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. $\square$

1.3 Direct sums

We are going to define direct sums in many ways in order to confuse students.

Definition ((Internal) direct sum)**.** Suppose V is a vector space over $\mathbb{F}$ and $U, W \subseteq V$ are subspaces. We say that V is the *(internal) direct sum* of U and W if

(i) $U + W = V$

(ii) $U \cap W = 0$.

We write $V = U \oplus W$.

Equivalently, this requires that every $\mathbf{v} \in V$ can be written uniquely as $\mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in U, \mathbf{w} \in W$. We say that U and W are *complementary subspaces* of V.

You will show in the example sheets that given any subspace $U \subseteq V$, U must have a complementary subspace in V.

Example. Let $V = \mathbb{R}^2$, and $U = \langle\begin{pmatrix}0\\1\end{pmatrix}\rangle$. Then $\langle\begin{pmatrix}1\\1\end{pmatrix}\rangle$ and $\langle\begin{pmatrix}1\\0\end{pmatrix}\rangle$ are both complementary subspaces to U in V.

Definition ((External) direct sum)**.** If U, W are vector spaces over $\mathbb{F}$, the *(external) direct sum* is

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

with addition and scalar multiplication componentwise:

$$(\mathbf{u}_1, \mathbf{w}_1) + (\mathbf{u}_2, \mathbf{w}_2) = (\mathbf{u}_1 + \mathbf{u}_2, \mathbf{w}_1 + \mathbf{w}_2), \quad \lambda(\mathbf{u}, \mathbf{w}) = (\lambda\mathbf{u}, \lambda\mathbf{w}).$$

(ii) $AE^n_{Ij}(\lambda)$ is obtained by adding $\lambda\times$ column i to column j.

(iii) $AT^n_i(\lambda)$ is obtained from A by rescaling the ith column by λ.

Multiplying on the left instead of the right would result in the same operations performed on the rows instead of the columns.

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

We are going to start with A, and then apply these operations to get it into this form.

Proof. We claim that there are elementary matrices $E^m_1, \cdots, E^m_a$ and $F^n_1, \cdots, F^n_b$ (these E are not necessarily the shears, but any elementary matrix) such that

$$E^m_1 \cdots E^m_a A F^n_1 \cdots F^n_b = \begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}$$

This suffices since the $E^m_i \in \mathrm{GL}_M(\mathbb{F})$ and $F^n_j \in \mathrm{GL}_n(\mathbb{F})$. 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. $\square$

It is an exercise to show that the row and column operations do not change the row rank or column rank, and deduce that they are equal.

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

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

So A^T represents ψ_L.

We also have

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

So

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

Definition (Left and right kernel)**.** The kernel of ψ_L is *left kernel* of ψ, while the kernel of ψ_R is the *right kernel* of ψ.

Then by definition, $\mathbf{v}$ is in the left kernel if $\psi(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{w} \in W$.

More generally, if $T \subseteq V$, then we write

$$T^\perp = \{\mathbf{w} \in W : \psi(\mathbf{t}, \mathbf{w}) = 0 \text{ for all } \mathbf{t} \in T\}.$$

Similarly, if $U \subseteq W$, then we write

$${}^\perp U = \{\mathbf{v} \in V : \psi(\mathbf{v}, \mathbf{u}) = 0 \text{ for all } \mathbf{u} \in U\}.$$

In particular, $V^\perp = \ker \psi_R$ and ${}^\perp W = \ker \psi_L$.

If we have a non-trivial left (or right) kernel, then in some sense, some elements in V (or W) are "useless", and we don't like these.

Definition (Non-degenerate bilinear form)**.** ψ is *non-degenerate* if the left and right kernels are both trivial. We say ψ is *degenerate* otherwise.

Definition (Rank of bilinear form)**.** If $\psi : V \to W$ is a bilinear form $\mathbb{F}$ on a finite-dimensional vector space V, then the *rank* of V is the rank of any matrix representing ϕ. This is well-defined since $r(P^TAQ) = r(A)$ if P and Q are invertible.

Alternatively, it is the rank of ψ_L (or ψ_R).

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.

We can understand this as saying if there are too many things in V (or W), then some of them are bound to be useless.

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. $\square$

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

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

To motivate this definition, we need a notion of volume. How can we define *volume* on a vector space? It should be clear that the "volume" cannot be uniquely determined, since it depends on what units we are using. For example, saying the volume is "1" is meaningless unless we provide the units, e.g. cm^3. So we have an axiomatic definition for what it means for something to denote a "volume".

Definition (Volume form)**.** A *volume form* on $\mathbb{F}^n$ is a function $d: \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$ that is

(i) Multilinear, i.e. for all i and all $\mathbf{v}_1, \cdots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \cdots, \mathbf{v}_n \in \mathbb{F}^n$, we have

$$\mathrm{d}(\mathbf{v}_1, \cdots, \mathbf{v}_{i-1}, \cdot, \mathbf{v}_{i+1}, \cdots, \mathbf{v}_n) \in (\mathbb{F}^n)^*.$$

(ii) Alternating, i.e. if $\mathbf{v}_i = \mathbf{v}_j$ for some $i \neq j$, then

$$d(\mathbf{v}_1, \cdots, \mathbf{v}_n) = 0.$$

We should think of $d(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ as the n-dimensional volume of the parallelopiped spanned by $\mathbf{v}_1, \cdots, \mathbf{v}_n$.

We can view $A \in \mathrm{Mat}_n(\mathbb{F})$ as n-many vectors in $\mathbb{F}^n$ by considering its columns $A = (A^{(1)}\ A^{(2)}\ \cdots\ A^{(n)})$, with $A^{(i)} \in \mathbb{F}^n$. Then we have

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

$\sigma = \sigma_1\sigma_2$, 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.

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

Our initial strategy is to identify basis-independent invariants for endomorphisms. For example, we will show that the rank, trace, determinant and characteristic polynomial are all such invariants.

Recall that the trace of a matrix $A \in \mathrm{Mat}_n(\mathbb{F})$ is the sum of the diagonal elements:

Definition (Trace)**.** The *trace* of a matrix of $A \in \mathrm{Mat}_n(\mathbb{F})$ is defined by

$$\operatorname{tr} A = \sum_{i=1}^{n} A_{ii}.$$

We want to show that the trace is an invariant. In fact, we will show a stronger statement (as well as the corresponding statement for determinants):

Lemma.

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

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

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

(iii) If $A, B \in \mathrm{Mat}_n(\mathbb{F})$ 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

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

This allows us to define the trace and determinant of an *endomorphism*.

Definition (Trace and determinant of endomorphism)**.** Let $\alpha \in \mathrm{End}(V)$, and A be a matrix representing α under any basis. Then the *trace* of α is $\operatorname{tr} \alpha = \operatorname{tr} A$, and the *determinant* is $\det \alpha = \det A$.

The lemma tells us that the determinant and trace are well-defined. We can also define the determinant without reference to a basis, by defining more general volume forms and define the determinant as a scaling factor.

The trace is slightly more tricky to define without basis, but in IB Analysis II example sheet 4, you will find that it is the directional derivative of the determinant at the origin.

To talk about the characteristic polynomial, we need to know what eigenvalues are.

We can use the last lemma and induction to show that any non-zero $f \in \mathbb{F}[t]$ can be written as

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

where $\lambda_1, \cdots, \lambda_k$ are all distinct, $a_i > 1$, and g is a polynomial with no roots in $\mathbb{F}$.

Hence we obtain the following:

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.

More importantly, we have the following:

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

We will not prove this.

We say $\mathbb{C}$ is an *algebraically closed field*.

It thus follows that every polynomial over $\mathbb{C}$ of degree $n > 0$ has precisely n roots, counted with multiplicity, since if we write $f(t) = g(t)\prod(t-\lambda_i)^{a_i}$ and g has no roots, then g is constant. So the number of roots is $\sum a_i = \deg f$, counted with multiplicity.

It also follows that every polynomial in $\mathbb{R}$ factors into linear polynomials and quadratic polynomials with no real roots (since complex roots of real polynomials come in complex conjugate pairs).

6.2.2 Minimal polynomial

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

$$f(A) = \sum_{i=0}^m a_i A^i, \quad f(\alpha) = \sum_{i=0}^m a_i \alpha^i$$

where $A^0 = I$ and $\alpha^0 = \iota$.

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

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

We can now re-state our diagonalizability theorem.

Theorem (Diagonalizability theorem 2.0)**.** Let $\alpha \in \operatorname{End}(V)$. 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. $\square$

Theorem. Let $\alpha, \beta \in \operatorname{End}(V)$ 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 $\alpha\beta = \beta\alpha$.

This is important in quantum mechanics. This means that if two operators do not commute, then they do not have a common eigenbasis. Hence we have the uncertainty principle.

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 $\beta|_{E_i} \in \operatorname{End}(E_i)$. 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. $\square$

6.3 The Cayley-Hamilton theorem

We will first state the theorem, and then prove it later.

Recall that $\chi_\alpha(t) = \det(t\iota - \alpha)$ for $\alpha \in \operatorname{End}(V)$. Our main theorem of the section (as you might have guessed from the title) is

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

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 $\chi_\alpha(A) = 0$. □

This proof suggests that we *really* ought to be able to just substitute in $t = \alpha$ and be done. In fact, we can do this, after we develop sufficient machinery. This will be done in the IB Groups, Rings and Modules course.

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 $M_\alpha \mid \chi_\alpha$.

with $\lambda, \mu \in \mathbb{C}$ distinct. We see that M_A determines the Jordan normal form of A, but χ_A does not.

Every 3×3 matrix in Jordan normal form is one of the six types. Here λ_1, λ_2 and λ_3 are distinct complex numbers.

Jordan normal form	χ_A	M_A
$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$	$(t-\lambda_1)(t-\lambda_2)(t-\lambda_3)$	$(t-\lambda_1)(t-\lambda_2)(t-\lambda_3)$
$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}$	$(t-\lambda_1)^2(t-\lambda_2)$	$(t-\lambda_1)(t-\lambda_2)$
$\begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{pmatrix}$	$(t-\lambda_1)^2(t-\lambda_2)$	$(t-\lambda_1)^2(t-\lambda_2)$
$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix}$	$(t-\lambda_1)^3$	$(t-\lambda_1)$
$\begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{pmatrix}$	$(t-\lambda_1)^3$	$(t-\lambda_1)^2$
$\begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{pmatrix}$	$(t-\lambda_1)^3$	$(t-\lambda_1)^3$

Notice that χ_A and M_A together determine the Jordan normal form of a 3×3 complex matrix. We do indeed need χ_A in the second case, since if we are given $M_A = (t-\lambda_1)(t-\lambda_2)$, we know one of the roots is double, but not which one.

In general, though, even χ_A and M_A together does not suffice.

We now want to understand the Jordan normal blocks better. Recall the definition

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

If $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ is the standard basis for $\mathbb{C}^n$, we have

$$J_n(0)(\mathbf{e}_1) = 0, \quad J_n(0)(\mathbf{e}_i) = \mathbf{e}_{i-1} \text{ for } 2 \leq i \leq n.$$

Thus we know

$$J_n(0)^k(\mathbf{e}_i) = \begin{cases} 0 & i \leq k \\ \mathbf{e}_{i-k} & k < i \leq n \end{cases}$$

In other words, for $k < n$, we have

$$(J_n(\lambda) - \lambda I)^k = J_n(0)^k = \begin{pmatrix} 0 & I_{n-k} \\ 0 & 0 \end{pmatrix}.$$

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

$$n((\alpha - \lambda\iota)^r) - n((\alpha - \lambda\iota)^{r-1}) = \text{ number of Jordan blocks } J_n(\lambda) \text{ with } n \geq r. \quad \square$$

We can interpret this result as follows: if $r \leq m$, when we take an additional power of $J_m(\lambda) - \lambda I_m$, we get from $\begin{pmatrix} 0 & I_{m-r} \\ 0 & 0 \end{pmatrix}$ to $\begin{pmatrix} 0 & I_{m-r-1} \\ 0 & 0 \end{pmatrix}$. So we kill off one more column in the matrix, and the nullity increase by one. This happens until $(J_m(\lambda) - \lambda I_m)^r = 0$, in which case increasing the power no longer affects the matrix. So when we look at the difference in nullity, we are counting the number of blocks that are affected by the increase in power, which is the number of blocks of size at least r.

We have now proved uniqueness, but existence is not yet clear. To show this, we will reduce it to the case where there is exactly one eigenvalue. This reduction is easy if the matrix is diagonalizable, because we can decompose the matrix into each eigenspace and then work in the corresponding eigenspace. In general, we need to work with "generalized eigenspaces".

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*.

We see that the diagonal matrix we get is not unique. We can re-scale our basis by any constant, and get an equivalent expression.

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. $\square$

Note that it follows that for the corresponding quadratic form q, we have

$$q\left(\sum_{i=1}^n a_i \mathbf{v}_i\right) = \sum_{i=1}^r a_i^2.$$

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

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

Now this theorem is a bit too strong, and we are going to fix that next lecture, by talking about Hermitian forms and sesquilinear forms. Before that, we do the equivalent result for real vector spaces.

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

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

Definition (Signature)**.** The *signature* of a bilinear form ϕ is the number $p - q$, where p and q are as above.

Of course, we can recover p and q from the signature and the rank of ϕ.

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

The above result was nice for real vector spaces. However, if ϕ is a bilinear form on a $\mathbb{C}$-vector space V, then $\phi(i\mathbf{v}, i\mathbf{v}) = -\phi(\mathbf{v}, \mathbf{v})$. So there can be no good notion of positive definiteness for complex bilinear forms. To make them work for complex vector spaces, we need to modify the definition slightly to obtain Hermitian forms.

Definition (Sesquilinear form)**.** Let V, W be complex vector spaces. Then a *sesquilinear form* is a function $\phi: V \times W \to \mathbb{C}$ such that

(i) $\phi(\lambda \mathbf{v}_1 + \mu \mathbf{v}_2, \mathbf{w}) = \bar{\lambda}\phi(\mathbf{v}_1, \mathbf{w}) + \bar{\mu}\phi(\mathbf{v}_2, \mathbf{w})$.

(ii) $\phi(\mathbf{v}, \lambda \mathbf{w}_1 + \mu \mathbf{w}_2) = \lambda\phi(\mathbf{v}, \mathbf{w}_1) + \mu\phi(\mathbf{v}\mathbf{w}_2)$.

for all $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in W$ and $\lambda, \mu \in \mathbb{C}$.

Note that some people have an opposite definition, where we have linearity in the first argument and conjugate linearity in the second.

These are called sesquilinear since "sesqui" means "one and a half", and this is linear in the second argument and "half linear" in the first.

Alternatively, to define a sesquilinear form, we can define a new complex vector space $\bar{V}$ structure on V by taking the same abelian group (i.e. the same underlying set and addition), but with the scalar multiplication $\mathbb{C} \times \bar{V} \to \bar{V}$ defined as

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

Then a sesquilinear form on $V \times W$ is a bilinear form on $\bar{V} \times W$. Alternatively, this is a linear map $W \to \bar{V}^*$.

Definition (Representation of sesquilinear form)**.** Let V, W be finite-dimensional complex vector spaces with basis $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ and $(\mathbf{w}_1, \cdots, \mathbf{w}_m)$ respectively, and $\phi: V \times W \to \mathbb{C}$ be a sesquilinear form. Then the matrix representing ϕ with respect to these bases is

$$A_{ij} = \phi(\mathbf{v}_i, \mathbf{w}_j).$$

for $1 \leq i \leq n, 1 \leq j \leq m$.

As usual, this determines the whole sesquilinear form. This follows from the analogous fact for the bilinear form on $\bar{V} \times W \to \mathbb{C}$. Let $\mathbf{v} = \sum \lambda_i \mathbf{v}_i$ and $W = \sum \mu_j w_j$. Then we have

$$\phi(\mathbf{v}, \mathbf{w}) = \sum_{i,j} \overline{\lambda}_i \mu_j \phi(\mathbf{v}_i, \mathbf{w}_j) = \lambda^\dagger A \mu.$$

8.2 Gram-Schmidt orthogonalization

As mentioned, we want to make sure every vector space has an orthonormal basis, and we can extend any orthonormal set to an orthonormal basis, at least in the case of finite-dimensional vector spaces. The idea is to start with an arbitrary basis, which we know exists, and produce an orthonormal basis out of it. The way to do this is the Gram-Schmidt process.

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.

Note that we are not requiring the set to be finite. We are just requiring it to be countable.

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.

(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

Adjoints

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

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

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

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

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

We can deduce similar results for complex finite-dimensional vector spaces, with the same proofs. In particular,

Proposition.

(i) If $A \in \text{Mat}_n(\mathbb{C})$ is Hermitian, then there exists a unitary matrix $U \in \text{Mat}_n(\mathbb{C})$ 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 $A, B \in \text{Mat}_n(\mathbb{C})$ 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.

That's all for self-adjoint matrices. How about unitary matrices?

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 $\alpha|_W \in \text{End}(W)$. 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. $\square$

This theorem and the analogous one for self-adjoint endomorphisms have a common generalization, at least for complex inner product spaces. The key fact that leads to the existence of an orthonormal basis of eigenvectors is that α and α^* commute. This is clearly a necessary condition, since if α is diagonalizable, then α^* is diagonal in the same basis (since it is just the transpose (and conjugate)), and hence they commute. It turns out this is also a sufficient condition, as you will show in example sheet 4.