = \dim V.$$
So we must have $\sum g_{\lambda_i} = \sum a_{\lambda_i}$. Since each g_{λ_i} is at most a_{λ_i}, they must be individually equal.
- (i) $\Leftrightarrow$ (iii): α is diagonalizable if and only if $M_\alpha(t)$ is a product of distinct linear factors if and only if $c_\lambda = 1$ for all eigenvalues λ. $\square$

Theorem (Jordan normal form theorem)**.** Every matrix $A \in \mathrm{Mat}_n(\mathbb{C})$ is similar to a matrix in Jordan normal form. Moreover, this Jordan normal form matrix is unique up to permutation of the blocks.

Theorem. Let $\alpha \in \mathrm{End}(V)$, and A in Jordan normal form representing α. Then the number of Jordan blocks $J_n(\lambda)$ in A with $n \geq r$ is
$$n((\alpha - \lambda\iota)^r) - n((\alpha - \lambda\iota)^{r-1}).$$

Proof. We work blockwise for
$$A = \begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_k}(\lambda_k) \end{pmatrix}.$$
We have previously computed
$$n((J_m(\lambda) - \lambda I_m)^r) = \begin{cases} r & r \leq m \\ m & r > m \end{cases}.$$
Hence we know
$$n((J_m(\lambda) - \lambda I_m)^r) - n((J_m(\lambda) - \lambda I_m)^{r-1}) = \begin{cases} 1 & r \leq m \\ 0 & \text{otherwise.} \end{cases}$$
It is also easy to see that for $\mu \neq \lambda$,
$$n((J_m(\mu) - \lambda I_m)^r) = n(J_m(\mu - \lambda)^r) = 0$$
Adding up for each block, for $r \geq 1$, we have
$$n((\alpha - \lambda\iota)^r) - n((\alpha - \lambda\iota)^{r-1}) = \text{ number of Jordan blocks } J_n(\lambda) \text{ with } n \geq r. \ \square$$

Theorem (Generalized eigenspace decomposition)**.** Let V be a finite-dimensional vector space $\mathbb{C}$ such that $\alpha \in \mathrm{End}(V)$. Suppose that

$$M_\alpha(t) = \prod_{i=1}^{k} (t - \lambda_i)^{c_i},$$

with $\lambda_1, \cdots, \lambda_k \in \mathbb{C}$ distinct. Then

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

where $V_i = \ker((\alpha - \lambda_i \iota)^{c_i})$ is the *generalized eigenspace*.

Proof. Let

$$p_j(t) = \prod_{i \neq j} (t - \lambda_i)^{c_i}.$$

Then $p_1, \cdots, p_k$ have no common factors, i.e. they are coprime. Thus by Euclid's algorithm, there exists $q_1, \cdots, q_k \in \mathbb{C}[t]$ such that

$$\sum p_i q_i = 1.$$

We now define the endomorphism

$$\pi_j = q_j(\alpha) p_j(\alpha)$$

for $j = 1, \cdots, k$.

Then $\sum \pi_j = \iota$. Since $M_\alpha(\alpha) = 0$ and $M_\alpha(t) = (t - \lambda_j \iota)^{c_j} p_j(t)$, we get

$$(\alpha - \lambda_j \iota)^{c_j} \pi_j = 0.$$

So $\operatorname{im} \pi_j \subseteq V_j$.

Now suppose $\mathbf{v} \in V$. Then

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

So

$$V = \sum V_j.$$

To show this is a direct sum, note that $\pi_i \pi_j = 0$, since the product contains $M_\alpha(\alpha)$ as a factor. So

$$\pi_i = \iota \pi_i = \left(\sum \pi_j \right) \pi_i = \pi_i^2.$$

So π is a projection, and $\pi_j|_{V_j} = \iota_{V_j}$. So if $\mathbf{v} = \sum \mathbf{v}_i$, then applying π_i to both sides gives $\mathbf{v}_i = \pi_i(\mathbf{v})$. Hence there is a unique way of writing $\mathbf{v}$ as a sum of things in V_i. So $V = \bigoplus V_j$ as claimed. □

7 Bilinear forms II

7.1 Symmetric bilinear forms and quadratic forms

Lemma. Let V be a finite-dimensional vector space over $\mathbb{F}$, and $\phi : V \times V \to \mathbb{F}$ is a symmetric bilinear form. Let $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ be a basis for V, and let M be the matrix representing ϕ with respect to this basis, i.e. $M_{ij} = \phi(\mathbf{e}_i, \mathbf{e}_j)$. Then ϕ is symmetric if and only if M is symmetric.

Proof. If ϕ is symmetric, then

$$M_{ij} = \phi(\mathbf{e}_i, \mathbf{e}_j) = \phi(\mathbf{e}_j, \mathbf{e}_i) = M_{ji}.$$

So $M^T = M$. So M is symmetric.

If M is symmetric, then

$$\begin{aligned}\phi(\mathbf{x}, \mathbf{y}) &= \phi\left(\sum x_i \mathbf{e}_i, \sum y_j \mathbf{e}_j\right)\\ &= \sum_{i,j} x_i M_{ij} y_j\\ &= \sum_{i,j}^{n} y_j M_{ji} x_i\\ &= \phi(\mathbf{y}, \mathbf{x}).\end{aligned}$$

□

Lemma. Let V is a finite-dimensional vector space, and $\phi : V \times V \to \mathbb{F}$ a bilinear form. Let $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ and $(\mathbf{f}_1, \cdots, \mathbf{f}_n)$ be bases of V such that

$$\mathbf{f}_i = \sum_{k=1}^{n} P_{ki} \mathbf{e}_k.$$

If A represents ϕ with respect to $(\mathbf{e}_i)$ and B represents ϕ with respect to $(\mathbf{f}_i)$, then

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

Proof. Special case of general formula proven before. □

Proposition (Polarization identity)**.** Suppose that $\operatorname{char} \mathbb{F} \neq 2$, i.e. $1 + 1 \neq 0$ on $\mathbb{F}$ (e.g. if $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$). If $q : V \to \mathbb{F}$ is a quadratic form, then there exists a *unique* symmetric bilinear form $\phi : V \times V \to \mathbb{F}$ such that

$$q(\mathbf{v}) = \phi(\mathbf{v}, \mathbf{v}).$$

Proof. Let $\psi : V \times V \to \mathbb{F}$ be a bilinear form such that $\psi(\mathbf{v}, \mathbf{v}) = q(\mathbf{v})$. We define $\phi : V \times V \to \mathbb{F}$ by

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

for all $\mathbf{v}, \mathbf{w} \in \mathbb{F}$. This is clearly a bilinear form, and it is also clearly symmetric and satisfies the condition we wants. So we have proved the existence part.

To prove uniqueness, we want to find out the values of $\phi(\mathbf{v}, \mathbf{w})$ in terms of what q tells us. Suppose ϕ is such a symmetric bilinear form. We compute

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

So we have

$$\phi(\mathbf{v}, \mathbf{w}) = \frac{1}{2}(q(\mathbf{v}+\mathbf{w}) - q(\mathbf{v}) - q(\mathbf{w})).$$

So it is determined by q, and hence unique. □

Theorem. Let V be a finite-dimensional vector space over $\mathbb{F}$, and $\phi : V \times V \to \mathbb{F}$ a symmetric bilinear form. Then there exists a basis $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ for V such that ϕ is represented by a diagonal matrix with respect to this basis.

Proof. We induct over $n = \dim V$. The cases $n = 0$ and $n = 1$ are trivial, since all matrices are diagonal.

Suppose we have proven the result for all spaces of dimension less than n. First consider the case where $\phi(\mathbf{v}, \mathbf{v}) = 0$ for all $\mathbf{v} \in V$. We want to show that we must have $\phi = 0$. This follows from the polarization identity, since this ϕ induces the zero quadratic form, and we know that there is a unique bilinear form that induces the zero quadratic form. Since we know that the zero bilinear form also induces the zero quadratic form, we must have $\phi = 0$. Then ϕ will be represented by the zero matrix with respect to any basis, which is trivially diagonal.

If not, pick $\mathbf{e}_1 \in V$ such that $\phi(\mathbf{e}_1, \mathbf{e}_1) \neq 0$. Let

$$U = \ker \phi(\mathbf{e}_1, \cdot) = \{\mathbf{u} \in V : \phi(\mathbf{e}_1, \mathbf{u}) = 0\}.$$

Since $\phi(\mathbf{e}_1, \cdot) \in V^* \setminus \{0\}$, we know that $\dim U = n - 1$ by the rank-nullity theorem.

Our objective is to find other basis elements $\mathbf{e}_2, \cdots, \mathbf{e}_n$ such that $\phi(\mathbf{e}_1, \mathbf{e}_j) = 0$ for all $j > 1$. For this to happen, we need to find them inside U.

Now consider $\phi|_{U\times U} : U \times U \to \mathbb{F}$, a symmetric bilinear form. By the induction hypothesis, we can find a basis $\mathbf{e}_2, \cdots, \mathbf{e}_n$ for U such that $\phi|_{U\times U}$ is represented by a diagonal matrix with respect to this basis.

Now by construction, $\phi(\mathbf{e}_i, \mathbf{e}_j) = 0$ for all $1 \leq i \neq j \leq n$ and $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ is a basis for V. So we're done. □

Theorem. Let ϕ be a symmetric bilinear form over a complex vector space V. Then there exists a basis $(\mathbf{v}_1, \cdots, \mathbf{v}_m)$ for V such that ϕ is represented by

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

with respect to this basis, where $r = r(\phi)$.

Proof. We've already shown that there exists a basis $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ such that $\phi(\mathbf{e}_i, \mathbf{e}_j) = \lambda_i \delta_{ij}$ for some λ_{ij}. By reordering the $\mathbf{e}_i$, we can assume that $\lambda_1, \cdots, \lambda_r \neq 0$ and $\lambda_{r+1}, \cdots, \lambda_n = 0$.

For each $1 \leq i \leq r$, there exists some μ_i such that $\mu_i^2 = \lambda_i$. For $r+1 \leq r \leq n$, we let $\mu_i = 1$ (or anything non-zero). We define

$$\mathbf{v}_i = \frac{\mathbf{e}_i}{\mu_i}.$$

Then

$$\phi(\mathbf{v}_i, \mathbf{v}_j) = \frac{1}{\mu_i \mu_j}\phi(\mathbf{e}_i, \mathbf{e}_j) = \begin{cases} 0 & i \neq j \text{ or } i = j > r \\ 1 & i = j < r. \end{cases}$$

So done. □

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

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

Theorem. Let ϕ be a symmetric bilinear form of a finite-dimensional vector space over $\mathbb{R}$. Then there exists a basis $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ for V such that ϕ is represented

$$\begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix},$$

with $p + q = r(\phi)$, $p, q \geq 0$. Equivalently, the corresponding quadratic forms is given by

$$q\left(\sum_{i=1}^n a_i \mathbf{v}_i\right) = \sum_{i=1}^p a_i^2 - \sum_{j=p+1}^{p+q} a_j^2.$$

Proof. We've already shown that there exists a basis $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ such that $\phi(\mathbf{e}_i, \mathbf{e}_j) = \lambda_i \delta_{ij}$ for some $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$. By reordering, we may assume

$$\begin{cases} \lambda_i > 0 & 1 \leq i \leq p \\ \lambda_i < 0 & p+1 \leq i \leq r \\ \lambda_i = 0 & i > r \end{cases}$$

We let μ_i be defined by

$$\mu_i = \begin{cases} \sqrt{\lambda_i} & 1 \leq i \leq p \\ \sqrt{-\lambda_i} & p+1 \leq i \leq r \\ 1 & i > r \end{cases}$$

Defining

$$\mathbf{v}_i = \frac{1}{\mu_i}\mathbf{e}_i,$$

we find that ϕ is indeed represented by

$$\begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix},$$

□

Theorem (Sylvester's law of inertia)**.** Let ϕ be a symmetric bilinear form on a finite-dimensional real vector space V. Then there exists unique non-negative integers p, q such that ϕ is represented by

$$\begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with respect to some basis.

Proof. We have already proved the existence part, and we just have to prove uniqueness. To do so, we characterize p and q in a basis-independent way. We already know that $p + q = r(\phi)$ does not depend on the basis. So it suffices to show p is unique.

To see that p is unique, we show that p is the largest dimension of a subspace $P \subseteq V$ such that $\phi|_{P\times P}$ is positive definite.

First we show we can find such at P. Suppose ϕ is represented by

$$\begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with respect to $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$. Then ϕ restricted to $\langle \mathbf{e}_1, \cdots, \mathbf{e}_p\rangle$ is represented by I_p with respect to $\mathbf{e}_1, \cdots, \mathbf{e}_p$. So ϕ restricted to this is positive definite.

Now suppose P is any subspace of V such that $\phi|_{P\times P}$ is positive definite. To show P has dimension at most p, we find a subspace complementary to P with dimension $n - p$.

Let $Q = \langle \mathbf{e}_{p+1}, \cdots, \mathbf{e}_n\rangle$. Then ϕ restricted to $Q \times Q$ is represented by

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

Now if $\mathbf{v} \in P \cap Q \setminus \{0\}$, then $\phi(\mathbf{v}, \mathbf{v}) > 0$ since $\mathbf{v} \in P \setminus \{0\}$ and $\phi(\mathbf{v}, \mathbf{v}) \leq 0$ since $\mathbf{v} \in Q$, which is a contradiction. So $P \cap Q = 0$.

We have

$$\dim V \geq \dim(P + Q) = \dim P + \dim Q = \dim P + (n - p).$$

Rearranging gives

$$\dim P \leq p.$$

A similar argument shows that q is the maximal dimension of a subspace $Q \subseteq V$ such that $\phi|_{Q\times Q}$ is negative definite. □

Corollary. Every real symmetric matrix is congruent to precisely one matrix of the form

$$\begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

7.2 Hermitian form

Lemma. Let $\phi : V \times V \to \mathbb{C}$ be a sesquilinear form on a finite-dimensional vector space over $\mathbb{C}$, and $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ a basis for V. Then ϕ is Hermitian if and only if the matrix A representing ϕ is Hermitian (i.e. $A = A^\dagger$).

Proof. If ϕ is Hermitian, then

$$A_{ij} = \phi(\mathbf{e}_i, \mathbf{e}_j) = \overline{\phi(\mathbf{e}_j, \mathbf{e}_i)} = A^\dagger_{ij}.$$

If A is Hermitian, then

$$\phi\left(\sum \lambda_i \mathbf{e}_i, \sum \mu_j \mathbf{e}_j\right) = \lambda^\dagger A \mu = \overline{\mu^\dagger A^\dagger \lambda} = \overline{\phi\left(\sum \mu_j \mathbf{e}_j, \sum \lambda_j \mathbf{e}_j\right)}.$$

So done. □

Proposition (Change of basis)**.** Let ϕ be a Hermitian form on a finite dimensional vector space V; $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ and $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ are bases for V such that

$$\mathbf{v}_i = \sum_{k=1}^n P_{ki} \mathbf{e}_k;$$

and A, B represent ϕ with respect to $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ and $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ respectively. Then

$$B = P^\dagger A P.$$

Proof. We have

$$\begin{aligned} B_{ij} &= \phi(\mathbf{v}_i, \mathbf{v}_j) \\ &= \phi\left(\sum P_{ki} \mathbf{e}_k, \sum P_{\ell j} \mathbf{e}_\ell\right) \\ &= \sum_{k,\ell=1}^n \bar{P}_{ki} P_{\ell j} A_{k\ell} \\ &= (P^\dagger A P)_{ij}. \end{aligned}$$

□

Lemma (Polarization identity (again))**.** A Hermitian form ϕ on V is determined by the function $\psi : \mathbf{v} \mapsto \phi(\mathbf{v}, \mathbf{v})$.

Proof. We have the following:

$$\begin{aligned} \psi(\mathbf{x} + \mathbf{y}) &= \phi(\mathbf{x}, \mathbf{x}) + \phi(\mathbf{x}, \mathbf{y}) + \phi(\mathbf{y}, \mathbf{x}) + \phi(\mathbf{y}, \mathbf{y}) \\ -\psi(\mathbf{x} - \mathbf{y}) &= -\phi(\mathbf{x}, \mathbf{x}) + \phi(\mathbf{x}, \mathbf{y}) + \phi(\mathbf{y}, \mathbf{x}) - \phi(\mathbf{y}, \mathbf{y}) \\ i\psi(\mathbf{x} - i\mathbf{y}) &= i\phi(\mathbf{x}, \mathbf{x}) + \phi(\mathbf{x}, \mathbf{y}) - \phi(\mathbf{y}, \mathbf{x}) + i\phi(\mathbf{y}, \mathbf{y}) \\ -i\psi(\mathbf{x} + i\mathbf{y}) &= -i\phi(\mathbf{x}, \mathbf{x}) + \phi(\mathbf{x}, \mathbf{y}) - \phi(\mathbf{y}, \mathbf{x}) - i\phi(\mathbf{y}, \mathbf{y}) \end{aligned}$$

So

$$\phi(\mathbf{x}, \mathbf{y}) = \frac{1}{4}(\psi(\mathbf{x} + \mathbf{y}) - \psi(\mathbf{x} - \mathbf{y}) + i\psi(\mathbf{x} - i\mathbf{y}) - i\psi(\mathbf{x} + i\mathbf{y})).$$

□

Theorem (Hermitian form of Sylvester's law of inertia)**.** Let V be a finite-dimensional complex vector space and ϕ a hermitian form on V. Then there exists unique non-negative integers p and q such that ϕ is represented by

$$\begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with respect to some basis.

Proof. Same as for symmetric forms over $\mathbb{R}$. □

8 Inner product spaces

8.1 Definitions and basic properties

Theorem (Cauchy-Schwarz inequality)**.** Let V be an inner product space and $\mathbf{v}, \mathbf{w} \in V$. Then

$$|(\mathbf{v}, \mathbf{w})| \leq \|\mathbf{v}\|\|\mathbf{w}\|.$$

Proof. If $\mathbf{w} = 0$, then this is trivial. Otherwise, since the norm is positive definite, for any λ, we get

$$0 \leq (\mathbf{v} - \lambda\mathbf{w}, \mathbf{v} - \lambda\mathbf{w}) = (\mathbf{v}, \mathbf{v}) - \bar{\lambda}(\mathbf{w}, \mathbf{v}) - \lambda(\mathbf{v}, \mathbf{w}) + |\lambda|^2(\mathbf{w}, \mathbf{w}).$$

We now pick a clever value of λ. We let

$$\lambda = \frac{(\mathbf{w}, \mathbf{v})}{(\mathbf{w}, \mathbf{w})}.$$

Then we get

$$0 \leq (\mathbf{v}, \mathbf{v}) - \frac{|(\mathbf{w}, \mathbf{v})|^2}{(\mathbf{w}, \mathbf{w})} - \frac{|(\mathbf{w}, \mathbf{v})|^2}{(\mathbf{w}, \mathbf{w})} + \frac{|(\mathbf{w}, \mathbf{v})|^2}{(\mathbf{w}, \mathbf{w})}.$$

So we get

$$|(\mathbf{w}, \mathbf{v})|^2 \leq (\mathbf{v}, \mathbf{v})(\mathbf{w}, \mathbf{w}).$$

So done. □

Corollary (Triangle inequality)**.** Let V be an inner product space and $\mathbf{v}, \mathbf{w} \in V$. Then

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|.$$

Proof. We compute

$$\begin{aligned}
\|\mathbf{v} + \mathbf{w}\|^2 &= (\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}) \\
&= (\mathbf{v}, \mathbf{v}) + (\mathbf{v}, \mathbf{w}) + (\mathbf{w}, \mathbf{v}) + (\mathbf{w}, \mathbf{w}) \\
&\leq \|\mathbf{v}\|^2 + 2\|\mathbf{v}\|\|\mathbf{w}\| + \|\mathbf{w}\|^2 \\
&= (\|\mathbf{v}\| + \|\mathbf{w}\|)^2.
\end{aligned}$$

So done. □

Lemma (Parseval's identity)**.** Let V be a finite-dimensional inner product space with an orthonormal basis $\mathbf{v}_1, \cdots, \mathbf{v}_n$, and $\mathbf{v}, \mathbf{w} \in V$. Then

$$(\mathbf{v}, \mathbf{w}) = \sum_{i=1}^{n} \overline{(\mathbf{v}_i, \mathbf{v})}(\mathbf{v}_i, \mathbf{w}).$$

In particular,

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

Proof.

$$\begin{aligned}(\mathbf{v}, \mathbf{w}) &= \left(\sum_{i=1}^n (\mathbf{v}_i, \mathbf{v})\mathbf{v}_i, \sum_{j=1}^n (\mathbf{v}_j, \mathbf{w})\mathbf{v}_j\right)\\ &= \sum_{i,j=1}^n \overline{(\mathbf{v}_i, \mathbf{v})}(\mathbf{v}_j, \mathbf{w})(\mathbf{v}_i, \mathbf{v}_j)\\ &= \sum_{i,j=1}^n \overline{(\mathbf{v}_i, \mathbf{v})}(\mathbf{v}_j, \mathbf{w})\delta_{ij}\\ &= \sum_{i=1}^n \overline{(\mathbf{v}_i, \mathbf{v})}(\mathbf{v}_i, \mathbf{w}).\end{aligned}$$

□

8.2 Gram-Schmidt orthogonalization

Theorem (Gram-Schmidt process)**.** Let V be an inner product space and $\mathbf{e}_1, \mathbf{e}_2, \cdots$ a linearly independent set. Then we can construct an orthonormal set $\mathbf{v}_1, \mathbf{v}_2, \cdots$ with the property that

$$\langle \mathbf{v}_1, \cdots, \mathbf{v}_k\rangle = \langle \mathbf{e}_1, \cdots, \mathbf{e}_k\rangle$$

for every k.

Proof. We construct it iteratively, and prove this by induction on k. The base case $k = 0$ is contentless.

Suppose we have already found $\mathbf{v}_1, \cdots, \mathbf{v}_k$ that satisfies the properties. We define

$$\mathbf{u}_{k+1} = \mathbf{e}_{k+1} - \sum_{i=1}^k (\mathbf{v}_i, \mathbf{e}_{i+1})\mathbf{v}_i.$$

We want to prove that this is orthogonal to all the other $\mathbf{v}_i$'s for $i \leq k$. We have

$$(\mathbf{v}_j, \mathbf{u}_{k+1}) = (\mathbf{v}_j, \mathbf{e}_{k+1}) - \sum_{i=1}^k (\mathbf{v}_i, \mathbf{e}_{k+1})\delta_{ij} = (\mathbf{v}_j, \mathbf{e}_{k+1}) - (\mathbf{v}_j, \mathbf{e}_{k+1}) = 0.$$

So it is orthogonal.

We want to argue that $\mathbf{u}_{k+1}$ is non-zero. Note that

$$\langle \mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{u}_{k+1}\rangle = \langle \mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{e}_{k+1}\rangle$$

since we can recover $\mathbf{e}_{k+1}$ from $\mathbf{v}_1, \cdots, \mathbf{v}_k$ and $\mathbf{u}_{k+1}$ by construction. We also know

$$\langle \mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{e}_{k+1}\rangle = \langle \mathbf{e}_1, \cdots, \mathbf{e}_k, \mathbf{e}_{k+1}\rangle$$

by assumption. We know $\langle \mathbf{e}_1, \cdots, \mathbf{e}_k, \mathbf{e}_{k+1}\rangle$ has dimension $k+1$ since the $\mathbf{e}_i$ are linearly independent. So we must have $\mathbf{u}_{k+1}$ non-zero, or else $\langle \mathbf{v}_1, \cdots, \mathbf{v}_k\rangle$ will be a set of size k spanning a space of dimension $k+1$, which is clearly nonsense.

Therefore, we can define

$$\mathbf{v}_{k+1} = \frac{\mathbf{u}_{k+1}}{\|\mathbf{u}_{k+1}\|}.$$

Then $\mathbf{v}_1, \cdots, \mathbf{v}_{k+1}$ is orthonormal and $\langle \mathbf{v}_1, \cdots, \mathbf{v}_{k+1}\rangle = \langle \mathbf{e}_1, \cdots, \mathbf{e}_{k+1}\rangle$ as required. □

Corollary. If V is a finite-dimensional inner product space, then any orthonormal set can be extended to an orthonormal basis.

Proof. Let $\mathbf{v}_1, \cdots, \mathbf{v}_k$ be an orthonormal set. Since this is linearly independent, we can extend it to a basis $(\mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{x}_{k+1}, \cdots, \mathbf{x}_n)$.

We now apply the Gram-Schmidt process to this basis to get an orthonormal basis of V, say $(\mathbf{u}_1, \cdots, \mathbf{u}_n)$. Moreover, we can check that the process does not modify our $\mathbf{v}_1, \cdots, \mathbf{v}_k$, i.e. $\mathbf{u}_i = \mathbf{v}_i$ for $1 \leq i \leq k$. So done. □

Proposition. Let V be a finite-dimensional inner product space, and $W \leq V$. Then

$$V = W \perp W^\perp.$$

Proof. There are three things to prove, and we know (iii) implies (ii). Also, (iii) is obvious by definition of $W^\perp$. So it remains to prove (i), i.e. $V = W + W^\perp$.

Let $\mathbf{w}_1, \cdots, \mathbf{w}_k$ be an orthonormal basis for W, and pick $\mathbf{v} \in V$. Now let

$$\mathbf{w} = \sum_{i=1}^k (\mathbf{w}_i, \mathbf{v})\mathbf{w}_i.$$

Clearly, we have $\mathbf{w} \in W$. So we need to show $\mathbf{v} - \mathbf{w} \in W^\perp$. For each j, we can compute

$$\begin{aligned}(\mathbf{w}_j, \mathbf{v} - \mathbf{w}) &= (\mathbf{w}_j, \mathbf{v}) - \sum_{i=1}^k (\mathbf{w}_i, \mathbf{v})(\mathbf{w}_j, \mathbf{w}_i)\\ &= (\mathbf{w}_j, \mathbf{v}) - \sum_{i=1}^k (\mathbf{w}_i, \mathbf{v})\delta_{ij}\\ &= 0.\end{aligned}$$

Hence for any λ_i, we have

$$\left(\sum \lambda_j \mathbf{w}_j, \mathbf{v} - \mathbf{w}\right) = 0.$$

So we have $\mathbf{v} - \mathbf{w} \in W^\perp$. So done. □

Proposition. Let V be a finite-dimensional inner product space and $W \leq V$. Let $(\mathbf{e}_1, \cdots, \mathbf{e}_k)$ be an orthonormal basis of W. Let π be the orthonormal projection of V onto W, i.e. $\pi : V \to W$ is a function that satisfies $\ker \pi = W^\perp$, $\pi|_W = \mathrm{id}$. Then

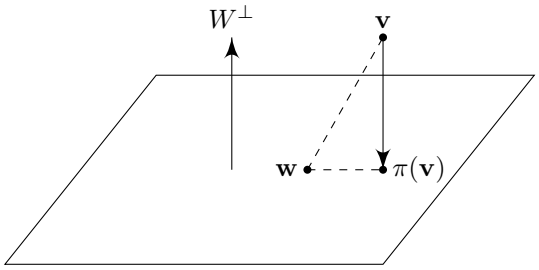
(i) π is given by the formula

$$\pi(\mathbf{v}) = \sum_{i=1}^k (\mathbf{e}_i, \mathbf{v})\mathbf{e}_i.$$

(ii) For all $\mathbf{v} \in V, \mathbf{w} \in W$, we have

$$\|\mathbf{v} - \pi(\mathbf{v})\| \leq \|\mathbf{v} - \mathbf{w}\|,$$

with equality if and only if $\pi(\mathbf{v}) = \mathbf{w}$. This says $\pi(\mathbf{v})$ is the point on W that is closest to $\mathbf{v}$.



Proof.

(i) Let $\mathbf{v} \in V$, and define

$$\mathbf{w} = \sum_{i=1} (\mathbf{e}_i, \mathbf{v})\mathbf{e}_i.$$

We want to show this is $\pi(\mathbf{v})$. We need to show $\mathbf{v} - \mathbf{w} \in W^\perp$. We can compute

$$(\mathbf{e}_j, \mathbf{v} - \mathbf{w}) = (\mathbf{e}_j, \mathbf{v}) - \sum_{i=1}^{k} (\mathbf{e}_i, \mathbf{v})(\mathbf{e}_j, \mathbf{e}_i) = 0.$$

So $\mathbf{v} - \mathbf{w}$ is orthogonal to every basis vector in $\mathbf{w}$, i.e. $\mathbf{v} - \mathbf{w} \in W^\perp$.So

$$\pi(\mathbf{v}) = \pi(\mathbf{w}) + \pi(\mathbf{v} - \mathbf{w}) = \mathbf{w}$$

as required.

(ii) This is just Pythagoras' theorem. Note that if $\mathbf{x}$ and $\mathbf{y}$ are orthogonal, then

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) \\ &= (\mathbf{x}, \mathbf{x}) + (\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{x}) + (\mathbf{y}.\mathbf{y}) \\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2. \end{aligned}$$

We apply this to our projection. For any $\mathbf{w} \in W$, we have

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v} - \pi(\mathbf{v})\|^2 + \|\pi(\mathbf{v}) - \mathbf{w}\|^2 \geq \|\mathbf{v} - \pi(\mathbf{v})\|^2$$

with equality if and only if $\|\pi(\mathbf{v}) - \mathbf{w}\| = 0$, i.e. $\pi(\mathbf{v}) = \mathbf{w}$. □

8.3 Adjoints, orthogonal and unitary maps

Lemma. Let V and W be finite-dimensional inner product spaces and $\alpha : V \to W$ is a linear map. Then there exists a unique linear map $\alpha^* : W \to V$ such that

$$(\alpha\mathbf{v}, \mathbf{w}) = (\mathbf{v}, \alpha^*\mathbf{w}) \qquad (*)$$

for all $\mathbf{v} \in V$, $\mathbf{w} \in W$.

Proof. There are two parts. We have to prove existence and uniqueness. We'll first prove it concretely using matrices, and then provide a conceptual reason of what this means.

Let $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ and $(\mathbf{w}_1, \cdots, \mathbf{w}_m)$ be orthonormal basis for V and W. Suppose α is represented by A.

To show uniqueness, suppose $\alpha^* : W \to V$ satisfies $(\alpha\mathbf{v}, \mathbf{w}) = (\mathbf{v}, \alpha^*\mathbf{w})$ for all $\mathbf{v} \in V$, $\mathbf{w} \in W$, then for all i, j, by definition, we know

$$\begin{aligned}(\mathbf{v}_i, \alpha^*(\mathbf{w}_j)) &= (\alpha(\mathbf{v}_i), \mathbf{w}_j)\\ &= \left(\sum_k A_{ki}\mathbf{w}_k, \mathbf{w}_j\right)\\ &= \sum_k \bar{A}_{ki}(\mathbf{w}_k, \mathbf{w}_j) = \bar{A}_{ji}.\end{aligned}$$

So we get

$$\alpha^*(\mathbf{w}_j) = \sum_i (\mathbf{v}_i, \alpha^*(\mathbf{w}_j))\mathbf{v}_i = \sum_i \bar{A}_{ji}\mathbf{v}_i.$$

Hence α^* must be represented by $A^\dagger$. So α^* is unique.

To show existence, all we have to do is to show $A^\dagger$ indeed works. Now let α^* be represented by $A^\dagger$. We can compute the two sides of $(*)$ for arbitrary $\mathbf{v}, \mathbf{w}$. We have

$$\begin{aligned}\left(\alpha\left(\sum \lambda_i\mathbf{v}_i\right), \sum \mu_j\mathbf{w}_j\right) &= \sum_{i,j} \bar{\lambda}_i\mu_j(\alpha(\mathbf{v}_i), \mathbf{w}_j)\\ &= \sum_{i,j} \bar{\lambda}_i\mu_j\left(\sum_k A_{ki}\mathbf{w}_k, \mathbf{w}_j\right)\\ &= \sum_{i,j} \bar{\lambda}_i\bar{A}_{ji}\mu_j.\end{aligned}$$

We can compute the other side and get

$$\begin{aligned}\left(\sum \lambda_i\mathbf{v}_i, \alpha^*\left(\sum \mu_j\mathbf{w}_j\right)\right) &= \sum_{i,j} \bar{\lambda}_i\mu_j\left(\mathbf{v}_i, \sum_k A^\dagger_{kj}\mathbf{v}_k\right)\\ &= \sum_{i,j} \bar{\lambda}_i\bar{A}_{ji}\mu_j.\end{aligned}$$

So done. □

Lemma. Let V be a finite-dimensional space and $\alpha \in \mathrm{End}(V)$. Then α is orthogonal if and only if $\alpha^{-1} = \alpha^*$.

Proof. ($\Leftarrow$) Suppose $\alpha^{-1} = \alpha^*$. If $\alpha^{-1} = \alpha^*$, then

$$(\alpha\mathbf{v}, \alpha\mathbf{v}) = (\mathbf{v}, \alpha^*\alpha\mathbf{v}) = (\mathbf{v}, \alpha^{-1}\alpha\mathbf{v}) = (\mathbf{v}, \mathbf{v}).$$

($\Rightarrow$) If α is orthogonal and $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ is an orthonormal basis for V, then for $1 \le i, j \le n$, we have

$$\delta_{ij} = (\mathbf{v}_i, \mathbf{v}_j) = (\alpha\mathbf{v}_i, \alpha\mathbf{v}_j) = (\mathbf{v}_i, \alpha^*\alpha\mathbf{v}_j).$$

So we know

$$\alpha^*\alpha(\mathbf{v}_j) = \sum_{i=1}^n (\mathbf{v}_i, \alpha^*\alpha\mathbf{v}_j))\mathbf{v}_i = \mathbf{v}_j.$$

So by linearity of $\alpha^*\alpha$, we know $\alpha^*\alpha = \mathrm{id}_V$. So $\alpha^* = \alpha^{-1}$. □

Corollary. $\alpha \in \mathrm{End}(V)$ is orthogonal if and only if α is represented by an orthogonal matrix, i.e. a matrix A such that $A^TA = AA^T = I$, with respect to any orthonormal basis.

Proof. Let $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ be an orthonormal basis for V. Then suppose α is represented by A. So α^* is represented by A^T. Then $A^* = A^{-1}$ if and only if $AA^T = A^TA = I$. □

Proposition. Let V be a finite-dimensional real inner product space and $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ is an orthonormal basis of V. Then there is a bijection

$$\begin{aligned} \mathrm{O}(V) &\to \{\text{orthonormal basis for } V\} \\ \alpha &\mapsto (\alpha(\mathbf{e}_1, \cdots, \mathbf{e}_n)). \end{aligned}$$

Proof. Same as the case for general vector spaces and general bases. □

Lemma. Let V be a finite dimensional complex inner product space and $\alpha \in \mathrm{End}(V)$. Then α is unitary if and only if α is invertible and $\alpha^* = \alpha^{-1}$.

Corollary. $\alpha \in \mathrm{End}(V)$ is unitary if and only if α is represented by a unitary matrix A with respect to any orthonormal basis, i.e. $A^{-1} = A^\dagger$.

Proposition. Let V be a finite-dimensional complex inner product space. Then there is a bijection

$$\begin{aligned} U(V) &\to \{\text{orthonormal basis of } V\} \\ \alpha &\mapsto \{\alpha(\mathbf{e}_1), \cdots, \alpha(\mathbf{e}_n)\}. \end{aligned}$$

8.4 Spectral theory

Lemma. Let V be a finite-dimensional inner product space, and $\alpha \in \mathrm{End}(V)$ self-adjoint. Then

(i) α has a real eigenvalue, and all eigenvalues of α are real.

(ii) Eigenvectors of α with distinct eigenvalues are orthogonal.

Proof. We are going to do real and complex cases separately.

(i) Suppose first V is a complex inner product space. Then by the fundamental theorem of algebra, α has an eigenvalue, say λ. We pick $\mathbf{v} \in V \setminus \{0\}$ such that $\alpha\mathbf{v} = \lambda\mathbf{v}$. Then

$$\bar{\lambda}(\mathbf{v}, \mathbf{v}) = (\lambda\mathbf{v}, \mathbf{v}) = (\alpha\mathbf{v}, \mathbf{v}) = (\mathbf{v}, \alpha\mathbf{v}) = (\mathbf{v}, \lambda\mathbf{v}) = \lambda(\mathbf{v}, \mathbf{v}).$$

Since $\mathbf{v} \neq \mathbf{0}$, we know $(\mathbf{v}, \mathbf{v}) \neq 0$. So $\lambda = \bar{\lambda}$.

For the real case, we pretend we are in the complex case. Let $\mathbf{e}_1, \cdots, \mathbf{e}_n$ be an orthonormal basis for V. Then α is represented by a symmetric matrix A (with respect to this basis). Since real symmetric matrices are Hermitian viewed as complex matrices, this gives a self-adjoint endomorphism of $\mathbb{C}^n$. By the complex case, A has real eigenvalues only. But the eigenvalues of A are the eigenvalues of α and $M_A(t) = M_\alpha(t)$. So done.

Alternatively, we can prove this without reducing to the complex case. We know every irreducible factor of $M_\alpha(t)$ in $\mathbb{R}[t]$ must have degree 1 or 2, since the roots are either real or come in complex conjugate pairs. Suppose $f(t)$ were an irreducible factor of degree 2. Then

$$\left(\frac{m_\alpha}{f}\right)(\alpha) \neq 0$$

since it has degree less than the minimal polynomial. So there is some $\mathbf{v} \in V$ such that

$$\left(\frac{M_\alpha}{f}\right)(\alpha)(\mathbf{v}) \neq \mathbf{0}.$$

So it must be that $f(\alpha)(\mathbf{v}) = \mathbf{0}$. Let $U = \langle \mathbf{v}, \alpha(\mathbf{v})\rangle$. Then this is an α-invariant subspace of V since f has degree 2.

Now $\alpha|_U \in \mathrm{End}(U)$ is self-adjoint. So if $(\mathbf{e}_1, \mathbf{e}_2)$ is an orthonormal basis of U, then α is represented by a real symmetric matrix, say

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix}$$

But then $\chi_{\alpha|_U}(t) = (t-a)^2 - b^2$, which has real roots, namely $a \pm b$. This is a contradiction, since $M_{\alpha|_U} = f$, but f is irreducible.

(ii) Now suppose $\alpha\mathbf{v} = \lambda\mathbf{v}$, $\alpha\mathbf{w} = \mu\mathbf{w}$ and $\lambda \neq \mu$. We need to show $(\mathbf{v}, \mathbf{w}) = 0$. We know

$$(\alpha\mathbf{v}, \mathbf{w}) = (\mathbf{v}, \alpha\mathbf{w})$$

by definition. This then gives

$$\lambda(\mathbf{v}, \mathbf{w}) = \mu(\mathbf{v}, \mathbf{w})$$

Since $\lambda \neq \mu$, we must have $(\mathbf{v}, \mathbf{w}) = 0$. □

Theorem. Let V be a finite-dimensional inner product space, and $\alpha \in \mathrm{End}(V)$ self-adjoint. Then V has an orthonormal basis of eigenvectors of α.

Proof. By the previous lemma, α has a real eigenvalue, say λ. Then we can find an eigenvector $\mathbf{v} \in V \setminus \{0\}$ such that $\alpha\mathbf{v} = \lambda\mathbf{v}$.

Let $U = \langle \mathbf{v}\rangle^\perp$. Then we can write

$$V = \langle \mathbf{v}\rangle \perp U.$$

We now want to prove α sends U into U. Suppose $\mathbf{u} \in U$. Then

$$(\mathbf{v}, \alpha(\mathbf{u})) = (\alpha\mathbf{v}, \mathbf{u}) = \lambda(\mathbf{v}, \mathbf{u}) = 0.$$

So $\alpha(\mathbf{u}) \in \langle \mathbf{v}\rangle^\perp = U$. So $\alpha|_U \in \mathrm{End}(U)$ and is self-adjoint.

By induction on $\dim V$, U has an orthonormal basis $(\mathbf{v}_2, \cdots, \mathbf{v}_n)$ of α eigenvectors. Now let

$$\mathbf{v}_1 = \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

Then $(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)$ is an orthonormal basis of eigenvectors for α. □

Corollary. Let V be a finite-dimensional vector space and α self-adjoint. Then V is the orthogonal (internal) direct sum of its α-eigenspaces.

Corollary. Let $A \in \mathrm{Mat}_n(\mathbb{R})$ be symmetric. Then there exists an orthogonal matrix P such that $P^TAP = P^{-1}AP$ is diagonal.

Proof. Let $(\,\cdot\,,\,\cdot\,)$ be the standard inner product on $\mathbb{R}^n$. Then A is self-adjoint as an endomorphism of $\mathbb{R}^n$. So $\mathbb{R}^n$ has an orthonormal basis of eigenvectors for A, say $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$. Taking $P = (\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n)$ gives the result. □

Corollary. Let V be a finite-dimensional real inner product space and $\psi : V \times V \to \mathbb{R}$ a symmetric bilinear form. Then there exists an orthonormal basis $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ for V with respect to which ψ is represented by a diagonal matrix.

Proof. Let $(\mathbf{u}_1, \cdots, \mathbf{u}_n)$ be any orthonormal basis for V. Then ψ is represented by a symmetric matrix A. Then there exists an orthogonal matrix P such that P^TAP is diagonal. Now let $\mathbf{v}_i = \sum P_{ki}\mathbf{u}_k$. Then $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ is an orthonormal basis since

$$\begin{aligned}(\mathbf{v}_i, \mathbf{v}_j) &= \Big(\sum P_{ki}\mathbf{u}_k, \sum P_{\ell j}\mathbf{u}_\ell\Big)\\ &= \sum P^T_{ik}(\mathbf{u}_k, \mathbf{u}_\ell)P_{\ell j}\\ &= [P^TP]_{ij}\\ &= \delta_{ij}.\end{aligned}$$

Also, ψ is represented by P^TAP with respect to $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$. □

Corollary. Let V be a finite-dimensional real vector space and ϕ, ψ symmetric bilinear forms on V such that ϕ is positive-definite. Then we can find a basis $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ for V such that both ϕ and ψ are represented by diagonal matrices with respect to this basis.

Proof. We use ϕ to define an inner product. Choose an orthonormal basis for V (equipped with ϕ) $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ with respect to which ψ is diagonal. Then ϕ is represented by I with respect to this basis, since $\psi(\mathbf{v}_i, \mathbf{v}_j) = \delta_{ij}$. So done. □

Corollary. If $A, B \in \mathrm{Mat}_n(\mathbb{R})$ are symmetric and A is positive definitive (i.e. $\mathbf{v}^TA\mathbf{v} > 0$ for all $\mathbf{v} \in \mathbb{R}^n \setminus \{0\}$). Then there exists an invertible matrix Q such that Q^TAQ and Q^TBQ are both diagonal.

Proposition.

(i) If $A \in \mathrm{Mat}_n(\mathbb{C})$ is Hermitian, then there exists a unitary matrix $U \in \mathrm{Mat}_n(\mathbb{C})$ such that
$$U^{-1}AU = U^\dagger AU$$
is diagonal.

(ii) If ψ is a Hermitian form on a finite-dimensional complex inner product space V, then there is an orthonormal basis for V diagonalizing ψ.

(iii) If ϕ, ψ are Hermitian forms on a finite-dimensional complex vector space and ϕ is positive definite, then there exists a basis for which ϕ and ψ are diagonalized.

(iv) Let $A, B \in \mathrm{Mat}_n(\mathbb{C})$ be Hermitian, and A positive definitive (i.e. $\mathbf{v}^\dagger A\mathbf{v} > 0$ for $\mathbf{v} \in V \setminus \{0\}$). Then there exists some invertible Q such that $Q^\dagger AQ$ and $Q^\dagger BQ$ are diagonal.

Theorem. Let V be a finite-dimensional complex vector space and $\alpha \in U(V)$ be unitary. Then V has an orthonormal basis of α eigenvectors.

Proof. By the fundamental theorem of algebra, there exists $\mathbf{v} \in V \setminus \{0\}$ and $\lambda \in \mathbb{C}$ such that $\alpha\mathbf{v} = \lambda\mathbf{v}$. Now consider $W = \langle\mathbf{v}\rangle^\perp$. Then

$$V = W \perp \langle\mathbf{v}\rangle.$$

We want to show α restricts to a (unitary) endomorphism of W. Let $\mathbf{w} \in W$. We need to show $\alpha(\mathbf{w})$ is orthogonal to $\mathbf{v}$. We have

$$(\alpha\mathbf{w}, \mathbf{v}) = (\mathbf{w}, \alpha^{-1}\mathbf{v}) = (\mathbf{w}, \lambda^{-1}\mathbf{v}) = 0.$$

So $\alpha(\mathbf{w}) \in W$ and $\alpha|_W \in \mathrm{End}(W)$. Also, $\alpha|_W$ is unitary since α is. So by induction on $\dim V$, W has an orthonormal basis of α eigenvectors. If we add $\mathbf{v}/\|\mathbf{v}\|$ to this basis, we get an orthonormal basis of V itself comprised of α eigenvectors. □