# Further Linear Algebra

Kexing Ying

May 15, 2020

## Contents

# 1 Vector Spaces

We continue from last year and examine some properties about vector spaces.

We will from this point forward write $\sum \mu S$ as a shorthand for $\sum_{s\in S} \mu_s s$ for some suitable set $S$ and indexed value $\mu$. We will also write $T \in \mathrm{End}(V)$ for $T$ is an endomorphism of $V$, that is, a linear map $T : V \to V$.

## 1.1 Algebraic & Geometric Multiplicities

We recall some basic definitions and properties of eigenvectors.

**Definition 1.1.** Let $V$ be some vector space, $T \in \mathrm{End}(V)$ a linear map and $\lambda$ an eigenvalue of $T$. Then the $\lambda$-eigenspace of $T$ is the subspace of $V$,

$$E_\lambda := \{v \in V \mid (\lambda I_V - T)v = \mathbf{0}\}.$$

We see that this is a subspace as it is the kernel of the linear map $\lambda I_V - T$.

**Theorem 1.** Let $V$ be some vector space, $T \in \mathrm{End}(V)$ a linear map. Suppose that $\{v_1, \cdots, v_k\}$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \cdots, \lambda_k$, then it is linearly independent.

*Proof.* We will prove by contrapositive. Suppose that $\{v_1, \cdots, v_k\}$ are eigenvectors that are linearly independent. Then by definition, there exists a minimal set of $\{\mu_i \mid i \in I\}$, such that $\sum_{i\in I} \mu_i v_i = 0$ (we see that $\mu_i \neq 0$ for all $i$ as otherwise it is not minimal). Now, let $j \in I$, then by rewriting, we have $v_j = \sum_{i\neq j} \mu_i' v_i$. Thus,

$$\lambda_j \sum_{i\neq j} \mu_i' v_i = \lambda_j v_j = T(v_j) = T\left(\sum_{i\neq j} \mu_i' v_i\right) = \sum_{i\neq j} \mu_i' T(v_i) = \sum_{i\neq j} \mu_i' \lambda_i v_i.$$

So, by rearranging, $0 = \sum_{i\neq j} (\lambda_i - \lambda_j)\mu_i' v_i$. Now, if for all $i \neq j$, $\lambda_i \neq \lambda_j$, we have found a smaller subset of $\{v_1, \cdots, v_k\}$ that is linearly dependent, contradicting our assumption, so there must be some $i$ such that $\lambda_i = \lambda_j$. □

**Corollary 1.1.** Let $V$ be a $n$-dimensional vector space. Then if the characteristic polynomial of the linear map $T \in \mathrm{End}(V)$ has $n$ distinct roots, then $T$ is diagonalisable.

We define *algebraic* and *geometric* multiplicity for eigenvalues.

**Definition 1.2** (Algebraic and Geometric Multiplicity)**.** Let $T \in \mathrm{End}(V)$ be a linear map with characteristic polynomial $\chi_T$, such that $\chi_T(\lambda) = 0$ (i.e. $\lambda$ is an eigenvalue of $T$).

The algebraic multiplicity of $\lambda$ is the number $a(\lambda)$ such that

$$\chi_T(x) = (x - \lambda)^{a(\lambda)} q(x),$$

for some polynomial $q(x)$ where $q(\lambda) \neq 0$.

The geometric multiplicity of $\lambda$ is

$$g(\lambda) = \dim E_\lambda.$$

**Proposition 1.1.** Let $T \in \mathrm{End}(V)$ be a linear map with an eigenvalue $\lambda$, then $g(\lambda) \leq a(\lambda)$.

*Proof.* Let $r = g(\lambda) = \dim E_\lambda$, then there exits linearly independent vectors $v_1, \cdots, v_r$ which forms a basis of $E_\lambda$. Suppose we extend this to a basis of $V$,

$$B = \{v_1, \cdots, v_r, w_1, \cdots, w_s\},$$

then by working out $T(b)$ for all $b \in B$, we find $T(v_i) = \lambda v_i$, and $T(w_i) = \sum \mu_i B$ so,

$$[T]_B = \left[\begin{array}{cccc|cccc} \lambda & 0 & \cdots & 0 & \mu_1(v_1) & \mu_2(v_1) & \cdots & \mu_s(v_3) \\ 0 & \lambda & \cdots & 0 & \mu_1(v_2) & \mu_2(v_2) & \cdots & \mu_s(v_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda & \mu_1(v_r) & \mu_2(v_r) & \cdots & \mu_s(v_r) \\ \hline & & & & \mu_1(w_1) & \mu_2(w_1) & \cdots & \mu_s(w_3) \\ & \mathbf{0} & & & \mu_1(w_2) & \mu_2(w_2) & \cdots & \mu_s(w_2) \\ & & & & \vdots & \vdots & \ddots & \vdots \\ & & & & \mu_1(w_s) & \mu_2(w_s) & \cdots & \mu_s(w_s) \end{array}\right].$$

We will refer to the four quadrants as $[\lambda], A, \mathbf{0}$ and $C$ respectively.

Thus, by considering the characteristic polynomial of this, we have

$$\chi_{[T]_B} = \det(xI - [T]_B) = (x - \lambda)^r \det(xI - C),$$

implying the algebraic multiplicity of $\lambda$ is at least $r$. □

**Theorem 2.** Let $\dim V = n$ and $T \in \mathrm{End}(V)$ be a linear map with distinct eigenvalues $\lambda_1, \cdots, \lambda_r$. Suppose that the characteristic polynomial of $T$ is

$$\chi_T = \prod_i (x - \lambda_i)^{a(\lambda_i)},$$

(so, $\sum_i a(\lambda_i) = n$). Then the following are equivalent,

1. $T$ is diagonalisable;
2. $\sum_i g(\lambda_i) = n$;
3. for all $i$, $g(\lambda_i) = a(\lambda_i)$.

*Proof.* $2 \iff 3$ is trivial so let us consider the other cases.

$1 \implies 2$. Suppose $T$ is diagonalisable, then there exists some $B$, a basis of $V$ consisting of eigenvectors of $T$. Then, we can partition $B$ into $F_{\lambda_i} := \{v \in B \mid T(v) = \lambda_i v\}$ for all eigenvalues of $T$. By noting that the subspace induced by $F_{\lambda_i}$ is a subspace of the $\lambda_i$ eigenspace $E_{\lambda_i}$, we have,

$$\sum_i g(\lambda_i) = \sum_i \dim E_{\lambda_i} \geq \sum_i \dim F_{\lambda_i} = n.$$

Now, as $\sum_i g(\lambda_i) \leq \sum_i a(\lambda_i) = n$ by the previous proposition, it follows $\sum_i g(\lambda_i) = n$.

$2 \implies 1$. Suppose $\sum_i g(\lambda_i) = n$. Let $B_i$ be the basis of $E_{\lambda_i}$ for all $\lambda_i$ an eigenvalue and let $B = \bigcup B_i$. We can straight away see that $|B| = n$ so it suffices to show that $B$ is linearly independent. Suppose otherwise, then there exists an index set $I \subseteq \{1, \cdots, r\}$,

$$\sum_{i \in I} \sum \mu_i B_i = 0$$

where $\sum \mu_i B_i \neq 0$ for all $i$. Now as $\sum \mu_i B_i \in E_{\lambda_i}$, this is a sum of eigenvectors with distinct eigenvalues. However, by theorem 1, these eigenvectors are there fore linearly independent, so they must be zero. # □

## 1.2 Direct Sums

Recall that we can add subspaces of a vector space together forming another subspace, that is, given $U_1, U_2 \leq V$, $U_1 + U_2 = \{u_1 + u_2 \mid u_1 \in U_1, u_2 \in U_2\} \leq V$. Direct sums is a particular case of this and is closely linked to *block-diagonal* matrices.

**Definition 1.3** (Direct Sum of Vector Space)**.** Let $V$ be a vector space with subspaces $U_1, \cdots, U_k$. We write

$$V = \bigoplus_{i=1}^{k} U_i$$

for the direct sum of subspaces if every $v \in V$, there exists unique $u_i \in U_i$ for all $i$ such that $v = \sum u_i$.

**Proposition 1.2.** Let $V$ be a vector space with subspaces $V_1, V_2$, then $V = V_1 \oplus V_2$ if and only if $V_1 \cap V_2 = \{0_v\}$ and $\dim V_1 + \dim V_2 = \dim V$.

*Proof.* Follow your nose. □

**Proposition 1.3.** Let $V$ be a vector space with subspaces $V_1, \cdots, V_k$, then $V = \bigoplus_{i=1}^{k} V_i$ if and only if $\sum_{i=1}^{k} \dim V_i = \dim V$ and if $B_i$ is a basis of $V_i$ then $\bigcup_{i=1}^{k} B_i$ is a basis of $V$.

*Proof.* ( $\implies$ ). Suppose $V = \bigoplus_{i=1}^{k} V_i$, then for all $i, j$, $V_i \cap V_j = \{0_V\}$, thus $B_i \cap B_j = \varnothing$ for all $i \neq j$, implying $\sum_{i=1}^{k} \dim V_i = \dim V$ and $|\bigcup B_i| = \dim V$ so it suffices to show that $\bigcup B_i$ is linearly independent. However, this is trivial as if $\bigcup B_i$ is linearly dependent, then there are two distinct ways of writing 0 as a sum of vectors in $V_i$. #

( $\impliedby$ ). Suppose $\sum_{i=1}^{k} \dim V_i = \dim V$ and $\bigcup_{i=1}^{k} B_i$ is a basis of $V$, then it follows,

$$V = V_1 + V_2 + \cdots V_k.$$

Suppose for contradiction there is two representations of $v \in V$ where $v = \sum v_i = \sum v_i'$. Then, $v = \sum_i \sum \mu_i B_i = \sum_i \sum \mu_i' B_i$, and thus, $0 = \sum_i \sum \mu_i B_i - \sum_i \sum \mu_i' B_i = \sum_i \sum (\mu_i - \mu_i') B_i$. By rewriting, $0 = \sum (\mu_i - \mu_i') \bigcup B_i$, implying $\bigcup B_i$ is linearly dependent. # □

**Definition 1.4** (Invariant Subspace)**.** Let $V$ be a vector space with subspace $W$ and let $T \in \mathrm{End}(V)$ be a linear map. We say $W$ is $T$-invariant if and only if

$$T(W) \subseteq W.$$

We write $T_W : W \to W$ as the restriction of $T$ to $W$.

A example of an invariant subspace is the eigenspace of a linear map since $T(E_\lambda) = \{T(v) \mid v \in E_\lambda\} = \{\lambda v\} \subseteq E_\lambda$.

**Theorem 3.** Let $T \in \mathrm{End}(V)$ be a linear map and suppose $V = \bigoplus V_i$ where for all $i$, $V_i$ is $T$-invariant. Let $B_i$ be a basis of $V_i$, and $A_i = [T_{V_i}]_{B_i}$, then

$$[T]_{\bigcup B_i} = \mathrm{diag}(A_1, A_2, \cdots, A_k).$$

*Proof.* Follows directly from the $T$-invariant property of $V_i$. □

From the proposition above, we see the close link between direct sums and block diagonal matrices. To further highlight the fact, from this point forward, we write $\bigoplus_{i=1}^k A_i = \mathrm{diag}(A_!, A_2, \cdots, A_k)$ where $A_i$ are block matrices.

**Corollary 3.1.** Let $A = \bigoplus_{i=1}^r A_i$ and let $\pi \in S_r$. Then $A \sim A' := \bigoplus_{i=1}^r A_{\pi(i)}$.

*Proof.* Let the vector space $V$ in the above theorem be the span of the columns of $A$ and $V_i$ the span of columns of $A_i$ with the missing entries filled with zero. Then it is not hard to see $V = \bigoplus V_i$. Now, by letting $T \in \mathrm{End}(V) : v \mapsto Av$, we see that for all $i$, $V_i$ is $T$-invariant and $T_{V_i} = v \mapsto A_i v$, so, by taking the basis $B$ to be the standard basis, we have $A = [T]_B$. Now, by permuting the standard basis by $\pi$, resulting in the basis $B'$, we have $A' = [T]_{B'}$, and so, by letting $P$ be the change of basis matrix from $B \to B'$, we have shown $A \sim A'$. □

## 1.3 Quotient Spaces

Just like other algebraic graphs we can construct a quotient structure on vector spaces.

Let $V$ be a vector space and $W \leq V$, then let $\sim_W \colon V \to V \to \mathrm{Prop}$ be the binary relation such that

$$v_1 \sim_W v_2 \iff v_1 + W = v_2 + W,$$

where $v + W = \{v + w \mid w \in W\}$ for all $v \in V$.

By manually checking, we find this is an equivalence relation and the set $V/\sim_W$ equipped with the natural addition and scalar multiplication form a vector space. We will write $V/W$ for this quotient space.

**Definition 1.5.** Given a quotient space $V/W$, there exists a linear map

$$q_W : V \to V/W : v \mapsto v + W.$$

**Proposition 1.4.** Let $V$ be a finite dimensional vector space with the subspace $W$, then $\dim V/W = \dim V - \dim W$.

*Proof.* Let $B_W$ be a basis of $W$ and $B_V$ the extension basis of $V$ from $B_W$. Then we easily see that $V/W \subseteq \mathrm{sp}(q_W(B_V \setminus B_W))$ as for all $v \in V$, $v = \sum \mu B_V$, so $v + W = q_W(v) = q_W(\sum \mu B_V) = \sum \mu q_W(B_W) + \sum \mu q_W(B_V \setminus B_W) = \sum 0_{V/W} + \sum \mu q_W(B_V \setminus B_W) \in \mathrm{sp}(q_W(B_V \setminus B_W))$.

Now suppose $q_W(B_V \setminus B_W)$ is not linearly independent in $V/W$, then, there exists $\mu$, $0 = \sum \mu q_W(B_V \setminus B_W) = q_W(\sum \mu(B_V \setminus B_W))$, so $\sum \mu(B_V \setminus B_W) \in \ker q_W = W$. If $\sum \mu(B_V \setminus B_W) = 0_V$, then $B_V \setminus B_W$ is not linearly dependent, a contradiction so, $\sum \mu(B_V \setminus B_W) \neq 0_V$. Now, as $\sum \mu(B_V \setminus B_W) \in W$, there is some $\lambda$, $\sum \mu(B_V \setminus B_W) = \sum \lambda B_W$, so $\sum \mu(B_V \setminus B_W) - \sum \lambda B_W = 0$ implying $B_V$ is not linearly independent. # □

With the above proposition, we have found a method to find a basis of a quotient space $V/W$ by extending the basis of $W$.

Let us now consider quotient spaces' relation with linear maps.

**Definition 1.6** (Quotient Map)**.** Let $V$ be a vector space and $W$ a subspace of $V$. Suppose $T \in \mathrm{End}(V)$ is a linear map and $W$ is $T$-invariant. Then there is an induced quotient map

$$\bar{T} : V/W \to V/W : q_W(v) \mapsto q_W(T(v)).$$

To see that this is well defined, let $u, v \in V$, $q_W(u) = q_W(v)$, then $u - v \in W$ implying $T(u - v) \in W$ as $W$ is $T$-invariant. Thus, $0_{V/W} = q_W(T(u - v)) = q_W(T(u) - T(v)) = q_W(T(u)) - q_W(T(v))$ implying $\bar{T}(u) = \bar{T}(v)$.

**Theorem 4.** Let $V$ be a vector space $W$ a subspace that is $T$-invariant for some $T \in \mathrm{End}(V)$ a linear map. Let $B_W$ be a basis of $W$, $B$ the extended basis of $V$ from $B_W$, and $\bar{B}$ the basis of $V/W$ as constructed by proposition 1.4. Then

$$[T]_B = \left[\begin{array}{c|c} [T_W]_{B_W} & A \\ \hline \mathbf{0} & [\bar{T}]_{\bar{B}} \end{array}\right],$$

where $A$ is some matrix.

*Proof.* Consider where $T(v)$ lands whenever $u \in B_w \subseteq W$, and where $\bar{T}(v)$ lands for the rest of the basis vectors. □

**Corollary 4.1.** Let $T \in \mathrm{End}(V)$ be a linear map and $W \leq V$ is $T$-invariant, then $\chi_T = \chi_{T_W} \chi_{\bar{T}}$ where $\chi_f$ denotes the characteristic polynomial of the linear map $f$.

## 1.4 Triangularisation Theorem

We have now arrived at the first major theorem of this course, that under certain conditions we can always triangularise matrices. We will in general work with upper triangular matrices when referring to triangular matrices.

**Proposition 1.5.** Let $A = [a_{i,j}], B = [b_{i,j}] \in M_n(\mathbb{F})$ be triangular, then

- $\chi_A(x) = \prod_{i=1}^n (x - a_{i,i})$;
- $\det A = \prod_{i=1}^n a_{i,i}$;
- $AB$ is also triangular with diagonal $a_{i,i} b_{i,i}$.

The Triangularisation theorem states:

**Theorem 5.** Let $V$ be a finite dimensional vector space over some field $\mathbb{F}$, and let $T \in \mathrm{End}(V)$ be a linear map. Suppose the the characteristic polynomial of $T$, $\chi_T$ factorises into a product of linear factors, i.e. there exists $\lambda_i \in \mathbb{F}$,

$$\chi_T(x) = \prod (x - \lambda_i),$$

then, there exists a basis $B$ of $V$ such that $[T]_B$ is upper triangular.

Straight away, we see a version of this in terms of matrices instead of linear maps in which the matrix is *similar* to a triangular matrix. We also note that, for some fields, such as the complex numbers $\mathbb{C}$, we can always triangularise any matrix (by *FTA*). This might not be the case for other fields such as the real numbers.

*Proof.* We induct on the dimension of $V$. The theorem is trivial when $\dim V = 1$, so let us consider the case when $\dim V = k + 1$ under the inductive hypothesis.

As $\chi_T$ factorises, $T$ has an eigenvalue $\lambda$ and some eigenvector $v \in V$. Let $W = \mathrm{sp}(v)$ be a $T$-invariant subspace of $V$. Then, by proposition 1.4, $V/W$ has dimension $k$ and we have the induced quotient map $\bar{T} : V/W \to V/W$. Now, by corollary 4.1, $\prod(x - \lambda_i) = \chi_T(x) = \chi_{\bar{T}}(x)\chi_{T_W}(x) = \chi_{T_W}(x)(x - \lambda)$. So, $\chi_{T_W}(x)$ is a polynomial of degree $k$ which factorises. Then by our inductive hypothesis, there exists a basis $\bar{B}$ such that $[\bar{T}]_{\bar{B}}$ is triangular. Then by theorem 4, we have found a basis $B$, $[T]_B$ is triangular. □

**Corollary 5.1.** Let $A \in M_n(\mathbb{C})$ with eigenvalues $\lambda_i$. Then $\sum g(\lambda_i)\lambda_i = \mathrm{tr}(A)$.

*Proof.* By the triangularisation theorem, $A = PQP^{-1}$ where $Q$ is triangular. As $A$ and $Q$ have the same eigenvalues, it suffices to show that $\mathrm{tr}(A) = \mathrm{tr}(Q)$. But this follows as $\mathrm{tr}(A) = \mathrm{tr}(PQP^{-1}) = \mathrm{tr}(P^{-1}PQ) = \mathrm{tr}(Q)$. □

# 2 Polynomials

## 2.1 Cayley-Hamilton Theorem

Recall that given a polynomial $p(x) = \sum_{i=0}^{n} a_i x^i$, we write $p(T)$ as the linear map $\sum_{i=0}^{n} a_i T^i$ for the linear map $T \in \mathrm{End}(V)$ and similarly for matrices. Then, the Cayley-Hamilton theorem states the famous result that, given a linear map $T \in \mathrm{End}(V)$, if $\chi_T$ is the characteristic polynomial of $T$, then $\chi_T(T) = 0$. We will prove this theorem within this chapter.

Straight away, we see that the result is trivial if the matrix in question is diagonal (or thus, similar to a diagonal matrix) since if $A = \mathrm{diag}(\lambda_i)$, $p(A) = \mathrm{diag}(p(\lambda_i))$ for any polynomial $p$. In fact, by similar argument, we find the theorem is also true for triangular matrices, and thus, by the triangularisation theorem, the Cayley-Hamilton theorem is true for vector spaces over the complex numbers (see problem sheet 3). However, this is less trivial for general matrices over arbitrary fields which we shall provide a proof here.

**Lemma 2.1.** Let $T \in \mathrm{End}(V)$ be a linear map such that there does not exists a proper non-trivial $T$-invariant subspace of $V$. Suppose $\dim V = n$, then the set $B := \{v, T(v), T^2(v), \cdots, T^{n-1}(v)\}$ forms a basis of $V$ for any non-zero $v \in V$.

*Proof.* As $|B| = n$, it suffices to show that it is linearly independent. Suppose otherwise, then there exists some $\mu$, such that $\sum \mu B = 0$. Then, we can choose $i$ such that $i$ is the largest number in which $T^i(v) = \sum_{i \neq j} \mu_j T^j(v)$. But then, for all $u \in \mathrm{sp}(B), u = \sum \lambda T^i(v)$, so $T(u) = T(\sum \lambda T^i(v)) = \sum \lambda T^{i+1}(v)$. Now, as $T^{n+1}(v) = T^{n+1-k}(T^k(v)) = T^{n+1-k}(\sum_{i \neq j} \mu_j T^j(v)) = \sum_{i \neq j} \mu_j T^{n+1-k+j}(v) \in \mathrm{sp}(B)$ as $n + 1 - k + j \leq n$ for all $j$ as $k$ is the largest. Thus, $\mathrm{sp}(B)$ is a proper and non-trivial $T$-invariant subspace. # □

*Proof.* (Cayley-Hamilton Theorem). Let $T \in \mathrm{End}(V)$ and $\chi_T$ be the characteristic polynomial of $T$. We will induct on the dimension of $V$, $n$.

The $n = 1$ case is trivial, so let us suppose the inductive hypothesis for dimensions $\leq k$ and we will prove this theorem for $n = k$.

Suppose first that there exists a proper and non-trivial $T$-invariant subspace $W$ of $V$ and suppose it has basis $B_W$. We can then extend this basis to a basis $B$ of $V$ such that

$$[T]_B = \left[\begin{array}{c|c} [T_W]_{B_W} & A \\ \hline \mathbf{0} & [\bar{T}]_{\bar{B}} \end{array}\right].$$

Now, we recall that $\chi_T = \chi_{T_W}\chi_{\bar{T}}$, so,

$$\begin{aligned}
\chi_T([T]_B) &= \chi_{T_W}([T]_B)\chi_{\bar{T}}([T]_B) \\
&= \left[\begin{array}{c|c} \chi_{T_W}([T_W]_{B_W}) & A \\ \hline \mathbf{0} & \chi_{T_W}([\bar{T}]_{\bar{B}}) \end{array}\right] \left[\begin{array}{c|c} \chi_{\bar{T}}([T_W]_{B_W}) & A \\ \hline \mathbf{0} & \chi_{\bar{T}}([\bar{T}]_{\bar{B}}) \end{array}\right] \\
&= \left[\begin{array}{c|c} \mathbf{0} & A \\ \hline \mathbf{0} & \chi_{T_W}([\bar{T}]_{\bar{B}}) \end{array}\right] \left[\begin{array}{c|c} \chi_{\bar{T}}([T_W]_{B_W}) & A \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \\
&= \mathbf{0},
\end{aligned}$$

where we write $A$ for arbitrary block matrix and the second to last equality is due to the inductive hypothesis.

Suppose now that there does not exists a non-trivial proper $T$-invariant subspace of $V$. Then by lemma 2.1, the set $B := \{v, T(v), \cdots, T^n(v)\}$ forms a basis of $V$. Now, we see that $[T]_B$ is a companion matrix resulting in $\chi_{[T]_B}(x) = \sum a_i x^i$, where $a_i$ are chosen such that $T^{n+1} = \sum -a_i T^i(v)$. But $T^{n+1} = \sum -a_i T^i(v) \iff \sum_{i=0}^{n+1} -a_i T^i(v) = 0$ where we let $a_{n+1} = 1$, so we have $\chi_{[T]_B}([T]_B(v)) = \sum_{i=0}^{n+1} -a_i T^i(v) = 0$.

Thus, by the law of excluded middle, we have Cayley-Hamilton. □

## 2.2 Some Theories on Polynomials

Let $\mathbb{F}$ be a field, then we denote the ring formed by the polynomials over $\mathbb{F}$ as $\mathbb{F}[X]$. And we will develop the theories of greatest common divisor, least common multiple, and polynomial prime factorisation for this ring.

**Theorem 6** (Euclidean Algorithm)**.** Let $f, g \in \mathbb{F}[X]$ such that $\deg g \geq 1$. Then there exists $q, r \in \mathbb{F}[X]$, $f = qg + r$ where $r = 0$ or $\deg r < \deg g$.

*Proof.* We induct on the degree of $f$, $n$. For $n = 0$, we can choose $q = 0$ and $r = f$ and we are done. Let's now assume $n = k + 1$ alongside the inductive hypothesis.

Let's write $f(x) = a_{k+1}x^{k+1} + \cdots$ and $g(x) = a_m x^m + \cdots$. Then, we can write $f_1 = f - a_{k+1}b_m^{-1}x^{n-m}g$, where $\deg f_1 \leq \deg f$. So by the inductive hypothesis, there exists some $q, r \in \mathbb{F}[X]$ such that $f_1 = qg + r$ and $\deg r \leq \deg g$. Then, $f = f_1 + a_{k+1}b_m^{-1}x^{n-m}g = (q + a_{k+1}b_m^{-1}x^{n-m})g + r$. □

**Definition 2.1** (Greatest Common Divisor)**.** Let $f, g \in \mathbb{F}[X] \setminus \{0\}$. Then we say $d \in \mathbb{F}[X]$ is the greatest common divisor of $f$ and $g$, $\gcd(f, g)$ if and only if $d \mid f, d \mid g$ and for all $e \mid f$ and $e \mid g$, $e \mid d$.

Straight away, we see that, unlike the integers, the greatest common divisor is of two polynomials is not unique as we can simply multiply the gcd by any scalar and receive another gcd. However, if we quotient out by this relation ($\sim: \mathbb{F}[X] \to \mathbb{F}[X] \to \text{Prop} : f, g \mapsto \exists \lambda \in \mathbb{F} \setminus \{0\}, f = \lambda g$), the gcd turns out to be unique (see problem sheet 3) and exists.

**Theorem 7.** If $f, g \in \mathbb{F}[X] \setminus \{0\}$, then the $\gcd(f, g)$ exists.

*Proof.* Same argument as the integers by repeatedly applying the Euclidean algorithm. □

**Definition 2.2** (Coprime)**.** We call two polynomials $f, g \in \mathbb{F}[X]$ to be coprime if and only if $\gcd(f, g) = 1$.

**Theorem 8** (Bezout's)**.** If $f, g, d \in \mathbb{F}[X]$ such that $d = \gcd(f, g)$, then there exists $r, s \in \mathbb{F}[X]$ such that $d = rf + sg$.

Now that we have established some basic properties about polynomials, we would like to consider what it might mean to be a prime polynomial.

**Definition 2.3** (Irreducible)**.** A polynomial $f \in \mathbb{F}[X]$ is irreducible over $\mathbb{F}$ if and only if $\deg f \geq 1$ and there does not exists $g, h \in \mathbb{F}[X]$, $\deg g, \deg h < \deg f$ such that $f = gh$.

**Remark.** We see that this definition of irreducibility is consistent with the one we have defined in ring theory since, $\langle f \rangle$ is not a maximal ideal if and only if there exists $g \in \mathbb{F}[X]$, $\langle f \rangle \subset \langle g \rangle \subset \mathbb{F}[X]$ and hence, $f \in \langle g \rangle$ implying there exists $h \in \mathbb{F}[X], f = gh$.

Given $p \in \mathbb{Q}[X]$, it is usually difficult to decide whether or not it is irreducible. However, there is some tools that can help us determine the irreducibility of some rational polynomials.

**Theorem 9.** Let $p \in \mathbb{Q}[X]$ be a monic polynomial with integer coefficients. Then,

- if $\alpha \in \mathbb{Q}$ is a root of $p$, then $\alpha \in \mathbb{Z}$;
- if $p$ is irreducible over $\mathbb{Q}$, then it has a monic factorisation $q, r$, where $q, r$ also have integer coefficients.

*Proof.* The first part follows easily while the other is Gauss' lemma. □

**Theorem 10.** Let $p \in \mathbb{F}[X]$ be irreducible, and $a, b \in \mathbb{F}[X]$, if $p \mid ab$, then $p \mid a$ or $p \mid b$.

*Proof.* Suppose $p \nmid a$, then $\gcd(p, q) = 1$ and by Bezout's, there exists $r, s \in \mathbb{F}[X]$ such that $rp + sa = 1$ so $b = rpb + sab$. Now, as $p \mid ab$, there exists $q \in \mathbb{F}[X]$ such that $ab = pq$ and so $b = (rb + sq)p$ and thus $p \mid b$. □

**Theorem 11** (Unique factorisation Theorem for Polynomials)**.** Let $f \in \mathbb{F}[X]$ with $\deg f \geq 1$, then there exists a unique sequence of polynomials $(p_i)_{i=1}^r \subset \mathbb{F}[X]$, such that $f = \prod p_i$.

*Proof.* Let us first prove existence. We induct on the degree of $f$. For $\deg f = 1$, the result is trivial so let use consider the theorem with $\deg f = k + 1$ with the inductive hypothesis. Now, if $f$ is irreducible, the result follows so suppose otherwise. Then $f$ can be factorised into two polynomials with degree less than that of $f$. But, by the inductive hypothesis, these two polynomial can be factorised, so, by multiplying their factors, can $f$ be factorised.

Let us now prove uniqueness. Suppose now $f = \prod_{i=1}^r p_i = \prod_{j=1}^s q_j$. Then, by considering that for all $i$ $p_i \mid q_j$ for some $j$, the result follows. □

Lastly, we conclude on defining the least common multiple for polynomials.

**Definition 2.4** (Least Common Mulitple)**.** Let $f, g \in \mathbb{F}[X]$, then the least common multiple of $f$ and $g$, $\operatorname{lcm}(f, g)$ is the polynomial $h$ such that $f \mid h, g \mid h$ and for all $k \in \mathbb{F}[X]$, if $f \mid k$ and $g \mid k$, then $h \mid k$.

Similarly to the greatest common multiple, we find the least common multiple exists and is unique (up to scalar multiplication).

## 2.3 Minimal Polynomial

We will in this section take a look at the minimal polynomial and some of its applications.

**Definition 2.5** (Minimal Polynomial)**.** The minimal polynomial of $T \in \mathrm{End}(V)$ is the non-zero monic polynomial $m_T(x)$ of least degree such that $m_T(T) = x \mapsto 0$.

Straight away, we see $m_T \mid \chi_T$ as, by the Euclidean algorithm, there exists $q, r \in \mathbb{F}[X]$ such that $\chi_T = qm_T + r$ where $\deg r < \deg m_T$ if $r \neq 0$. By evaluating with $T$ on both sides, we have $0 = r(T)$, so $r$ annihilates $T$. But this contradicts the minimality of $m_T$, so $r = 0$. This can be generalised to any annihilators of $T$.

**Lemma 2.2.** Let $T \in \mathrm{End}(V)$ and let $p \in \mathbb{F}[X]$ such that $p(T) = 0$. Then, a minimal polynomial of $T$ divides $p$.

*Proof.* By the Euclidean algorithm, there exists, $q, r \in \mathbb{F}[X]$, where $r = 0$ or $\deg r < \deg m_T$, such that $p = qm_T + r$. Thus, $0 = p(T) = q(T)m_T(T) + r(T) = r(T)$. # (Minimality of $m_T$.) □

Furthermore, we can deduce that the minimal polynomial is unique.

**Theorem 12.** Let $T \in \mathrm{End}(V)$ and let $m_T$ and $m'_T$ be minimal polynomials of $T$, then $m_T = m'_T$.

*Proof.* By the previous lemma, we have $m_T \mid m'_T$ so there exists $p \in \mathbb{F}[X]$, such that $m_T = pm'_T$. Now, as both $m_T$ and $m'_T$ are the minimal polynomials, they must have the same degree, thus $\deg p = 0$ and hence, is a constant. But, since both $m_T$ and $m'_T$ are monic, $p = 1$, so $m_T = m'_T$. □

As one can imagine, the minimal polynomial for matrices is defined similarly and alike many other properties, is shared for similar matrices.

**Theorem 13.** Let $A, B \in M_n\mathbb{F}$ and $A \sim B$. Then the minimal polynomial of $A$, $m_A$ is the same as the minimal polynomial of $B$, $m_B$.

*Proof.* Suppose $A = P^{-1}BP$. Then, $0 = m_A(A) = m_A(P^{-1}BP) = P^{-1}m_A(B)P$, so $m_A(B) = 0$ as $P$ is invertible. By symmetry, $m_B(A) = 0$ and thus, the result follows by divisibility. □

As the minimal polynomial is pretty powerful, it will be helpful to be able to compute the minimal polynomial. We will now develop some tools to help us find this minimal polynomial.

**Theorem 14.** Let $T \in \mathrm{End}(V)$, then $\lambda$ is an eigenvalue of $T$ if and only if $m_T(\lambda) = 0$.

*Proof.* ( $\implies$ ) If $\lambda$ is an eigenvalue of $T$ then there exists non-zero $v \in V$, $T(v) = \lambda v$, then $0 = m_T(T)(v) = \sum a_i T^i(v) = \sum a_i \lambda^i v = m_T(\lambda)v$. Now, as $v$ is non-zero, $m_T(\lambda) = 0$.

( $\impliedby$ ) Backwards direction follows straight away as if $m_T(\lambda) = 0$, as $m_T \mid \chi_T$ there exists some polynomial $p$ such that $\chi_T = m_T p$ and thus, $\chi_T(\lambda) = m_T(\lambda)p(\lambda) = 0$, and hence, is an eigenvalue. □

With that, in order to find the minimal polynomial, it will be insightful to find the characteristic polynomial as the minimal polynomial shares roots and divides it (and then you can check all the cases).

**Proposition 2.1.** The minimal polynomial of the companion matrix of some polynomial $p$ is $p$.

Before we can prove proposition 2.1, let us prove another useful lemma.

**Lemma 2.3.** Let $A \in M_n(\mathbb{F})$ and suppose there exists some $v \in \mathbb{F}^n$, such that the set $S := \{A^i v \mid i \leq k\}$ is linearly independent for some $k < n$. Then all polynomials that annihilates $A$ has degree at least $k+1$.

*Proof.* Suppose there exists $p \in \mathbb{F}[X]$ with degree $l \leq k$ such that $p(A) = 0$. Then $0 = p(A)v = \sum A^i v$. # □

*Proof.* (Proposition 2.1). By the previous lemma, it suffices to find some $v$ such that $\{C(p)^i v \mid i \leq n-1\}$ is linearly independent. Straight away, we see $v = e_1$ suffices so we are done. □

While, we proved that the minimal polynomial share linear factors with the characteristic polynomial, we hope the same is true for all irreducible factors. This turns out to be true.

**Theorem 15.** Let $T \in \mathrm{End}(V)$, then, for all $p \in \mathbb{F}[X]$ such that $p$ is a irreducible factor of $\chi_T$, $p \mid m_T$.

*Proof.* We will prove this theorem by cases on whether there exists a proper non-trivial $T$-invariant subspace of $V$, so first, let us suppose such a subspace $W$ exists.

Let us induct on the dimension of $V$, n. The theorem is trivial for $n = 1$, so let us consider the case for $n = k+1$ with the inductive hypothesis holding for all $n = m \leq k$. Again, as $W$ is $T$-invariant, we have

$$[T]_B = \left[\begin{array}{c|c} [T_W]_{B_W} & A \\ \hline \mathbf{0} & [\bar{T}]_{\bar{B}} \end{array}\right].$$

Now, suppose we write $m_T(x) = \sum a_i x^i$, then,

$$\begin{aligned} 0 = m_T([T]_B) &= \sum a_i [T]_B^i \\ &= \left[\begin{array}{c|c} \sum a_i [T_W]_{B_W}^i & A' \\ \hline \mathbf{0} & \sum a_i [\bar{T}]_{\bar{B}}^i \end{array}\right] \\ &= \left[\begin{array}{c|c} m_T([T_W]_{B_W}) & A' \\ \hline \mathbf{0} & m_T([\bar{T}]_{\bar{B}}) \end{array}\right], \end{aligned}$$

so $m_T$ annihilates both $[T_W]_{B_W}$ and $[\bar{T}]_{\bar{B}}$. Now, by considering $p \mid \chi_T = \chi_{[T_W]_{B_W}} \chi_{[\bar{T}]_{\bar{B}}}$, where $p$ is irreducible, $p \mid \chi_{[T_W]_{B_W}}$ or $p \mid \chi_{[\bar{T}]_{\bar{B}}}$. Either way, by the inductive hypothesis, $p \mid m_{[T_W]_{B_W}}$ or $p \mid m_{[\bar{T}]_{\bar{B}}}$, both of which divides $m_T$ as $m_T$ annihilates their respecctive matrices.

Let us now consider the case in which there does not exists a proper non-trivial $T$-invariant subspace of $V$. But, then, similar to the proof of the Cayley-Hamilton theorem, we can construct some basis $B$ such that $[T]_B$ is the companion matrix of $\chi_{[T]_B}$. Now as the companion matrix has the same minimal polynomial as the characteristic polynomial, we are done! □

Before we end this section, I would like to prove a powerful result regarding diagonalisability.

**Theorem 16.** Let $T \in \mathrm{End}(V)$, and suppose there exists a non-zero polynomial $p$ of degree greater than 0 such that $p$ annihilates $T$ and $p$ can be factored into distinct roots. Then $T$ is diagonalisable.

*Proof.* Suppose we write $p(x) = \prod(x-\lambda_i)$, then, it suffices to show that $\dim\ker(\prod(T-\lambda_i)) \leq \sum \dim\ker(T-\lambda_i)$ as the kernel of $\prod(T-\lambda_i)$ is $V$ while the latter is the sum of the dimensions of the eigenspaces. This can be proved by showing $\dim\ker T_1T_2 \leq \dim\ker T_1 + \dim\ker T_2$ and applying induction. □

**Proposition 2.2.** Let $T_1, T_2 \in \mathrm{End}(V)$, then $\dim\ker T_1T_2 \leq \dim\ker T_1 + \dim\ker T_2$.

*Proof.* It is easy to see that $\ker T_2 \leq \ker T_1T_2$ so it suffices to show $\dim\ker T_1T_2/\ker T_2 \leq \dim\ker T_1$ since for $W \leq V$, $\dim V/W = \dim V - \dim W$.

Consider the map $\phi : \ker T_1T_2/\ker T_2 \to \dim\ker T_1 : v + \ker T_2 \mapsto T_2(v)$. $\phi$ has well-defined range as for all $v \in \ker T_1T_2$, $T_2(v) \in \ker T_1$. We will now show $\phi$ is well-defined overall and injective all at once. Let $v_1, v_2 \in \ker T_1T_2$, then,

$$v_1 + \ker T_2 = v_2 + \ker T_2 \iff v_1 - v_2 \in \ker T_2 \iff T_2(v_1 - v_2) = 0 \iff T_2(v_1) = T_2(v_2).$$

Furthermore, it is easy to see that $\phi$ is a linear map since

$$\begin{aligned}\phi((v_1 + \ker T_2) + (v_2 + \ker T_2)) &= \phi((v_1 + v_2) + \ker T_2)\\ &= T_2(v_1 + v_2) = T_2(v_1) + T_2(v_2)\\ &= \phi(v_1 + \ker T_2) + \phi(v_2 + \ker T_2).\end{aligned}$$

so $\phi$ is a injective linear map. Now, as the image of an linearly independent set under a injective linear map is also linearly independent, we can construct an linearly independent set in $\ker T_1$ with cardinality $\dim\ker T_1T_2/\ker T_2$ by taking the image of its basis over $\phi$. Thus, $\dim\ker T_1T_2/\ker T_2 \leq \dim\ker T_2$ and the result follows from proposition 1.4. □

# 3 Canonical Forms of Vector Spaces

As mentioned in the introduction, the highlight of this course are the canonical form theorems on general vector spaces (that is we would like to show any matrix over some field is similar to a particularly "nice" block diagonal matrix). This will allow us to deduce properties about general matrices and correspondingly, linear maps.

In order to achieve this, we will first develop some theories on how to decompose a general vector space, that is, given vector space $V$, we would like to find $V_i$ such that $V = \bigoplus V_i$.

## 3.1 Primary Decomposition

**Theorem 17** (Primary Decomposition Theorem)**.** Let $V$ be a finite dimensional vector space of the field $\mathbb{F}$, and let $T \in \mathrm{End}(V)$ with minimal polynomial $m_T$. Suppose $m_T$ has the irreducible factors $p_i$ such that,

$$m_T = \prod_{i=1}^{k} p_i^{n_i},$$

where $p_i \neq p_j$ for all $i \neq j$. Then,

- $\mathrm{V} = \bigoplus_{i=1}^{k} \ker(p_i(T)^{n_i})$;
- $\ker(p_i(T)^{n_i})$ is $T$-invariant;
- each restriction of $T$ on $\ker(p_i(T)^{n_i})$ has minimal polynomial $p_i^{n_i}$.

Straight away, we see that this decomposition is unique by the unique factorisation theorem, so, we can call it a canonical decomposition. Furthermore, if all factors of $m_T$ are linear, the factorisation becomes,

$$m_T = \prod_{i=1}^{k} (x - \lambda_i)^{n_i},$$

for distinct $\lambda_i$. In this cases, the individual decomposition becomes $\ker(T - \lambda I)^{n_i}$ which are the *generalised eigenspace* of $T$. Lastly, we notice a direct corollary of this theorem is theorem 16, so we have alternatively an (arguably) easier proof of this.

Before we can prove the primary decomposition theorem, let us prove the following lemma.

**Lemma 3.1.** Let $T \in \mathrm{End}(V)$ and suppose $p_1, p_2 \in \mathbb{F}[X]$ are coprime such that $p_1(T)p_2(T) = 0$, then $V = \ker p_1(T) \oplus \ker p_2(T)$. Furthermore, if $m_T = p_1p_2$, then the restriction of $T$ to $\ker p_i(T)$ has minimal polynomial $p_i$ for $i = 1, 2$.

*Proof.* By Bezout's, there exists $s, t \in \mathbb{F}[X]$ such that $1 = sp_1 + tp_2$. So, by evaluating at $T$, we have $I = s(T)p_1(T) + t(T)p_2(T)$. Then for all $v \in V$, $v = Iv = s(T)p_1(T)(v) + t(T)p_2(T)(v) = v_1 + v_2$. Now, as $p_1p_2 = 0$, we have $p_2(v_1) = p_2(s(T)p_1(T)(v)) = p_2s(T)p_2p_1(T)(v) = 0$, we have $v_1 \in \ker p_2(T)$ and similarly, $v_2 \in \ker p_1(T)$, and so, $V = \ker p_1(T) + \ker p_2(T)$. Also, suppose $v \in V_1 \cap V_2$, then $v = Iv = s(T)p_1(T)(v) + t(T)p_2(T)(v) = 0$, so $V = \ker p_1(T) \oplus \ker p_2(T)$.

Now, suppose $m_T = p_1p_2$ and let us denote the minimal polynomial of $T$ restricted on $\ker p_i(T)$ as $m_i$ for $i = 1, 2$. By definition $p_i(T_i) := p_i(T\,|_{\ker p_i(T)}) = 0$, so $m_i \mid p_i$. So, as $p_1$

and $p_2$ are coprime, so are $m_1$ and $m_2$, thus, $m_T = \text{lcm}(m_1, m_2) = m_1 m_2$ and the result follows. □

We can see straight away how this lemma can help us prove the theorem.

*Proof.* (Primary Decomposition Theorem). Let us write $m_T = \prod_{i=1}^{k} p_i^{n_i}$ and we will induct on $k$. The theorem is trivial for for the base case so we begin by letting $k = n + 1$. Then, by considering the previous lemma on $p_{n+1}$ and $\prod_{i=1}^{n} p_i$, the theorem follows. □

## 3.2 Jordan Canonical Form

Recall that the triangularisation theorem is limited in many ways. Not only does it not apply for matrices who's characteristic polynomial cannot be factored, for those matrices the theorem does apply, the triangularisation is not unique. We attempt to improve upon this with the Jordan canonical form theorem.

**Definition 3.1** (Jordan Blocks)**.** Let $\mathbb{F}$ be a field and let $\lambda \in \mathbb{F}$. Then $J_n(\lambda) \in M_n(\mathbb{F})$ is a Jordan block if and only if

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

Jordan blocks are a nice form of matrices and we can derive several properties straight away.

**Proposition 3.1.** Let $J = J_n(\lambda)$ be some Jordan block, then

- $\chi_J(x) = (x - \lambda)^n = m_J$;
- $\lambda$ is the only eigenvalue and $a(\lambda) = n, g(\lambda) = 1$;
- $(J - \lambda I)e_{i+1} = e_i$ for $i = 1, \cdots, n-1$ and $(J - \lambda I)e_1 = 0$.

*Proof.* Obvious. □

Before we state the Jordan canonical form theorem, let us recall some properties of block diagonal matrices. Let $A_i$ have characteristic polynomial $\chi_i$ and let $A = \bigoplus A_i$, then

- $\chi_A = \prod \chi_i$;
- $m_A = \text{lcm}(m_i)$;
- for all $\lambda$ and eigenvalue of $A$, $\dim E_\lambda(A) = \sum \dim E_\lambda(A_i)$;
- for all $q \in \mathbb{F}[X]$, $q(A) = \bigoplus q(A_i)$.

With that, we can state the Jordan canonical form theorem.

**Theorem 18** (Jordan canonical form theorem)**.** Let $A \in M_n(\mathbb{F})$ and the characteristic polynomial is a product of linear factors over $\mathbb{F}$. Then, there exists some $J$, $A \sim J = \bigoplus J_{n_i}(\lambda_i)$ and this decomposition is unique up to the ordering of the Jordan blocks.

We call this unique decomposition the Jordan canonical form of $A$.

**Remark.** The condition on which the characteristic polynomial must be factorisable is necessary as otherwise it would not even be triangulisable. Also, we note that the $\lambda_i$ is not necessarily distinct.

Of course, there is an equivalent theorem for linear maps. In fact, we shall prove the linear map version now.

**Theorem 19.** Let $V$ be a $n$-dimensional vector space over $\mathbb{F}$ and $T \in \mathrm{End}(V)$ such that $\chi_T$ is a product of linear factors over $\mathbb{F}$. Then, there exists a basis $B$ of $V$ such that $[T]_B = J = \bigoplus J_{n_i}(\lambda_i)$, and furthermore, $J$ is uniquely determined by $T$ (up to ordering).

*Proof.* We shall provide a proof for uniqueness first.

Suppose that $T$ has only one eigenvalue $\lambda$ and

$$[T]_B \sim J = \bigoplus J_{n_i}(\lambda)$$

for some basis of $V$, $B$. Let us write $J = \bigoplus_{i=1}^{r} J_i(\lambda)^{a_i}$ where $a_i$ is the number of $\lambda$-blocks with size $i$. Then, let us define

$$m_i := \dim \mathrm{Im}(J - \lambda I)^i.$$

Now, by considering $J_r(0)^i$, for $i = 1, \cdots, r$ we have

$$\begin{aligned} m_{r-1} &= a_r, \\ m_{r-2} &= a_{r-1} + 2a_r, \\ m_{r-3} &= a_{r-2} + 2a_{r-1} + 3a_r, \end{aligned}$$

and so on. In general, we see

$$m_i = \sum_{j=1}^{r-i} j a_{i+j}$$

for $i = 0, \cdots, r-1$. Thus, $m_i$ uniquely determines $a_i$; and as $m_i$ is computable and is uniquely determined by $(J - \lambda I)^i$, we have $a_i$ is unique.

Now, let us relax the number of eigenvalues condition. Suppose $T$ has at least two eigenvalues with one of them being $\lambda$, then we can write

$$[T]_B \sim J = J_\lambda \oplus L,$$

where $L$ is the direct sum of all the other Jordan blocks. Now, as $L - \lambda I$ is invertible (easy to see as the determinant is non-zero), it has full rank $l$. Thus, by defining

$$r_i = \dim \mathrm{Im}(J - \lambda I)^i,$$

we have $r_i = \dim \mathrm{Im}(J_\lambda - \lambda I)^i + l$, allowing us to compute $m_i = \dim \mathrm{Im}(J_\lambda - \lambda I)^i = r_i - l$. $\square$

In most cases, it might not be necessary to go through the steps as described above to find the Jordan canonical form of a matrix.

**Proposition 3.2.** Let $A \in M_n(\mathbb{F}) \sim J$ where $J$ is in Jordan Canonical Form. Suppose we write

$$J = (J_{n_1}(\lambda) \oplus \cdots \oplus J_{n_\alpha}(\lambda)) \oplus (J_{m_1}(\mu) \oplus \cdots \oplus J_{m_\beta}(\mu)) \oplus \cdots,$$

where $\lambda, \mu, \cdots$ are eigenvalues of $A$, then,

- the sum of the size of the $\lambda$-blocks equals the algebraic multiplicity of $\lambda$, i.e. $n_1 + \cdots + n_\alpha = a(\lambda)$;
- the number of $\lambda$-blocks equals the geometric multiplicity of $\lambda$, i.e. $\alpha = g(\lambda)$;
- $\max\{n_1, \cdots, n_\alpha\} = r$ where $(x-\lambda)^r$ is the highest power of $x - \lambda$ dividing $m_A$.

*Proof.* The proofs are rather straight forward.

- Obvious by considering the characteristic polynomial of $J_k(\lambda)$.
- True by considering each $\lambda$-block has geometric multiplicity one, and that $J$ and $A$ share the same $\lambda$-eigenspace.
- Since for all $n$, $J_n(\lambda)$ has minimal polynomial $(x-\lambda)^n$, we have the power dividing $m_J$ is $\operatorname{lcm}((x-\lambda)^{n_i}, \cdots, (x-\lambda)^{n_\alpha})$ which equals $(x-\lambda)^{\max\{n_1, \cdots, n_\alpha\}}$.

□

Now, let us finish the proof of the JCF theorem by proving that a linear map is similar to its JCF (we can say this since we had proved uniqueness).

*Proof.* First we reduce the proof to the case where $T \in \operatorname{End}(V)$ has only one eigenvalue. Suppose

$$m_T(x) = \prod_{i=1}^{n} (x - \lambda_i)^{n_i},$$

where $\lambda_i$ are distinct. By the primary decomposition theorem

$$V = \bigoplus_{i=1}^{n} \ker(T - \lambda_i I).$$

Then, if $B_i$ is a basis of $\ker(T - \lambda_i I)$, then $\bigcup_i B_i$ is a basis of $V$. Furthermore, as $\ker(T - \lambda_i I)$ are $T$-invariant, we can write

$$[T]_B = \bigoplus_{i=1}^{n} [T_{V_i}]_{B_i},$$

and for all $i$, $[T_{V_i}]_{B_i}$ has minimal polynomial $(x-\lambda_i)^{n_i}$, that is $[T_{V_i}]_{B_i}$ has only one eigenvalue. Hence, it suffices to show the theorem whenever $T$ has only one eigenvalue $\lambda$.

If $T$ only has one eigenvalue $\lambda$, then $\chi_T(x) = (x-\lambda)^n$ where $n = \dim V$. Suppose we define $S := T - \lambda I$, then straight away, we see $S^n = 0$ by Cayley-Hamilton so $S$ is nilpotent, and hence, it suffices to show that $S$ is similar to a JCF with 0-Jordan blocks.

By considering the 0-Jordan blocks are cyclic matrices, that is, we see $[S]_B = J_1(0) \oplus \cdots$ where $B := \{v_1, \cdots, v_n\}$ if and only if $S(v_i) = v_{i+1}$ for all $i = 0, \cdots n-1$ and $S(v_{n_1}) = 0$,

we can write the basis $B$ as the union

$$\bigcup_{i=1}^{k}\{v_1, S(v_1), \cdots, S^{n_i-1}(v)\}.$$

So, it suffices to find such $\{v_1, \cdots, v_k\}$ (we call basis generated using $S$ by this set a Jordan basis of $V$).

To prove this, we shall induct on $n = \dim V$. The base case is trivial so assume the existence of $\{v_1, \cdots, v_k\}$ for all $\dim V < n$. Consider $\mathrm{Im}(S) = S(V)$ is a proper subspace of $V$ (as $S$ is nilpotent). So, by the inductive hypothesis, there exists a Jordan basis of $S(V)$ generated by $U = \{u_1, \cdots, u_r\}$. Now as, for all $u_i \in U$, $u_i \in S(V)$, there exists $v_i \in V$, such that $S(v_i) = u_i$. Furthermore, as $S^{m_i}(u_i) = 0$, we see that $S^{m_i-1}(u_i) \in \ker S$, and thus, we can extend this to a basis of $\ker S$ by adding $w_1, \cdots, w_s$.

Finally, by letting $B = \{v_1, \cdots, v_r, w_1 \cdots, w_s\}$, we see that $B$ generates a Jordan basis if and only if $S^i(B)$ linearly independent (since $\dim \ker S + \dim \mathrm{Im}S = n$). Suppose then there exists $\mu$ such that $\sum \mu S^i(B) = 0$. Then, by applying $S$ on both sides, we receive $\sum \mu' S^i(U) = 0$ implying the non-vanishing terms has zero coefficient. Furthermore, we see the vanishing vectors is a basis of $\ker S$, so they also have zero coefficient, and hence $\mu = 0$ and $S^i(B)$ is linearly independent. Thus, we have found a Jordan basis of $S$, so we are done! □

Now that we have proved that the JCF theorem, we would like to compute the Jordan basis of $V$ such that $[T]_B$ is in JCF for some appropriate $T \in \mathrm{End}(V)$.

Let $S \in \mathrm{End}(V)$ be nilpotent,

1. Compute the chain of subspaces

   $$V \geq S(V) \geq S^2(V) \geq \cdots \geq S^r(V) \geq 0$$

   where $S^{r+1}(V) = 0$.

2. Find a basis of $S^r(V)$ and for each $i$, add $v_i$ to $U$ where $S(v_i) = u_i$. Furthermore, if needed, extend $U$ such that it has span greater than $\ker S^{r-1}$. This results in $U$ to be a Jordan basis of $S^{r-1}(V)$.

3. Repeat the above step until we get a Jordan basis of $V$.

## 3.3 Cyclic Decomposition

So far we have seen a satisfactory canonical form for matrices who's characteristic polynomial factors into linear roots. However, this is often not the case. So, in the next few sections, we shall develop theories for the *rational canonical form* theorem.

**Definition 3.2.** Let $V$ be a finite dimensional vector space over some field $\mathbb{F}$, $T \in \mathrm{End}(V)$, $0 \neq v \in V$ and define the subspace

$$\begin{aligned} Z(v, T) :=& \{p(T)(v)) \mid p \in \mathbb{F}[X]\} \\ =& \mathrm{sp}(v, T(v), T^2(v), \cdots), \end{aligned}$$

and we call it the $T$-cyclic subspace of $V$ generated by $v$.

Straight away, we see that $Z(v,T)$ is $T$-invariant and thus, we can restrict $T$ by $Z(v,T)$ (denoted by $T_v$). Also, if $v$ is an eigenvector of $T$, then $Z(v,T) = \mathrm{sp}(v)$.

**Definition 3.3** ($T$-annihilator)**.** The $T$-annihilator of $v$ is the smallest degree, monic polynomial $m_v(x)$ such that $m_v(T)(v) = 0$.

Consider the sequence $v, T(v), T^2(v), \cdots$. As $V$ is finite dimensional there exists some $k$, where $k$ is the smallest natural number such that $T^k(v) \in \mathrm{sp}(v, \cdots, T^{k-1}(v))$. Then, there exists $\mu$, such that $T^k(v) = -\sum \mu T^i(v)$, and so, $(\sum^n \mu T^i)v = 0$. Hence, by the choice of $k$, $m_v(x) = x^k \sum \mu x^i$.

**Proposition 3.3.** Given $k$ as defined above, $B = \{v, \cdots, T^{k-1}v\}$ is a basis of $Z(v,T)$. Furthermore, $[T_v]_B$ is the companion matrix of $m_v$ and the minimal polynomial of $T_v$ is $m_v$.

*Proof.* Obvious except perhaps for the last part. However, we have proved that the minimal polynomial of the companion matrix is the polynomial it is associated with, that is $m_v$. □

**Proposition 3.4.** Let $A \in M_n(\mathbb{F})$. For all $v \in \mathbb{F}^n$, suppose we let $k$ be the largest number such that $\{v, Av, \cdots, A^k v\}$ is linearly independent, then, $k \leq \dim \mathrm{Im}\, A$.

*Proof.* Suppose $k > \dim \mathrm{Im}\, A$, for some $v$. Then we see that $A$ is similar to a matrix with rank greater than $A$ by considering $PAP^{-1}$ where $P$ is the change of basis matrix from the standard basis to the basis extended from $\{v, \cdots, A^k v\}$. But this is a contradiction as similar matrices to $A$ has the same rank as $A$. □

We can easily extend the above proposition to block diagonal matrices.

**Proposition 3.5.** Let $A \in M_n(\mathbb{F})$ and suppose $A = \bigoplus A_i$. Then, for all $v \in \mathbb{F}^n$, if $k$ is the largest number such that $\{v, Av, \cdots, A^k v\}$ is linearly independent, then, $k \leq \max\{\dim \mathrm{Im}\, A_i\}$.

*Proof.* Suppose we write $v = (v_i)$ such that $Av = (A_i v_i)$, then, since by assumption, there exists $\mu \neq 0$, $\sum_{j=0}^{k+1} \mu A^j v = 0$, so, by restricting $\mu$ to the appropriate dimensions, $\sum_{j=0}^{k+1} \mu' A_i^j v = 0$. Now, by proposition 3.4, $k \leq \dim \mathrm{Im}\, A_i$ and thus, by considering the above argument for all $i$, $k \leq \max\{\dim \mathrm{Im}\, A_i\}$. □

With the above set-up, let us state the *cyclic decomposition theorem*.

**Theorem 20** (Cyclic Decomposition Theorem)**.** Let $V$ be a finite dimensional vector space over the field $\mathbb{F}$ and let $T \in \mathrm{End}(V)$. Then if $m_T = f^k$ for some irreducible $f \in \mathbb{F}[X]$, there exists some $v_1, \cdots, v_r \in V$ such that

$$V = \bigoplus_{i=1}^{r} Z(v_i, T),$$

where $Z(v_i, T)$ has $T$-annihilator $f^{k_i}$ for all $i = 1, \cdots, r$ and $k = k_1 \geq k_2 \geq \cdots \geq k_r$. Furthermore, $k_i$ for all $i$ is uniquely determined by $T$.

Straight away, we see that this theorem can be generalised to arbitrary endomorphism by the primary decomposition theorem, that is, by the same argument as the proof of the JCF theorem, the general case can be reduced to the above case. Furthermore, by considering that the change of basis matrix of $T$ with respect to its cyclic subspace is the companion matrix of its annihilator, we have the following corollary.

**Corollary 20.1.** Let $T$ be as described in theorem 20. Then there exists a basis $B$ of $V$ such that

$$[T]_B = \bigoplus_{i=1}^{r} C(f^{k_i}).$$

In fact, the cyclic decomposition theorem implies the JCF theorem. By the same argument as before, we can simply consider the nilpotent case. Suppose $A$ is some nilpotent matrix, then there exists some $k$ such that $m_A(x) = x^k$, so by the cyclic decomposition theorem, $A \sim \bigoplus C(x^{k_i})$. But $C(x^{k_i}) = (J_{k_i}(0))^T \sim J_{k_i}(0)$, so we are done!

As we have seen, the cyclic decomposition theorem is a powerful theorem. However, unlike the proof of the JCF and the primary decomposition theorems, as we shall see, the proof the the cyclic decomposition theorem is non-constructive, that is, we will not end up with an algorithm for computing the cyclic decomposition basis in the end, so it is less useful in that regard.

*Proof.* (Cyclic Decomposition Theorem). We induct on the dimension of $V$. Again the theorem is trivial for the base case so let us consider the case whenever $\dim V = n$ where the cyclic decomposition theorem is true for all $\dim V' = k < n$.

By assumption, there exists some $f \in \mathbb{F}[X]$, $k \in \mathbb{N}$ such that $m_T = f^k$ and $f$ is irreducible. By the minimality of $m_T$, there exists some $v_1 \in V$ such that $f(T)^{k-1}(v_1) \neq 0$. Furthermore, by proposition 3.3, we have the $T$-annihilator of $v_1$ is $f^k$ (since $m_{T_{v_1}} \mid m_T = f^k$ and for all $l < k$, $f^l$ is not the $T$-annihilator). Thus, by defining $Z_1 = Z(v_1, T)$, we have the first component of the decomposition.

Now let $\bar{V} = V/Z_1$ and let $\bar{T} : \bar{V} \to \bar{V}$ be the quotient map of $T$. By recalling that we can decompose the matrix representation of $T$ as

$$[T]_B = \left[\begin{array}{c|c} [T_{Z_1}]_{B_{Z_1}} & A \\ \hline \mathbf{0} & [\bar{T}]_{\bar{B}} \end{array}\right]$$

we see that the $m_{\bar{T}} \mid m_T = f^k$, so $m_{\bar{T}} = f^l$ for some $l \leq k$. Then, as $\dim \bar{V} < \dim V$, we can apply the inductive hypothesis to $\bar{T}$, so there exist $\bar{w_i} := w_i + Z_1$ such that $\bar{V} = \bigoplus Z(\bar{w_i}, \bar{T})$ and $\bar{w_i}$ has $\bar{T}$-annihilator $f_i^k$ where $i = 2, \cdots r$ and $n - \dim Z_1 = k_2 \geq \cdots k_r$. However, theses $\bar{w_i}$ are vectors in the quotient space so we need to find some ways to "lift" them back to $V$ such that they preserve their properties in $\bar{V}$.

I claim that there exists some $v_2 \in Z_1 + w_2$ such that $v_2$ has $T$-annihilator $f^{k_2}$. To see this, we considering that, as $f^{k_2}$ is the $\bar{T}$ annihilator of $\bar{w}_2$, by definition $f(\bar{T})^{k_2}(\bar{w_2}) = 0_{\bar{V}}$ and so, for all $v \in \bar{w}_2$, $v + Z_1 = \bar{w}_2$, thus,

$$0_{\bar{V}} = f(\bar{T})^{k_2}(v + Z_1) = f(T)^{k_2}(v) + Z_1,$$

implying $f(T)^{k_2}(v) \in Z_1$. Hence, as $Z_1$ is a cyclic subspace of $V$, there exists some $g \in \mathbb{F}[X]$ such that $f(T)^{k_2}(v) = g(T)(v_1)$. Then, $0 = 0v = f(T)^k(v) = f(T)^{k-k_2}g(T)(v_1)$ and so, $f(T)^{k-k_2}g(T)(v_1)$ is a $T$-annihilator of $v_1$. But, as $m_{v_1} = f^k$, $f_k \mid f^{k-k_2}g$, and by the irreducibility of $f$, there exists some $h \in \mathbb{F}[X]$ such that $g = f^{k_2}h$. Finally, by defining $v_2 := v - h(T)(v_1)$, we have

$$f(T)^{k_2}(v_2) = f(T)^{k_2}(v - h(T)(v_1)) = f(T)^{k_2}(v) - g(T)(v_2) = 0,$$

and hence $f^{k_2}$ is a $T$-annihilator of $v_2$. (We see that this is part of the proof that uses the axiom of choice making it non-constructive.)

By similar process, we see that for all $i$, there exists $v_i \in w_i + Z_1$ with $T$-annihilator $f^{k_1}$. So it suffices to show that $V = \bigoplus Z_i$ where $Z_i = Z(v_i, T)$ and thus, we need two show $\dim V = \sum \dim Z_i$ and $V = \sum Z_i$.

To show that $\dim V = \sum \dim Z_i$, let us consider the subspaces $Z_i$ and $\bar{Z}_i := \{\bar{z} \mid z \in Z_i\} = Z(\bar{w}_i, \bar{T})$. By recalling our inductive hypothesis and the above claim, $Z_i$ and $\bar{Z}_i$ has annihilator $f^{k_i}$. Thus, as the dimension of a cyclic subspace is the degree of its annihilator, we have $\dim Z_i = \dim \bar{Z}_i = k_i \deg f$. Now, since by the inductive hypothesis we have $\bar{V} = \bigoplus \bar{Z}_i$, so $\dim \bar{V} = \sum_{i=2}^r \dim \bar{Z}_i = \sum_{i=2}^r \dim Z_i$, and hence, $\dim V = \dim Z_1 + \dim \bar{V} = \dim Z_1 + \sum \dim_{i=2}^r Z_i = \sum \dim Z_i$.

Lastly, it remains to show that $V = \sum Z_i$. But for all $v \in V$, either $v \in Z_1$, or $\bar{v} \in \bigoplus_{i=2}^r \bar{Z}_i \setminus \{0\}$, since $V/Z_1 = \bigoplus_{i=2}^r \bar{Z}_i$. Thus, it is the sum of non-trivial vectors $u_i$ for $u_i \in Z_i$, and hence, we are done with the existence part of the proof!

To show uniqueness, we use a similar argument which we had used for the uniqueness of the JCF theorem. Let $V = \bigoplus_{i=1}^r Z_i$ where $Z_i$ has $T$-annihilator $f^{k_i}$ where $k = k_1 \geq \cdots \geq k_r$ and suppose $n_j$ is the number of $Z_i$ with $T$-annihilator $f^j$. Then by applying $f(T)^{k-1}$ to $V = \bigoplus_{i=1}^r Z_i$, we have

$$f(T)^{k-1}(V) = \bigoplus_{i=1}^{n_k} f(T)^{k-1}(Z_i)$$

since all smaller $Z_i$ will be annihilated by $f(T)^{k-1}$. Now, as each subspace $f(T)^{k-1}(Z_i)$ is cyclic and has $T$-annihilator $f(t)$, it has dimension $\deg f$. Thus, $\dim f(T)^{k-1}(V) = n_k \deg f$. Similarly, we see $\dim f(T)^{k-2}(V) = 2n_k \deg f + n_{k-1} \deg f$ and so forth, thus, uniquely determining $n_j$ for all $j$. □

## 3.4 Rational Canonical Form

With that, we can finally state and prove the rational canonical form theorem.

**Theorem 21** (Rational Canonical Form Theorem)**.** Let $V$ be a finite dimensional vector space over some field $\mathbb{F}$, and let $T \in \mathrm{End}(V)$ with $m_T = \prod_{i=1}^t f_i^{k_i}$ where $f_i$ are distinct irreducible polynomials in $\mathbb{F}[X]$. Then there exists some basis $B$ of $V$ such that

$$[T]_B = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{r_i} C(f_i^{k_{i_{r_j}}}),$$

where for all $i = 1, \cdots, t$, $k_i = k_{i_1} \geq k_{i_2} \geq \cdots \geq k_{i_{r_i}}$, and further more, these $k_{i_j}$ for all $j = 1, \cdots, r_i$ are uniquely determined.

We call this decomposition the rational canonical form of $T$ and for all $i, j$, $f_i^{k_{ij}}$ the elementary divisors of $T$.

*Proof.* This proof is rather straight forward in comparison to the cyclic decomposition theorem by considering corollary 20.1.

By the primary decomposition theorem, if $V_i = \ker f_i^{k_i}$, we have $V = \bigoplus V_i$, and furthermore, each restriction $T_{V_i}$ has minimlal polynomila $f_i^{k_i}$. So, by corollary 20.1, there exists some $B_i$, a basis of $V_i$, such that

$$[T_{V_i}]_{B_i} = \bigoplus_{j=1}^{r_i} C(f_i^{k_{ir_j}})),$$

and thus, by letting $B = \bigcup B_i$,

$$[T]_B = \bigoplus_{i=1}^{t} \bigoplus_{j=1}^{r_i} C(f_i^{k_{ir_j}}).$$

□

In some literature, the rational canonical form refers to a different decomposition where the elementary divisors are not necessarily irreducible; instead it shows that every $n \times n$ matrix over $\mathbb{F}$ is similar to a unique matrix of the form $\bigoplus C(g_i)$ where $g_i \in \mathbb{F}[X]$ and for all $i$, $g_i \mid g_{i+1}$ (see problem sheet 7).

Recall the general linear group $G = GL_n(\mathbb{F})$ and for all $g \in G$ we will denote its conjugation class by $[g] := \{h \in G \mid g \sim h\}$. Then, by the rational canonical form theorem, as the RCF of a matrix is unique, there exists a bijection between the conjugacy classes of $GL_n(\mathbb{F})$ and the set of $n \times n$ invertible RCF matrices.

# 4 Geometry

## 4.1 Dual Space

**Definition 4.1** (Linear Functional)**.** Let $V$ be a vector space over some field $\mathbb{F}$. Then, a linear functional is a linear map[1] $\phi : V \to \mathbb{F}$.

Some typical examples of linear functionals are the zero map $\mathbf{0}$, the projection map $\pi_i$ and the trace function $\mathrm{tr} : M_n(\mathbb{F}) \to \mathbb{F} : A \mapsto \mathrm{tr}(A)$. We also observe that linear functionals are closed under addition and scalar multiplication. This is helpful as we would like to create a vector space from these linear functionals.

**Definition 4.2** (Dual Space)**.** Let $V$ be a vector space over some field $\mathbb{F}$. Then the dual space $V^*$ of $V$, is the vector space consisting of the linear functionals of $V$ with natural addition and scalar multiplication.

**Proposition 4.1.** Let $V$ be a finite dimensional vector space with dimension $n$. Then $\dim V^* = n$.

*Proof.* Let $B := \{v_1, \cdots, v_n\}$ be a basis of $V$ and let us define the linear functionals $\phi_i : V \to \mathbb{F} : \sum_{j=1}^{n} \lambda_j v_j \mapsto \lambda_j$ (it is obvious that $\phi_i$ is well-defined and is a linear map). Now, by defining $B^* := \{\phi_1, \cdots, \phi_n\}$, I claim that $B^*$ spans $V^*$ and is linearly independent.

Let $\phi \in V^*$ and suppose we denote $w_i := \phi(v_i)$. Then for all $v = \sum \mu v_i \in V$, $\phi(v) = \phi(\sum \mu v_i) = \sum \mu \phi(v_i) w_i = \sum w_i \phi_i(\mu v_i) = \sum w_i \phi_i(\sum \mu v_j) = \sum w_i \phi_i(v)$, where the second to last equality is true as all other components but $\mu v_i$ evaluates to zero.

Now, let us prove linear independence. Suppose that there exists $\mu$ such that $\sum \mu \phi_i = 0$. Then $\sum \mu \phi_i(v_j) = 0$ for some $j = 1, \cdots, n$. But then, $0 = \sum \mu \phi_i(v_j) = \mu_j$, and so $\mu = 0$ as $j$ was chosen arbitrarily. □

The basis $B^*$ in the proof above is in general refereed to as the dual basis of $B$.

**Definition 4.3** (Annihilators)**.** Let $V$ be a vector space over the field $\mathbb{F}$, then for some $X \subseteq V$, the annihilator of $X$ is

$$X^\circ = \{\phi \in V^* \mid \phi(X) = 0\}.$$

It is easy to see that for all $X \subseteq V$, the annihilator of $X$ is a subspace of the dual space of $V$.

**Proposition 4.2.** Let $V$ be a vector space over the field $\mathbb{F}$ with subspace $W$, then $\dim W^\circ = \dim V - \dim W$.

*Proof.* Let $r = \dim W$ and $B := \{w_1, \cdots, w_r\}$ be a basis of $W$. Furthermore, suppose we extend $B$ to a basis of $V$, $B' := \{w_1, \cdots, w_r, v_1, \cdots, v_s\}$. Then, there exists a dual basis of $B'$, $\{\phi_1, \cdots, \phi_r, \sigma_1, \cdots, \sigma_s\}$. Consider that, for all $w \in W, w = \sum \mu B' = \sum \mu w_1$ for some $\mu$, that is, the coefficients for $v_i = 0$, we have $\sigma_i(W) = 0$, so $\sigma_i \in W^\circ$. Now, as it is trivial that $\{\sigma_i \mid i\}$ is linearly independent, it remains to show this set spans $W^\circ$. Let $\sigma \in W^\circ$,

[1] We recall that a field is a vector space over itself.

then there exists $\lambda, \mu$ such that $\sigma = \sum \lambda\phi_i + \sum \mu\sigma_i$. Then, as $\sigma$ is a $W$-annihilator, for all $i$, $0 = \sigma(w_i) = \sum \lambda\phi_j(w_i) + \sum \mu\sigma_j(w_i) = \sum \lambda\phi_j(w_i) = \lambda_i\phi_i(w_i)$ implying $\lambda_i = 0$. Thus, $\sigma = \sum \mu\sigma_i$, and hence, $\{\sigma_i \mid i\}$ spans $W^\circ$ and is a basis of $W^\circ$. □

Alternatively, one can prove the above proposition by showing $W^\circ \cong (V/W)^*$ or by considering rank-nullity on the map $r : V^* \to W^* : \phi \mapsto \phi\,|_W$.

## 4.2 Inner Product Spaces

**Definition 4.4** (Inner Product)**.** Let $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$, and let $V$ be a vector space over $\mathbb{F}$. An inner product on $V$, is a map, $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{F}$, such that,

- for all $\lambda_1, \lambda_2 \in \mathbb{F}$, $v_1, v_2, w \in V$ $\langle\lambda_1 v_1 + \lambda_2 v_2, w\rangle = \lambda_1\langle v_1, w\rangle + \lambda_2\langle v_2, w\rangle$ (left-linear);
- for all $v, w \in V$, $\langle v, w\rangle = \overline{\langle w, v\rangle}$ where $\bar{z}$ is the complex conjugate of $z$ (Hermitian symmetry);
- for all $v \in V \setminus \{0\}$, $\langle v, v\rangle > 0$ (positive definite).

A vector space equipped with an inner product is a (real or complex) inner product space.

**Remark.** The first two conditions of an inner product forms a Hermitian form. Therefore, we can say that an inner product is simply a positive definite Hermitian form.

Straight away we see that $\langle 0, v\rangle = 0$ and $\langle v, \lambda_1 w_1 + \lambda_2 w_2\rangle = \overline{\lambda_1}\langle v, w_1\rangle + \overline{\lambda_2}\langle v, w_2\rangle$. Furthermore, we see that if $\mathbb{F} = \mathbb{R}$, then the inner product is symmetric and right-linear as well. Lastly, we see that if $\langle u, v\rangle = \langle u, w\rangle$ for all $u \in V$, then $v = w$ (since, $\langle v - w, v - w\rangle = \langle v, v - w\rangle - \langle w, v - w\rangle = \overline{\langle v - w, v\rangle} - \overline{\langle v - w, w\rangle} = \overline{\langle v - w, v\rangle - \langle v - w, w\rangle} = 0$).

Some typical examples of inner products are of course the dot product on $\mathbb{R}^n$ or $\mathbb{C}^n$, $\langle\cdot,\cdot\rangle : \mathcal{C}([a,b],\mathbb{R})^2 \to \mathbb{R} : (f,g) \mapsto \int_a^b fg$, and $\langle\cdot,\cdot\rangle : M_{m,n}(\mathbb{C})^2 \to \mathbb{C} : (A,B) \mapsto \operatorname{tr}(B^T\bar{A})$.

There is a very natural way of defining the matrix of a inner product on a vector space with respect to some basis. Let $V$ be a finite dimensional inner product space and suppose $B = \{v_1, \cdots, v_n\}$ is a basis of $V$, then the matrix associated with the inner product is defined to be the matrix $A$ with $[A]_{i,j} = (v_i, v_j)$. Straight away, we see that $A$ is Hermitian, that is $A^T = \overline{A}$ since $a_{j,i} = \overline{a_{i,j}}$ and we find $\langle u, v\rangle = [u]_B^T A\overline{[v]}_B$.

**Definition 4.5.** A Hermitian matrix $A$ such that $x^T Ax > 0$ for all $x \neq 0$ is called positive definite.

**Proposition 4.3.** A Hermitian matrix $A$ has only real eigenvalues and it is positive definite if and only if every eigenvalue of $A$ is positive.

*Proof.* See problem sheet 8. □

We recall from analysis that the inner product on the vector space $V$ induces a norm with $\|v\| := \sqrt{\langle v, v\rangle}$ which in turn induced a metric on $V$ with $d(x,y) := \|x - y\|$. This results in the classical inequalities – Cauchy-Schwarz and the Triangle inequality.

**Proposition 4.4.** Let $V$ be a inner product space and let $v, w \in V$, then

- $|\langle v, w\rangle| \leq \|v\|\|w\|$;
- $\|v + w\| \leq \|v\| + \|w\|$;
- $\|v - w\| \leq \|u - w\| + \|w - v\|$.

*Proof.* Same proof as last year. □

Let us now consider the connection between the dual spaces and the inner product spaces.

**Proposition 4.5.** Let $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$ and suppose $V$ is an inner product space over $\mathbb{F}$. Then for $v \in V$, define $f_v : V \to \mathbb{F}$ by,

$$f_v(w) = \langle w, v\rangle,$$

for all $w \in V$. Then $V^* = \{f_v \mid v \in V\}$.

$\{f_v \mid v \in V\} \subseteq V^*$ trivially so it suffices to show that $V^* \subseteq \{f_v \mid v \in V\}$. To prove this, we might instinctively try to show that $\{f_v \mid v \in B\}$ for $B$ some basis of $V$ is linearly independent. However, we immediately becomes stuck as inner products are not necessarily right linear. So instead, let us consider the following vector space.

Given a vector space $V = (V, \mathbb{F}, +, \times)$, denote $\bar{V}$ as the vector space $(V, \mathbb{F}, +, \otimes)$ where

$$\otimes : \mathbb{F} \times V \to V : (\lambda, v) \mapsto \bar{\lambda} \times v.$$

It is a routine check to see that $\bar{V}$ is a vector space, and since $\sum \mu B = 0 \iff \sum \bar{\mu} B = 0$, we have $\dim V = \dim \bar{V}$.

*Proof.* (Proposition 4.5). Consider $\pi : \bar{V} \to V^* : v \mapsto f_v$. We see that $\pi$ is a linear map since for all $\lambda, \mu \in \mathbb{F}, v, w \in \bar{V}$,

$$\begin{aligned}
\pi(\lambda \otimes v + \mu \otimes w) &= u \mapsto \langle u, \bar{\lambda} v + \bar{\mu} w\rangle \\
&= u \mapsto \overline{\langle \bar{\lambda} v + \bar{\mu} w, u\rangle} \\
&= u \mapsto \overline{\bar{\lambda}\langle v, u\rangle} + \bar{\mu}\langle w, u\rangle \\
&= u \mapsto \lambda\langle u, v\rangle + \mu\langle u, w\rangle \\
&= \lambda\pi(v) + \mu\pi(u).
\end{aligned}$$

Furthermore, as $\ker \pi = \{0\}$, $\pi$ is injective and by the rank-nullity theorem $\dim \operatorname{Im}(\pi) = \dim \bar{V} = \dim V^*$ implying $\pi$ is surjective (since if otherwise, we can find some $f \notin \operatorname{Im}(\pi)$ implying $\operatorname{Im}(\pi) < \operatorname{sp}(B \cup \{f\}) \leq V^*$ #). Hence, $\pi$ is bijective and is a vector space isomorphism between $\bar{V}$ and $V^*$. But, as $\{f_v \mid v \in V\} = \operatorname{Im}(\pi) = V^*$, we are done! □

Let us now extend some definitions regarding orthogonality from last year to inner product spaces.

**Definition 4.6** (Orthogonal)**.** Let $V$ be an inner product space. We say vectors $v, w \in V$ are orthogonal if and only if $\langle u, v\rangle = 0$.

**Definition 4.7** (Orthogonal and Orthonormal Set)**.** A set of vectors $S := \{v_1, \cdots, v_n\} \subseteq V$ where $V$ is an inner product space is called orthogonal if and only if for all $i, j = 1 \cdots n$, if $i \neq j$ then $\langle v_i, v_j \rangle = 0$. Furthermore, if $S$ is orthogonal and for all $i = 1 \cdots n$, $\|v_i\| = \sqrt{\langle v_i, v_i \rangle} = 1$, we call $S$ an orthonormal set.

We see that the definition of an orthonormal set can be represented nicely with the Kronecker delta where $\langle v_i, v_j \rangle = \delta_{ij}$.

**Definition 4.8** (Orthogonal Complement)**.** Let $V$ be an inner product space and suppose $W \subseteq V$, then the orthogonal complement of $W$ is

$$W^\perp = \{v \in V \mid \langle v, w \rangle = 0, \forall w \in W\}.$$

**Proposition 4.6.** Let $V$ be a finite dimensional inner product space and suppose $W \leq V$ is a subspace of $V$, then $V = W \oplus W^\perp$.

*Proof.* Recall the annihilator space $W^\circ = \{f \in V^* \mid f(w) = 0, \forall w \in W\} = \{f_v \in V^* \mid v \in W^\perp\}$, so $\dim W^\perp = \dim W^\circ = \dim V - \dim W$. Thus, it remains to show that $W \cap W^\perp = \{0\}$. But this is trivial as for all $v \in W \cap W^\perp$, $\langle v, v \rangle = 0$, so $v = 0$. □

A direct consequence of the above proposition is an important result we knew from last year.

**Theorem 22.** Let $V$ be a finite dimensional inner product space then,

- $V$ has an orthonormal basis;
- any orthonormal set $S := \{v_1, \cdots, v_r\}$ can be extended to an orthonormal basis.

*Proof.* We prove that $V$ has an orthonormal basis by induction on $\dim V = n$. The theorem is trivial for $n = 1$ so let us assume the theorem for inner products spaces with dimension less than $n$. Let $v_1 \in V$ be a vector with $\|v_1\| = 1$ and suppose we define $W = \text{sp}(v_1)$. Then, by proposition 4.6,

$$V = W \oplus W^\perp,$$

implying $\dim W^\perp = n - 1$. So by the inductive hypothesis, $W^\perp$ has an orthonormal basis $B$. Now, by the properties of direct sum, we have $B \cup \{v_1\}$ is a basis of $V$, and as $v_1$ is orthogonal to all elements of $W^\perp$, $B \cup \{v_1\}$ is in fact an orthonormal basis.

Similarly for the second part of the theorem, we let $W = \text{sp}(S)$ and we have $V = W \oplus W^\perp$. Now, by invoking part 1, we have a orthonormal basis of $W^\perp$, $B$. Thus, $B \cup S$ is an orthonormal basis of $V$. □

Of course, this theorem was already a direct result of last years linear algebra module where we found an algorithm for finding such orthonormal bases – the Gram-Schmidt process. We shall quickly recall the process here.

1. Start with any basis of $V$, $\{v_1, \cdots, v_n\}$;
2. Let $u_1 = v_1/\|v_1\|$;
3. Define $u_{k+1} = (v_{k+1} - \sum_{i=1}^{k} \langle v_k, u_i \rangle u_i)/\|(v_{k+1} - \sum_{i=1}^{k} \langle v_k, u_i \rangle u_i)\|$ for all $k$.

Then, $\{u_1, \cdots, u_n\}$, is a orthonormal basis of $V$ with the nice property that $\text{sp}(u_1, \cdots, u_i) = \text{sp}(v_1, \cdots, v_i)$.

**Proposition 4.7.** Let $V$ be a finite dimensional inner product space with an orthonormal basis $B := \{u_1, \cdots, u_n\}$. Then for all $v \in V$, $v = \sum_{i=1}^{n} \langle v, u_i \rangle u_i$, and we call elements of the set $\{\langle v, u_i \rangle \mid i = 1, \cdots, n\}$ the Fourier coefficients of $v$. Furthermore, $\|v\|^2 = \sum_{i=1}^{n} \langle v, u_i \rangle^2$.

*Proof.* Since $B$ is a basis, there exists some $\lambda$ such that $v = \sum \lambda u_i$. Now, by taking the inner product of $v$ with $u_i$, we have $\langle v, u_i \rangle = \langle \sum \lambda u_j, u_i \rangle = \sum \lambda \langle u_j, u_i \rangle = \lambda_i$ and we are done!

For the second part, we see that $\|v\|^2 = \langle v, v \rangle = \langle \sum_{i=1}^{n} \langle v, u_i \rangle u_i, \sum_{j=1}^{n} \langle v, u_j \rangle u_j \rangle = \sum_{i=1}^{n} \langle v, u_i \rangle \langle u_i, \sum_{j=1}^{n} \langle v, u_j \rangle u_j \rangle = \sum_{i=1}^{n} \langle v, u_i \rangle \overline{\sum_{j=1}^{n} \langle v, u_j \rangle \langle u_j, u_i \rangle} = \sum_{i=1}^{n} \langle v, u_i \rangle \overline{\langle v, u_i \rangle} = \sum_{i=1}^{n} \langle v, u_i \rangle^2$. □

As can be imagined by the name "Fourier coefficients", there is a connection between the above proposition and the Fourier series. Consider the vector space $V$ over $\mathbb{R}$ (works for $\mathbb{C}$ also) consisting of continuous functions with domain $[0, \pi]$, i.e. $V = \mathcal{C}([0, \pi], \mathbb{R})$. We recall the natural inner product on continuous functions,

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R} : (f, g) \mapsto \frac{2}{\pi} \int_0^{\pi} fg.$$

By manually checking, we find the set $\{1/2, \cos x, \cos 2x, \cdots\}$ is an orthonormal set of $V$. Consider that for all $f \in V$,

$$\langle f, \cos nx \rangle = \frac{2}{\pi} \int_0^{\pi} f(x) \cos nx dx,$$

we have found the Fourier cosine coefficients of the Fourier series. The similar can be done to the sine coefficients.

**Definition 4.9** (Orthogonal Projection Map)**.** Let $V$ be an inner product space and suppose $v, w \in V \setminus \{0\}$. Then the projection of $v$ along $w$ is the vector $\frac{\langle v, w \rangle}{\langle w, w \rangle} w$. More generally, for all $W \leq V$, recall that $V = W \oplus W^{\perp}$; so there exist unique $w \in W$, $w' \in W^{\perp}$ such that $v = w + w'$. Thus, we define the orthogonal projection map $\pi_W : V \to W$ such that $\pi_W(v) = w$.

To see why the first vector is called a projection, we consider two vectors in Euclidean spaces equipped with the dot product as the inner product. By drawing the vectors explicitly, we see that the projection of $v$ along $w$ is exactly what the name describes, that is the projection of $v$ in the direction of $w$.

**Lemma 4.1.** Let $V$ be a inner product space and suppose $u, v \in V$ such that $\langle u, v \rangle = 0$, then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.

*Proof.* Follows straight away from definition.

$$\|u + v\|^2 = \langle u + v, u + v \rangle = \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle = \langle u, u \rangle + \langle v, v \rangle + \|u\|^2 + \|v\|^2.$$

□

**Proposition 4.8.** Let $V$ be an inner product space and suppose $v \in V \setminus \{0\}$ and $W \leq V$. Then $\pi_W(v)$ is the vector in $W$ closest (with respect to the induced metric) to $v$, that is for all $w \in W$, $\|v - \pi_W(v)\| \leq \|v - w\|$. Furthermore, if $\{v_1, \cdots, v_r\}$ is an orthonormal basis of $W$, then

$$\pi_W(v) = \sum \langle v, v_i \rangle v_i.$$

*Proof.* Let $w \in W$, then $\|v - w\|^2 = \|(v - \pi_W(v)) + (\pi_W(v) - w)\|^2$. Since $v - \pi_W(v) \in W^\perp$ and $\pi_W(v) - w \in W$ we have $\|v - w\|^2 = \|v - \pi_W(v)\|^2 + \|\pi_W(v) - w\|^2 \geq \|v - \pi_W(v)\|^2$.

For the second part, we consider that the decomposition of $v = w + w'$ is unique, thus, it suffices to show that $v - \sum \langle v, v_i \rangle v_i \in W^\perp$. Let $u \in W$, then

$$\langle v - \sum \langle v, v_i \rangle v_i, u \rangle = \langle v, u \rangle - \sum \langle v, v_i \rangle \langle v_i, u \rangle.$$

As $\{v_1, \cdots, v_r\}$ is an orthonormal basis of $W$, $u = \sum \langle u, v_i \rangle v_i$, so,

$$\langle v - \sum \langle v, v_i \rangle v_i, u \rangle = \langle v, u \rangle - \sum_i \langle v, v_i \rangle \sum_j \overline{\langle w, v_j \rangle} \langle v_i, v_j \rangle = \langle v, u \rangle - \sum_i \langle v, v_i \rangle \overline{\langle w, v_i \rangle},$$

since $\langle v_i, v_j \rangle = \delta_{ij}$. Now, by considering

$$\langle v, u \rangle = \langle v, \sum \langle u, v_i \rangle v_i \rangle = \sum \overline{\langle w, v_i \rangle} \langle v, v_i \rangle,$$

and hence, $\langle v - \sum \langle v, v_i \rangle v_i, u \rangle = 0$ for all $u \in W$ implying $v - \sum \langle v, v_i \rangle v_i \in W^\perp$. □

Lastly, let us revisit some properties about the change of orthonormal basis from last year.

**Proposition 4.9.** Let $V$ be an inner product space with orthonormal bases $E := \{e_1, \cdots, e_n\}$ and $F := \{f_1, \cdots, f_n\}$. Furthermore if $P = (p_{ij})$ is the change of basis matrix such that

$$f_i = \sum_{j=1}^{n} p_{ji} e_i,$$

Then $P^T \bar{P} = I$ and we will call $P$ orthogonal if $P$ is a real matrix and unitary if it is a complex matrix.

*Proof.* Consider for all $r, s$,

$$\begin{aligned} \langle f_r, f_s \rangle &= \langle \sum_{i=1}^{n} p_{ir} e_i, \sum_{j=1}^{n} p_{js} e_j \rangle = \sum_i p_{ir} \sum_j \bar{p}_{js} \overline{\langle e_j, e_i \rangle} \\ &= \sum_i p_{ir} \bar{p}_{is} = \sum_i p^T_{ri} \bar{p}_{is} = [P^T \bar{P}]_{rs} \end{aligned}$$

Now, since $\langle f_r, f_s \rangle = 1$ is $r = s$ and 0 otherwise, we have $P^T \bar{P} = I$. □

We have thus far defined orthogonality for matrices but alike all definitions for matrices, there is an equivalent definition for linear maps. It turns out that theses orthogonal linear maps preserves distance, that is, if $T \in \text{End}(V) : v \mapsto Pv$ where $P$ is orthogonal, then

$\|T(v)\| = \|v\|$. This follows as $\|T(v)\|^2 = \|Pv\|^2 = (Pv)^T\overline{Pv} = v^T P^T \bar{P}\bar{v} = v^T v = \|v\|^2$. Alike that of metric spaces, we call such maps that preserves distance *isometries*.

By checking the axioms, we find that these isometries forms the well-known groups under composition – the orthogonal group and the unitary group, which are

$$O(n, \mathbb{R}) := \{P \in M_n(\mathbb{R}) \mid P^T P = I\},$$

and

$$U(n, \mathbb{C}) := \{P \in M_n(\mathbb{C}) \mid P^T \bar{P} = I\},$$

respectively.

## 4.3 Linear Maps on Inner Product Spaces

Recall the *spectral theorem* for symmetric matrices from first year – for all $A \in M_n(\mathbb{R})$ such that $A = A^T$, there exists some $P \in M_n(\mathbb{R})$ an orthogonal matrix such that $P^{-1}AP$ is diagonal. We will in this section generalise this result for arbitrary Inner product spaces.

**Proposition 4.10.** Let $V$ be a finite dimensional inner product space and let $T \in \mathrm{End}(V)$. Then, there exists an unique linear map $T^* \in \mathrm{End}(V)$ such that

$$\langle T(u), v\rangle = \langle u, T^*(v)\rangle,$$

for all $u, v \in V$.

*Proof.* Let $v \in V$ and let us define $h : V \to \mathbb{F} : u \mapsto \langle T(u), v\rangle$. Clearly, $h$ is linear, so $h$ is an element of the dual space of $V$, $V^*$. Then, by proposition 4.5, there exists uniquely a $v' \in V$ such that $h = f_{v'}$ where $f_{v'} : V \to \mathbb{F} : u \mapsto \langle u, v'\rangle$. So, let use define $T^* : V \to V : v \to v'$ so $\langle T(u), v\rangle = \langle u, T^*(v)\rangle$, and thus, it suffices to show $T^*$ is a linear map.

Let $\alpha, \beta \in \mathbb{F}$ and $v_1, v_2 \in V$, then

$$\begin{aligned}\langle u, T^*(\alpha v_1, \beta v_2)\rangle &= \langle T(u), \alpha v_1, \beta v_2\rangle = \bar{\alpha}\langle T(u), v_1\rangle + \bar{\beta}\langle T(u), v_2\rangle \\ &= \bar{\alpha}\langle u, T^*(v_1)\rangle + \bar{\beta}\langle u, T^*(v_2)\rangle = \langle u, \alpha T^*(v_1) + \beta T^*(v_2)\rangle.\end{aligned}$$

Since $u$ was choosen arbitrary, $T^*(\alpha v_1, \beta v_2) = \alpha T^*(v_1) + \beta T^*(v_2)$, and so $T^*$ is a linear map. □

**Definition 4.10** (Adjoint)**.** We call $T^*$ as described above the adjoint of $T$ and if $T = T^*$ then we say $T$ is self-adjoint.

The property of a linear map being self-adjoint correlates with the notion of symmetric in matrices. We see that if $T(v) = Av$ and the inner product of $u, v$ is $u^T v$, then $\langle T(u), v\rangle = \langle Au, v\rangle = (Au)^T v = u^T A^T v = \langle u, A^T v\rangle$, so $T^*(v) = A^T v$ and $T$ is self-adjoint if and only if $A^T = A$, that is $A$ is symmetric.

**Proposition 4.11.** Let $V$ be an inner product space with an orthonormal basis $B := \{v_1, \cdots v_n\}$. Let $T \in \mathrm{End}(V)$ such that $[T]_B = A = (a_{ij})$. Then $[T^*]_B = \bar{A}^T$.

*Proof.* Consider that $T(v_i) = \sum_j \langle T(v_i), v_j \rangle v_j$, and so $a_{ij} = \langle T(v_j), v_i \rangle$. Now, as $\langle T(v_j), v_i \rangle = \langle v_j, T^*(v_i) \rangle$, $([T^*]_B)_{ij} = \langle T^*(v_j), v_i \rangle = \overline{\langle v_i, T^*(v_j) \rangle} = \overline{\langle T(v_i), v_j \rangle} = \bar{a}_{ji}$. So, $[T^*]_B = \bar{A}^T$ as required. □

With this proposition, we can conclude that a self-adjoint matrix is either real symmetric or complex Hermitian.

**Theorem 23** (The Spectral Theorem)**.** Let $V$ be a finite dimensional inner product space and let $T \in \mathrm{End}(V)$ be self-adjoint, then $V$ has an orthonormal basis of eigenvectors of $T$.

This is a generalisation of last year's spectral theorem where we saw that real symmetric matrices are diagonalisable by some orthonormal matrix. Of course, we also see that, for complex matrices, we will require the matrix to be unitary.

Before proving the spectral theorem, let us first prove the following useful lemma.

**Lemma 4.2.** Let $V$ be a finite dimensional inner product space and let $T \in \mathrm{End}(V)$ be self-adjoint. Then,

- all eigenvalues of $T$ are real;
- the eigenvectors of distinct eigenvalues are orthogonal to each other;
- if $W \leq V$, is $T$-invariant, then so it $W^\perp$.

*Proof.*

1. Let $\lambda$ be an eigenvalue of $T$ with the eigenvector $v$, then $\lambda\langle v, v \rangle = \langle \lambda v, v \rangle = \langle T(v), v \rangle = \langle v, T(v) \rangle = \langle v, \lambda v \rangle = \bar{\lambda}\langle v, v \rangle$. Now as $v$ is an eigenvector, $v \neq 0$ so $\lambda = \bar{\lambda}$ and hence, $\lambda \in \mathbb{R}$.
2. Let $\lambda, \mu \in \mathbb{R}$ be distinct eigenvalues with eigenvectors $v_1, v_2$. Then $\lambda\langle v_1, v_2 \rangle = \langle \lambda v_1, v_2 \rangle = \langle T(v_1), v_2 \rangle = \langle v_1, T(v_2) \rangle = \langle v_1, \mu v_2 \rangle = \bar{\mu}\langle v_1, v_2 \rangle = \mu\langle v_1, v_2 \rangle$. Now as $\lambda \neq \mu$, we have $\langle v_1, v_2 \rangle = 0$.
3. Let $w' \in W^\perp$, then, for all $w \in W$, $\langle T(w'), w \rangle = \langle w', T(w) \rangle$. Now, as $W$ is $T$-invariant, $T(w) \in W$, so, $\langle w', T(w) \rangle = 0$ and hence, $\langle T(w'), w \rangle = 0$ and so $T(w') \in W^\perp$. Thus, $W^\perp$ is $T$-invariant.

□

With the above lemma, the spectral theorem follows straight away.

*Proof.* (The Spectral Theorem). We induct on the dimension of $V$, $\dim V = n$. If $n = 1$ then the theorem is trivial so let us assume the inductive hypothesis for all $\dim V \leq k$ and suppose $n = k + 1$.

By the fundamental theorem of algebra, $T$ must have some eigenvalue in $\mathbb{C}$ and by the above lemma, this eigenvalue is in fact real. So, suppose we denote this by $\lambda$ and let $E_\lambda$ the the eigenspace corresponding to $\lambda$. If $\dim E_\lambda = k + 1$ then every vector is an eigenvector of $T$ so we can simply choose the elementary basis; so, let us suppose otherwise – $\dim E_\lambda \leq k$.

Then, as $E_\lambda \leq V$ is $T$-invariant, we have $T\,|_{E_\lambda}$ is self-adjoint, and hence, by the inductive hypothesis, there exists some orthonormal basis of $B$ such that $[T\,|_{E_\lambda}]_B$ is diagonal.

Furthermore, by the above lemma, so is $E_\lambda^\perp$ $T$-invariant, therefore, $T\mid_{E_\lambda^\perp}$ is self-adjoint. Since $E_\lambda \oplus E_\lambda^\perp = V$, we have $\dim E_\lambda^\perp = \dim V - \dim E_\lambda \leq k$, and hence, by the inductive hypothesis, there exists some orthonormal basis $B'$ of $E_\lambda^\perp$ such that $[T\mid_{E_\lambda^\perp}]_{B'}$ is diagonal. With that, we have $B \cup B'$ forms a basis of $V$ and is in fact, orthonormal by the above lemma. Furthermore, $[T]_{B\cup B'}$ is diagonal since, $[T]_{B\cup B'} = [T\mid_{E_\lambda}]_B \oplus [T\mid_{E_\lambda^\perp}]_{B'}$. □

As the above proof is constructive, we can construct an algorithm to find such an orthonormal basis of a self-adjoint endomorphism.

1. Compute the eigenspaces of $T$;
2. Find an orthonormal basis of each eigenspace (by Gram-Schmidt);
3. Union over all the orthonormal basis.

The resulting basis is an orthonormal basis of $V$ since $V = \bigoplus E_{\lambda_i}$ (by primary decomposition and the fundamental theorem of algebra) and by part 2 of lemma 4.2.

## 4.4 Bilinear Forms

We have so far examined geometric properties of vector spaces over real and complex fields with the inner product. We will now generalise this to arbitrary fields with the bilinear forms and later the quadratic forms.

While working with the inner product space, we were required the positive definite axiom, i.e. for all $v \in V \setminus \{0\}$, $\langle v, v\rangle > 0$. Straight away, we see that this axiom cannot be be generalised to fields that are not totally ordered and so, we will drop this axiom in our definition.

**Definition 4.11** (Bilinear Form)**.** Let $V$ be a vector space of some field $\mathbb{F}$. Then the map $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{F}$ is a bilinear form if and only if it is both left and right linear, that is, for all $\alpha, \beta \in \mathbb{F}$, $v_1, v_2, w \in V$, $\langle \alpha v_1 + \beta v_2, w\rangle = \alpha\langle v_1, w\rangle + \beta\langle v_2, w\rangle$.

Beware that the bilinear form is not a strict generalisation of the inner product as we see that a inner product is not necessarily right linear.

**Proposition 4.12.** For any field $\mathbb{F}$ let $V = \mathbb{F}^n$ and suppose $A \in M_n(\mathbb{F})$, then by defining $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{F} : (u, v) \mapsto u^T A v$, $\langle\cdot,\cdot\rangle$ is a bilinear form.

The statement is true staright away by properties of matrix multiplications and it turns out, this definition gives all the bilinear forms of $\mathbb{F}^n$. By defining $(A)_{ij} := \langle v_i, v_j\rangle$, for some basis of $V$, $B$, one can immediately see that this matrix captures all bilinear forms; and for $\mathbb{F}^n$ one can simply choose $B$ as the standard basis.

**Definition 4.12** (Symmetric and Skew-Symmetric)**.** A bilinear form $\langle\cdot,\cdot\rangle$ on the vector space $V$ is symmetric if and only if for all $u, v \in V$, $\langle u, v\rangle = \langle v, u\rangle$; and is skew-symmetric if and only if for all $u, v \in V$, $\langle u, v\rangle = -\langle v, u\rangle$.

By definition, we see that, for a skew-symmetric bilinear form, for all $v$, we have $\langle v, v\rangle = -\langle v, v\rangle$, that is $2\langle v, v\rangle = 0$. This does *not* imply that $\langle v, v\rangle = 0$ since, for example in $\mathbb{F}_2$, it is possible for $\langle v, v\rangle = 1$. However, we can say that $\langle v, v\rangle = 0$ if and only if $2 \neq 0$ in

the particular field (that is $\mathbb{F}$ has characteristic[2] 2) as non-zero elements of a field has multiplicative inverses.

**Proposition 4.13.** Let $V$ be a vector space over the field $\mathbb{F}$ with the skew-symmetric bilinear form $\langle \cdot, \cdot \rangle$. If $\text{char}(\mathbb{F}) \neq 2$, then for all $v \in V$, $\langle v, v \rangle = 0$.

*Proof.* See above. □

A nice result about the symmetric and skew-symmetric bilinear forms is the following.

**Proposition 4.14.** Let $V$ be a vector space over the field $\mathbb{F}$ with the bilinear form $\langle \cdot, \cdot \rangle$. Then, $\langle \cdot, \cdot \rangle$ is symmetric or skew-symmetric if and only if for all $u, v \in V$, $\langle u, v \rangle = 0 \iff \langle v, u \rangle = 0$.

Since the above proposition demonstrates that the symmetric and skew-symmetric bilinear forms are the only bilinear forms that satisfy the particular restriction, we shall only consider symmetric and skew-symmetric bilinear forms when extending the notion of orthogonality into bilinear forms.

**Definition 4.13.** Let $V$ be a vector space over the field $\mathbb{F}$ with the symmetric/skew-symmetric bilinear form $\langle \cdot, \cdot \rangle$. Then given $W \leq V$, we define the perpendicular subspace of $W$ to be

$$W^{\perp} := \{v \in V \mid \langle v, w \rangle = 0, \forall w \in W\}.$$

Lastly, to compensate for our inability to restrict the positive definite axiom onto our bilinear forms, we define the following notion.

**Definition 4.14** (Non-Degenerate)**.** A bilinear form $\langle \cdot, \cdot \rangle$ on the vector space $V$ is non-degenerate if and only if $V^{\perp} = \{0_V\}$.

There is a easy way to see whether or not a bilinear form is degenerate, that is, it is degenerate if and only if the matrix associated with the bilinear form with respect to any basis is degenerate, i.e. it is not invertible (see problem sheet). With this, in combination with Gram-Schmidt, we can immediately conclude that all inner products are non-degenerate.

The non-degenerate condition on bilinear forms results in similar notions as that of the inner product and allows us to prove similar propositions which we had proved for inner product spaces.

**Proposition 4.15.** Let $V$ be a finite dimensional vector space over the field $\mathbb{F}$ equipped with the non-degenerate bilinear form $\langle \cdot, \cdot \rangle$. Then, for all $v \in V$, $f_v : V \to \mathbb{F} : u \mapsto \langle v, u \rangle$ is a linear functional in the dual space $V^*$ and $\phi : V \to V^* : v \mapsto f_v$ forms an vector space isomorphism. Furthermore, for all $W \leq V$, $\dim W^{\perp} = \dim V - \dim W$.

[2]We recall from *groups and rings* that the characteristic of a ring is the smallest natural number $n$ such that the sum of $n$ 1s equals 0 (the characteristic is defined to be 0 if such $n$ does not exist).

*Proof.* $\phi$ is trivially linear and so it suffices to show it is bijective. We recall that a linear map is injective if and only if it has a trivial kernel, and so, since $\langle \cdot, \cdot \rangle$ is non-degenerate, for all $v \in \ker \phi$, $\phi(v) = f_v = 0$, and so, for all $u \in V$, $\langle v, u \rangle = 0$ implying $v = 0$. Now, as $\dim V = \dim V^*$, we have $\phi$ is an isomorphism. Then, the second part of the proof follows straight away by rank-nullity. □

Note that for inner product spaces, we also showed that $V = W \oplus W^\perp$; this is not necessarily true for non-degenerate bilinear forms. To find an example this we can simply consider any skew-symmetric bilinear forms and so $\langle v, v \rangle = 0$ and hence, for all $w \in W \leq V$, $w \in W \cap W^\perp$.

For bilinear forms, we unfortunately cannot in general find a orthonormal basis of the vector space as, say for example, any skew-symmetric matrix, for all $v \in V$, $\langle v, v \rangle = 0$. So, to formulate this question properly such that there exists a nice basis which we can work with, let us first consider the following.

**Proposition 4.16.** Let $V$ be a vector space over $\mathbb{F}$ equipped with the non-degenerate bilinear form $\langle \cdot, \cdot \rangle$. Suppose we have the bases of $V$, $B_1, B_2$ and let $P$ be the change of basis matrix from $B_2$ to $B_1$, then, if $A$ is the matrix associated with the bilinear form with respect to the basis $B_1$, then $P^T AP$ is the matrix associated with the bilinear form with respect to the basis $B_2$.

*Proof.* For all $u, v \in V$, we have

$$\langle u, v \rangle = [u]_{B_1}^T [v]_{B_1} = (P[u]_{B_2})^T A (P[v]_{B_2}) = [u]_{B_2}^T (P^T AP)[v]_{B_2},$$

and so, $P^T AP$ is the matrix associated with the bilinear form with respect to the basis $B_2$. □

**Definition 4.15** (Congruent)**.** Let $\mathbb{F}$ be some field and let $A, B \in M_n(\mathbb{F})$, then we say $A, B$ are congruent if and only if there exists some $P \in M_n(\mathbb{F})$ such that $B = P^T AP$.

By routinely checking, we see that congruence if an equivalence relation and, of course, all the nice properties associated with equivalence relations.

**Theorem 24.** Let $V$ be a finite dimensional over the field $\mathbb{F}$ where $\text{char}(\mathbb{F}) \neq 2$ and suppose $\langle \cdot, \cdot \rangle$ is a non-degenerate skew-symmetric bilinear form on $V$, then

- there exists $m \in \mathbb{N}$ such that $\dim V = 2m$;
- there exists a basis $B := \{e_1, f_1, \cdots, e_m, f_m\}$ such that the matrix associated with the bilinear form with respect to $B$ is $J_m = \bigoplus_{i=1}^m A$ where

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

  that is $\langle e_i, f_i \rangle = 1$, $\langle e_i, f_i \rangle = -1$ and $\langle e_i, e_i \rangle = \langle f_i, f_i \rangle = \langle e_i, f_j \rangle = \langle f_i, e_j \rangle = 0$ for all $i \neq j$.

This basis is often called the symplectic basis and the non-degenerate skew-symmetric bilinear form itself is called a symplectic form.

*Proof.* We induct on the dimension of $V$. If $\dim V = 1$, then we we cannot have a skew-symmetric bilinear form on $V$, and thus, by *exfalso*, we have the statement is true for $\dim V = 1$. Let us now suppose the inductive hypothesis and let $\dim V = n > 1$. We can choose an arbitrary vector $e_1$ and then, as $\dim \mathrm{sp}\{e_1\}^\perp = \dim V - \dim \mathrm{sp}\{e_1\} = n - 1 < n$, $V \setminus \mathrm{sp}\{e_1\}^\perp$ is non-empty, and so, we can simply choose an arbitrary vector $f_1 \in V \setminus \mathrm{sp}\{e_1\}^\perp$ and scale it such that $\langle e_1, f_1\rangle = 1$. We see that $\{e_1, f_1\}$ is linearly independent as otherwise, there exists some $\lambda \in \mathbb{F}$ such that $f_1 = \lambda e_1$ and so, $\langle e_1, f_1\rangle = \lambda\langle e_1, e_1\rangle = 0$ as the bilinear form is skew-symmetric. #

With the above procedure, if $\dim V = 2$, then we are done, so suppose $\dim V > 2$, and let $W := \mathrm{sp}\{e_1, f_1\}$. We see that it suffices to show that $V = W \oplus W^\perp$ since, if this is the case, we can simply apply the inductive hypothesis to $W^\perp$ (we know that the restricted bilinear form on $W^\perp$ is also non-degenerate since if there exists a vector $v$ such that for all $w \in W^\perp$, then $v \in (W^\perp)^\perp = W$ contradicting the direct sum condition). Let $w \in W \cap W^\perp$, then there exists $\alpha, \beta \in \mathbb{F}$ such that $w = \alpha e_1 + \beta e_2$ and by considering $0 = \langle e_1, w\rangle = \alpha\langle e_1, e_1\rangle + \beta\langle e_1, f_1\rangle$ and $0 = \langle f_1, w\rangle = \alpha\langle f_1, e_1\rangle + \beta\langle f_1, f_1\rangle = -\alpha$, we have $\alpha = \beta = 0$ and so $w = 0$ and hence, $V = W \oplus W^\perp$. □

With this theorem, we can conclude that there exists a unique symplectic form on any even dimensional vector spaces up to congruence.

**Lemma 4.3.** Let $V$ be a finite dimensional over the field $\mathbb{F}$ where $\mathrm{char}(\mathbb{F}) \neq 2$ and suppose $\langle\cdot,\cdot\rangle$ is a non-degenerate symmetric bilinear form on $V$, then there exists some $v \in V$ such that $\langle v, v\rangle \neq 0$.

*Proof.* Suppose otherwise that for all $v \in V$ such that $\langle v, v\rangle = 0$, then, for all $u, v \in V$, $0 = \langle u + w, u + w\rangle = \langle u, u\rangle + \langle w, w\rangle + 2\langle u, w\rangle = 2\langle u, w\rangle$. So, as the characteristic of $\mathbb{F}$ is not 2, $\langle u, w\rangle = 0$ for all $u, v \in V$. But this means that the bilinear form is degenerate. # □

**Theorem 25.** Let $V$ be a finite dimensional over the field $\mathbb{F}$ where $\mathrm{char}(\mathbb{F}) \neq 2$ and suppose $\langle\cdot,\cdot\rangle$ is a non-degenerate symmetric bilinear form on $V$, then $V$ has an orthogonal basis $B := \{v_1, \cdots, v_n\}$, i.e. $B$ is a basis such that $\langle v_i, v_j\rangle = 0$ whenever $i \neq j$.

*Proof.* We induct on the $\dim V = n$. The result is clear for $n = 1$ so let us assume the inductive hypothesis for all $n \leq k$. By the above lemma, let $v_1 \in V$ such that $\langle v_1, v_1\rangle \neq 0$. Then, by defining $W := \mathrm{sp}\{v_1\}$ it suffices to show that $V = W \oplus W^\perp$ since if this is the case, we can apply the inductive hypothesis to $W^\perp$. Suppose $w \in W \cap W^\perp$ then, there exists $\lambda \in \mathbb{F}$ such that $w = \lambda v_1$ and $0 = \langle v_1, w\rangle = \lambda\langle v_1, v_1\rangle$. Since $\langle v_1, v_1\rangle \neq 0$ we have $\lambda = 0$ and so $w = 0$ and $W \cap W^\perp = \varnothing$ so $V = W \oplus W^\perp$. □

## 4.5 Quadratic Forms

As can be seen from the problem sheets, the classification of the congruence of bilinear forms is a subtle concept and depends on the field we are working with. To solve this problem, we shall consider the theory of quadratic forms. The quadratic forms can also be applied to study conics in arbitrary fields, however, this is beyond the scope of this course. Let us from this point forward assume that the field we are working with $\mathbb{F}$ does not have characteristic 2.

**Definition 4.16** (Quadratic Form)**.** Let $V$ be a finite dimensional vector space over $\mathbb{F}$. Then, $Q : V \to \mathbb{F} : v \mapsto \langle v, v\rangle$ is a quadratic form on $V$ if $\langle \cdot, \cdot \rangle$ is a symmetric bilinear form. We say $Q$ is non-degenerate if and only if $\langle \cdot, \cdot \rangle$ is non-degenerate.

With this, for every symmetric bilinear form on $V$, there is a corresponding quadratic form. The reverse of this is also true – there is a unique correspondence between symmetric bilinear forms and quadratic forms using the fact that $\langle u, v\rangle = (Q(u + v) - Q(u) - Q(v))/2$.

To see why the quadratic form is named as such, consider the general quadratic form on $\mathbb{F}^n$. As we had shown before, we have every symmetric bilinear form can be represented by $\langle u, v\rangle = u^T Av$ for some $A \in M_n(\mathbb{F})$ such that $A^T = A$. Hence, $Q(x) = x^T Ax = \sum_{i,j} a_{ij} X_i x_j = \sum_{i=1}^n a_{ii} x_i^2 + 2\sum_{i<j} a_{ij} x_i x_j$, which is a homogeneous quadratic polynomial (over the variables $\{x_1, \cdots, x_n\}$).

**Definition 4.17.** Let $Q$ and $Q'$ be quadratic forms on $\mathbb{F}^n$ such that $Q(x) = x^T Ax$ for some symmetric $A \in M_n(\mathbb{F})$ and $Q'(x) = x^T P^T APx$ for some invertible $P \in M_n(\mathbb{F})$, then we call $Q$ and $Q'$ equivalent.

As the name suggests, by checking, we see that this forms an equivalence relation on the quadratic forms.

A result that immediately follows from theorem 25 is that, if a quadratic form $Q$ on $\mathbb{F}^n$ is non-degenerate, then it is equivalent to some quadratic form $Q'(x_1, x_2, \cdots, x_n) = \sum_{i=1}^n a_i x_i^2$. With this in mind, we will now try to classify (up to equivalence) all non-degenerate quadratic forms over $\mathbb{C}$ and $\mathbb{R}$ (and some information over $\mathbb{Q}$).

**Theorem 26.** Let $V = \mathbb{F}^n$, $Q : V \to \mathbb{F}$ be a non-degenerate quadratic form. Then,

- if $\mathbb{F} = \mathbb{C}$, then, $Q$ is equivalent to the form $Q_0(x) = \sum_{i=1}^n x_i^2$ for all $x \in \mathbb{C}^n$;
- if $\mathbb{F} = \mathbb{R}$, then $Q$ is equivalent to a unique quadratic form $Q_{p,q}$ where $Q_{p,q}(x) = (\sum_{i=1}^p x_i^2) - (\sum_{i=p+1}^n x_i^2)$ and $p + q = n$ (this is known as Sylvester's law of inertia);
- if $\mathbb{F} = \mathbb{Q}$, then there are infinitely many inequivalent quadratic forms on $\mathbb{Q}^n$.

*Proof.* (Part 1) As mentioned before, we have $Q$ is equivalent to some quadratic form $Q'(x) = \sum_{i=1}^n a_i x_i^2 = x^T Ax$ where $A = \text{diag}(a_i)$. Then, by simply choosing $P = \text{diag}(\sqrt{a_i})$, we have $P^T = P$ and furthermore, $P^T I_n P = \text{diag}(a_i) = A$ and hence, by transitivity, $Q$ is equivalent to $Q_0 = x \mapsto x^T x = \sum x_i^2$.

(Part 2) Unlike working with $\mathbb{C}$, it is not necessary that $\sqrt{a_i}$ is in the reals, we we have to approach differently. Again, let us write $Q'$ to be the equivalent quadratic form such that $Q'(x) = \sum a_i x_i^2$ and suppose we order $\{a_i \mid i\}$ such that $a_1, \cdots, a_p > 0$ and $a_{p+1}, \cdots, a_{p+q=n} < 0$. Then,

$$\text{diag}(a_1, \cdots, a_n) = P^T(I_p \oplus -I_q)P,$$

where $P = \text{diag}(\sqrt{|a_i|})$ and hence, the existence of the result by transitivity.

Now, let us prove the uniqueness of this equivalence. Suppose otherwise, then there exists some $p', q'$ such that $Q \sim Q_{p,q}$ and $Q \sim Q_{p',q'}$. Let $\langle \cdot, \cdot \rangle$ be the symmetric bilinear form corresponding to $Q$. Then, as $Q \sim Q_{p,q}$, there exists an orthogonal basis of $V$, $B := \{v_i \mid i\}$,

such that $\langle v_i, v_i \rangle = 1$ if $1 \leq i \leq p$ and $-1$ if $p+1 \leq i \leq p+q = n$. Furthermore, as $Q \sim Q_{p',q'}$, there exists another orthogonal basis of $V$, $B' := \{w_i \mid i\}$ such that $\langle w_i, w_i \rangle = 1$ if $1 \leq i \leq p'$ and $-1$ if $p'+1 \leq i \leq p'+q' = n$. Then, by defining $U := \text{sp}\{v_1, \cdots, v_p\}$, and $W := \text{sp}\{w_{p'+1}, \cdots, w_n\}$, for all $u \in U \setminus \{0\}$, $Q(u) > 0$ and similarly, for all $w \in W \setminus \{0\}$, $Q(w) < 0$, thus, $U \cap W = \{0\}$. Hence, we have $\dim U \cap W \leq n$ and so $p' \geq p$. By the same argument, we can show that $p' \leq p$ and thus $p' = p$ and hence, the equivalence is unique.

(Part 3) We call a positive integer $d$ square-free if $d$ is a product of distinct primes. For a square-free integers $d$, we define the quadratic form $Q_d : \mathbb{Q}^n \to \mathbb{Q} : x \mapsto \sum_{i=1}^{n-1} x_i^2 + dx_n^2$. Then, suppose $Q_d \sim Q_{d'}$ for some square-free $d, d'$ and so $A_d \equiv A_{d'}$ (we denote congruence by $\equiv$) where $A_d = \text{diag}(1, \cdots, 1, d)$, and thus, there exists some $P \in GL_n(\mathbb{Q})$ such that $A_{d'} = P^T A_d P$. With that, by taking the determinant on both sides, we have $d' = (\det P^2)d$, implying $d' = d$ as they are square-free. It is easy to see that there is a infinite sequence of square-free integers and so, we have constructed an infinite sequence of inequivalent quadratic forms. □

Before ending this module, let us consider some applications of bilinear and quadratic forms. The theory of bilinear and quadratic forms occur naturally in many parts of mathematics most notably – geometry, number theory and algebra. In special relativity, quadratic forms $Q_{3,1}$ occur in the Minkowski space-time ($\mathbb{R}^4$). In number theory, a classical question asks that, given a rational quadratic form, $Q : \mathbb{Q}^n \to \mathbb{Q}$, and a rational number $K$, does the equation $Q(x) = K$ have a solution on $\mathbb{Q}$. Indeed, if we restrict the equation to the integers, we are asking for the solutions to a quadratic homogeneous diophantine equation. In group theory, just as we have done for inner product, we can define isometries on bilinear forms and hence receiving isometry groups on these forms.

**Definition 4.18** (Isometry)**.** Let $f : V^2 \to \mathbb{F}$ be a non-degenerate symmetric/skew-symmetric bilinear form. Then, an isometry of $f$ is a linear map $T \in \text{End}(V)$ such that for all $u, v \in V$, $f(u, v) = f(T(u), T(v))$.

With this definition, the set of isometrics, $I(V, f) := \{T \mid T \text{ is an isometry}\}$ forms a subgroup of $GL(V)$ and in particular, if $f$ is skew-symmetric, then this group is unique up to equivalence known as the symplectic group $Sp(V, f)$.