Linear Algebra

Notes by Finley Cooper

21st November 2025

Contents

1 Vector Spaces

1.1 Definitions

For this lecture course, $\mathbb{F}$ will always be field.

Definition. (Vector Space) A $\mathbb{F}$-*vector space* (or a vector space over $\mathbb{F}$) is an abelian group $(V, +, \mathbf{0})$ equipped with a function

$$\begin{aligned} \mathbb{F} \times V &\to V \\ (\lambda, v) &\to v \end{aligned}$$

which we call scalar multiplication such that $\forall v, w \in V, \forall \lambda, \mu \in \mathbb{F}$

(i) $(\lambda + \mu)v = \lambda v + \mu v$
(ii) $\lambda(v + w) = \lambda v + \lambda w$
(iii) $\lambda(\mu v) = (\lambda\mu)v$
(iv) $1 \cdot v = v \cdot 1 = v$

Remember that $\mathbf{0}$ and 0 are not the same thing. 0 is an element in the field $\mathbb{F}$ and $\mathbf{0}$ is the additive identity in V.

For an example consider $\mathbb{F}^n$ n-dimensional column vectors with entries in $\mathbb{F}$. We also have the example of a vector space $\mathbb{C}^n$ which is a complex vector space, but also a real vector space (taking either $\mathbb{C}$ or $\mathbb{R}$ as the underlying scalar field).

We also can see that $M_{m\times n}(\mathbb{F})$ form a vector space with m rows and n columns.
For any non-empty set X, we denote $\mathbb{F}^X$ as the space of functions from X to $\mathbb{F}$ equipped with operations such that:

$$\begin{aligned} f + g \text{ is given by } (f + g)(x) &= f(x) + g(x) \\ \lambda f \text{ is given by } (\lambda f)(x) &= \lambda f(x) \end{aligned}$$

Proposition. For all $v \in V$ we have that $0 \cdot v = \mathbf{0}$ and $(-1) \cdot v = -v$ where $-v$ denotes the additive inverse of v.

Proof. Trivial.

Definition. (Subspace) A *subspace* of a $\mathbb{F}$-vector space V is a subset $U \subseteq V$ which is a $\mathbb{F}$-vector space itself under the same operations as V. Equivalently, $(U, +)$ is a subgroup of $(V, +)$ and $\forall \lambda \in \mathbb{F}$, $\forall u \in U$ we have that $\lambda u \in U$.

Remark. Axioms (i)-(iv) are always automatically inherited into all subspaces.

Proposition. (Subspace test) Let V be a $\mathbb{F}$-vector space and $U \subseteq V$ then U is a subspace of V if and only if,

(i) U is nonempty.
(ii) $\forall \lambda \in \mathbb{F}$ and $\forall u, w \in U$ we have that $u + \lambda w \in U$.

Proof. If U is a subspace then U satisfies (i) and (ii) since it contains 0 and is closed. Conversely suppose that $U \subseteq V$ satisfies (i) and (ii). Taking $\lambda = -1$ so $\forall u, w \in V$, $u - w \in U$ hence $(U, +)$ is a subgroup of $(V, +)$ by the subgroup test. Finally taking $u = \mathbf{0}$ so we have that $\forall w \in U, \forall \lambda \in \mathbb{F}$ we have that $\lambda w \in U$. So U is a subspace of V. □

We notate U by $U \leq V$.

For some examples

(i)
$$\left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : x + y + z = t \right\} \subseteq \mathbb{R}^3,$$
for fixed $t \in \mathbb{R}$ is a subspace of $\mathbb{R}^3$ iff $t = 0$.

(ii) Take $\mathbb{R}^{\mathbb{R}}$ as all the functions from $\mathbb{R}$ to $\mathbb{R}$ then the set of continuous functions is a subspace.

(iii) Also we have that $C^\infty(\mathbb{R})$, the set of infintely differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ is a subspace of $\mathbb{R}^{\mathbb{R}}$ and the subspace of continuous functions.

(iv) A further subspace of all of those subspaces is the set of polynomial functions.

Lemma. For $U, W \leq V$ we have that $U \cap W \leq V$.

Proof. We'll use the subspace test. Both U, W are subspaces so they contain $\mathbf{0}$ hence $\mathbf{0} \in U \cap W$ so $U \cap W$ is nonempty. Secondly take $x, y \in U \cap W$ with $\lambda \in \mathbb{F}$. Then $U \leq V$ and $x, y \in U$ so $x + \lambda y \in U$. Similarly with W so $x + \lambda y \in W$ hence we have that $x + \lambda y \in U \cap W$ hence $U \cap W \leq V$ □

Remark. This does not apply for subspaces, in fact from IA Groups, we know it doesn't even hold for the underlying abelian group.

Definition. (Subspace sum) For $U, W \leq V$, the *subspace sum* of U, W is
$$U + W = \{u + w : u \in U, w \in W\}.$$

Lemma. If $U, W \leq V$ then $U + W \leq V$.

Proof. Simple application of the subspace test.

Remark. $U + W$ is the smallest subgroup of U, W in terms of inclusion, i.e. if K is such that $U \subseteq K$ and $W \subseteq K$ then $U + W \subseteq K$.

1.2 Linear maps, isomorphisms, and quotients

Definition. (Linear map) For V, W $\mathbb{F}$-vector spaces. A *linear map* from V to W is a group homomorphism, φ, from $(V, +)$ to $(W, +)$ such that $\forall v \in V$
$$\varphi(\lambda v) = \lambda \varphi(v)$$

Equivalently to show any function $\alpha : V \to W$ is a linear map we just need to show that $\forall u, w \in V$, $\forall \lambda \in \mathbb{F}$ we have

$$\alpha(u + \lambda w) = \alpha(u) + \lambda\alpha(w).$$

For some examples of linear maps

(i) $V = \mathbb{F}^n, W = \mathbb{F}^m$ $A \in M_{m\times n}(\mathbb{F})$. Then let $\alpha : V \to W$ be given by $\alpha(v) = Av$. Then α is linear.

(ii) $\alpha : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ defined by taking the derivative.

(iii) $\alpha : C(\mathbb{R}) \to \mathbb{R}$ defined by taking the integral from 0 to 1.

(iv) X any nonempty set, $x_0 \in X$,

$$\begin{aligned} \alpha : \mathbb{F}^X &\to \mathbb{F} \\ f &\to f(x_0) \end{aligned}$$

(v) For any V, W the identity mapping from V to V is linear and so is the zero map from V to W.

(vi) The composition of two linear maps is linear.

(vii) For a non-example squaring in $\mathbb{R}$ is not linear. Similiarly adding constants is not linear, since linear maps preserve the zero vector.

Definition. (Isomorphism) A linear map $\alpha : V \to W$ is an *isomorphism* if it is bijective. We say that V and W are isomorphic, if there exists an isomorphism from $V \to W$ and denote this by $V \cong W$.

An example is the vector space $V = \mathbb{F}^4$ and $W = M_{2\times 2}(\mathbb{F})$ we can define the map

$$\begin{aligned} \alpha : V &\to W \\ \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} &\to \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{aligned}$$

Then α is an isomorphism.

Proposition. If $\alpha : V \to W$ is an isomorphism then $\alpha^{-1} : W \to V$ is also an isomorphism.

Proof. Clearly α^{-1} is a bijection. We need to prove that α^{-1} is linear. Take $w_1, w_2 \in W$ and $\lambda \in \mathbb{F}$. So we can write $w_i = \alpha(v_i)$ for $i = 1, 2$. Then

$$\alpha^{-1}(w_1 + \lambda w_2) = \alpha^{-1}(\alpha(v_1) + \lambda\alpha(v_2)) = \alpha^{-1}(\alpha(v_1 + \lambda v_2)) = v_1 + \lambda v_2 = \alpha^{-1}(w_1) + \lambda\alpha^{-1}(w_2)$$

. Hence α^{-1} is linear, so α^{-1} is an isomorphism. □

Definition. (Kernal) Let V, W be $\mathbb{F}$-vector spaces. Then the *kernel* of the linear map $\alpha : V \to W$ is

$$\ker(\alpha) = \{v \in V : \alpha(v) = \mathbf{0}_W\} \subseteq V$$

Definition. (Image) Let V, W be $\mathbb{F}$-vector spaces. Then the *image* of a linear map $\alpha : V \to W$ is

$$\mathrm{im}(\alpha) = \{\alpha(v) : v \in V\} \subseteq W$$

Lemma. For a linear map $\alpha : V \to W$ the following hold.

(i) $\ker \alpha \leq V$ and $\mathrm{im}\, \alpha \leq W$

(ii) α is surjective if and only if $\mathrm{im}\, \alpha = W$

(iii) α is injective if and only if $\ker \alpha = \{\mathbf{0}_V\}$

Proof. $\mathbf{0}_V + \mathbf{0}_V = \mathbf{0}_V$, so applying α to both sides any using the fact that α is linear gives that $\alpha(\mathbf{0}_V) = \mathbf{0}_W$. So $\ker \alpha$ is nonempty. The rest of the proof is a simple application of the subspace test.

The second statement is immediate from the definition.

For the final statement suppose α injective. Suppose $v \in \ker \alpha$. Then $\alpha(v) = \mathbf{0}_W = \alpha(\mathbf{0}_w)$ so $v = \mathbf{0}_V$ by injectivity. Hence $\ker \alpha$ is trivial. Conversely suppose that $\ker \alpha = \{0_V\}$ Let $u, v \in V$ and suppose that $\alpha(u) = \alpha(v)$. The $\alpha(u - v) = \mathbf{0}_W$, so $u - v \in \ker \alpha$, so $u = v$. □

For V a $\mathbb{F}$-vector space, $W \leq V$ write

$$\frac{V}{W} = \{v + W : v \in V\}$$

as the left cosets of W in V. Recall that two cosets $v + V$ and $u + W$ are the same coset if and only if $v - u \in W$.

Proposition. V/W is an $\mathbb{F}$-vector space under operations

$$(u + W) + (v + W) = (u + v) + W$$
$$\lambda(v + W) = (\lambda v) + W$$

We call V/W the quotient space of V by W.

Proof. The proof is long and requires a lot of vector space axioms so we'll just sketch out the proof.

We check that operations are well-defined, so for $u, \overline{u}, v, \overline{v} \in V$ and $\lambda \in \mathbb{F}$ if

$$u + W = \overline{u} + W, \quad v + W = \overline{v} + W$$

then

$$(u + v) + W = (\overline{u} + \overline{w}) + W$$

and

$$(\lambda u) + W = (\lambda \overline{u}) + W$$

The vector space axioms are inherited from V. □

Proposition. (Quotient map) The function $\pi_W : V \to \frac{V}{W}$ called a *quotient map* is given by

$$\pi_W(v) = v + W$$

is a well-defined, surjective, linear map with $\ker \pi_W = W$.

Proof. Surjectivity is clear. For linearity let $u, v \in V$ and $\lambda \in \mathbb{F}$. Then

$$\begin{aligned}\pi_W(u + \lambda v) &= (u + \lambda v) + W \\ &= (u + W) + (\lambda v + W) \\ &= (u + W) + \lambda(v + W) \\ &= \pi_W(u) + \lambda \pi_W(v)\end{aligned}$$

For $v \in V$, we have that $v \in \ker \pi_W \iff \pi_W(v) = \mathbf{0}_{V/W}$. So $v + W = \mathbf{0}_V + W$ so finally $v = v - \mathbf{0}_V \in W$. □

Theorem. (First isomorphism theorem) Let V, W be $\mathbb{F}$-vector spaces and $\alpha : V \to W$ linear. Then there is an isomorphism

$$\overline{\alpha} : \frac{V}{\ker \alpha} \to \operatorname{im} \alpha$$

given by $\overline{\alpha}(v + \ker \alpha) = \alpha(v)$

Proof. For $u, v \in V$,

$$u + K = v = K \iff u - v \in K \iff \alpha(u - v) = \mathbf{0}_W \iff \alpha(u) = \alpha(v) \iff \overline{\alpha}(u + \ker \alpha) = \overline{\alpha}(v + \ker \alpha)$$

The forward direction shows that $\overline{\alpha}$ is well-defined, and the converse shows that $\overline{\alpha}$ is injective. For surjectivity given $w \in \operatorname{im} \alpha$, there exists some $v \in V$ s.t. $w = \alpha(v)$. Then $w = \overline{\alpha}(v + \ker \alpha)$. Finally for linearity given $u, v \in V$, $\lambda \in \mathbb{F}$,

$$\begin{aligned}\overline{\alpha}((u + \ker \alpha) + \lambda(v + \ker \alpha)) &= \overline{\alpha}((u + \lambda v) + \ker \alpha) \\ &= \alpha(u + \lambda v) \\ &= \alpha(u) + \lambda \alpha(v) \\ &= \overline{\alpha}(u + \ker \alpha) + \lambda \overline{\alpha}(v + \ker \alpha)\end{aligned}$$

So $\overline{\alpha}$ is linear hence is an isomorphism □

1.3 Basis

Definition. (Span) Let V be a $\mathbb{F}$-vector space. Then the *span* of some subset $S \subseteq V$ is

$$\langle S \rangle = \left\{ \sum_{s \in S} \lambda_s \cdot s : \lambda_s \in \mathbb{F} \right\}$$

where $\sum$ denotes finite sums. An expression the form above is called a *linear combination* of S.

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

Definition. (Finite-dimensional) For a vector space V we say that it is *finite-dimensional* if there exists a finite spanning set.

We'll give some simple remarks without proof.

(i) $\langle S \rangle \leq V$ and conversely if $W \leq V$ and $S \subseteq W$ then $\langle S \rangle \leq W$.

(ii) If $S, T \subseteq W$ and S spans V and $S \subseteq \langle V \rangle$ then T spans V.

(iii) By convention $\langle \emptyset \rangle = \{\mathbf{0}_V\}$.

(iv) $\langle S \cup T \rangle = \langle S \rangle + \langle T \rangle$

For an example consider $V = \mathbb{R}^3$ and consider the sets

$$S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \right\}$$

$$T = \left\{ \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 4 \end{pmatrix} \right\}$$

Then $\langle S \rangle = \langle T \rangle = \left\{ \begin{pmatrix} x \\ y \\ 2y \end{pmatrix} : x, y \in \mathbb{R} \right\} \leq \mathbb{R}^3$.

For a second example consider $V = \mathbb{R}^{\mathbb{N}}$ and set $T = \{\delta_n : n \in \mathbb{N}\}$. This is not a spanning set, since we require infinitely many elements from T to make an element in V. In fact we can write that

$$\langle T \rangle = \{f \in \mathbb{R}^{\mathbb{N}} : f(n) = 0 \text{ for all but finitely many terms}\}.$$

Definition. (Linear Independence) A subset $S \subseteq V$ is called *linearly independent* if, for all finite linear combinations

$$\sum_{s \in S} \lambda_s s \quad \text{of S}$$

if the sum is the zero vector in V the $\lambda_s = 0$ for all $s \in S$.

If S is not linearly indepedent we say that S is linearly dependent.

We'll make some more remarks

(i) If $\mathbf{0} \in S$ then S is not linearly independent.

(ii) If we have a finite set, then to show linearly independent, we only need to consider the linear combination of all elements, not all finite lienar combinations.

(iii) However is S is infinite, then we have to consider every possible finite subset of S and show it's linearly independent.

(iv) Every subset of a linearly independent set is itself linearly indepedent.

Definition. (Basis) A subset $S \subseteq V$ is a *basis* for V if S is linearly independent and a spanning set.

For an example consider $e_i \in \mathbb{F}^n$ be given by

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{with the 1 in the } i\text{th entry}$$

then the set $\{e_i : 1 \leq i \leq n\}$ is the standard basis for $\mathbb{F}^n$.

For $P(\mathbb{R})$ the set of real polynomial functions and let $p_n \in P(\mathbb{R})$ be given by $p_n(x) = x^n$, then $\{p_n : n \in \mathbb{Z}_{\geq 0}\}$ is a basis for $P(\mathbb{R})$.

Proposition. If $S \subseteq V$ is a finite spanning set, then there exists a subset $S' \subseteq S$ such that S' is a basis.

Proof. If S is linearly independent then we're done. Otherwise write $S = \{v_1, \ldots, v_n\}$. Then there exists $\lambda_1, \ldots, \lambda_n$ such that $\lambda_1 v_1 + \cdots \lambda_n v_n = \mathbf{0}$ wlog suppose that λ_n is nonzero. Then

$$v_n = -\frac{1}{\lambda_n} \sum_{i=1}^{n-1} \lambda_i v_i$$

so v_n is in the span of the other vectors. Hence $S \setminus \{v_n\}$ is still a spanning set. Repeat which the set is linearly independent, must terminate since the set is finite and the empty set is not a spanning set. □

Corollary. Every finite-dimensional vector space has a finite basis.

Proof. Trivial application of the proposition □

Theorem. (Steinitz Exchange Lemma) Let $S, T \subseteq V$ finite with S linearly independent and T a spanning set of V. Then
(i) $|S| \leq |T|$,
(ii) and there exists $T' \subseteq T$ which has size $|T'| = |T| - |S|$ and $S \cup T'$ spans V.

Proof. To come later...

Let's look at some consequences of the lemma first.

Corollary. For a finite-dimensional vector space V,
(i) Every basis for V is finite.
(ii) All finite basis have the same size.

Proof. V has a finite basis B, suppose we have some other basis B' infinite. Let $B'' \subseteq B'$ with $|B''| = |B| + 1$ then $|B''|$ is linearly independent, so applying (i) of the Steinitz exchange lemma with $S = B''$ and $T = B$ we get a contradiction.

For the second part, let B_1, B_2 be finite basis for V then apply Steinitz symmetrically since both are spanning set and linearly independent, so we get that $|B_1| \geq |B_2|$ and $|B_1| \geq |B_2|$ so $|B_1| = |B_2|$. □

Definition. (Dimension) For a vector space V the *dimension* of V is the size of any basis. We write this as $\dim V$.

This definition is well-defined by the previous corollary.

For an example $\dim \mathbb{F}^n = n$ since we've shown the standard basis has size n. As a complex vector space $\mathbb{C}$ is one-dimensional as a complex vector space and two-dimension as a real vector space, with basis $\{1\}$ and $\{1, i\}$ repectively.

Corollary. For a vector space V let $S, T \subseteq V$ finite, with S linearly independent and T a spanning set, then

$$|S| \leq \dim V \leq |T|$$

with equality if and only if S spans or V is linearly independent respectively.

Proof. The inequalities are immediate from Steinitz. If S is a basis then $|S| = \dim V$ from the previous corollary. Conversely if $|S| = \dim V$ and let B be a basis for V so we have that $|B| = |S|$ so B is a spanning set. So we can apply Steinitz (ii) to B so there exists $B' \subseteq B$ with $|B'| = |B| - |S| = 0$ and $S \cup B' = S \cup \emptyset$ spans V. So S is a basis. Similiar we have a very similar proof for equality in V. □

We will not prove that every vector space has a basis, however some non-finitely dimensional vector spaces have an infinite basis, for example $P(\mathbb{R})$.

Proposition. If V is a finite-dimensional vector space, then if $U \leq V$ then U is finite-dimensional, namely, $\dim U \leq \dim V$ with equality if and only if $U = V$.

Proof. If $U = \{\mathbf{0}\}$, we're done. Otherwise let $\mathbf{0} \neq u_1 \in U$. Then $\{u_1\} \subseteq U$ is linearly indepedent. Repeating, after repeating k times suppose we have $\{u_1, \ldots, u_k\}$ linearly indepedent with $k \leq \dim(V)$ by the previously corollary. If the set spans U we're done, if not we'll add another vector, u_{k+1} outside of the span of our space. If $\{u_1, \ldots, u_{k+1}\}$ is not linearly indepedent, we can write $\mathbf{0}$ non-trivially, so

$$\sum_{i=1}^{k+1} \lambda_i u_i = \mathbf{0}$$

with $\lambda_{k+1} \neq 0$ since $\{u_1, \ldots, u_k\}$ linearly indepedent. Thus we have that

$$u_{k+1} = -\frac{1}{\lambda_{k+1}} \left(\sum_{i=1}^{k} \lambda_i u_i \right)$$

this process must terminate after at most $\dim V$ many steps, by the previous corollary. If $\dim U = \dim V$ apply the previous corollary with S being any basis for U. □

Proposition. (Extending a basis) Let $U \leq V$. For any basis B_U of U there exists a basis B_V of V such that $B_U \subseteq B_V$.

Proof. Apply the second result from Steinitz with $S = B_U$ and T is any basis for V. We obtain that $T' \subseteq T$ s.t.

$$|T'| = |T| - |S| = \dim V - \dim U$$

and $B_V = B_U \cup T'$ spans V. But we have that

$$|B_V| \leq |B_U| + |T'| = \dim V$$

so by the previous corollary, B_V is a basis for V. □

Now we'll finally prove the Steinitz exchange lemma.

Proof. Let $S = \{u_1, \ldots, u_m\}$, $T = \{v_1, \ldots, v_n\}$ with $|T| = m$ and $|T| = n$. If S is empty then we're done. Otherwise there exists $\lambda_i \in \mathbb{F}$ such that

$$u_1 = \sum_{i=1}^{n} \lambda_i v_i$$

so by renumbering we can say that $\lambda_1 \neq 0$. Then

$$v_1 = \frac{1}{\lambda_1}\left(u_1 - \sum_{i=2}^{n} \lambda_i v_i\right)$$

So $\{u_1, v_2, \ldots, v_n\}$ spans V. After repeating k times with $k < m$ suppose $\{u_1, \ldots, u_k, v_{k+1}, \ldots, v_n\}$ spans V, then there exists $\lambda_i, \mu_j \in \mathbb{F}$ such that

$$u_{k+1} = \sum_{j=1}^{k} \mu_j u_j + \sum_{i=k+1}^{n} \lambda_i v_i$$

If for all $\lambda_i = 0$ then

$$\left(\sum_{j=1}^{k} \mu_j u_j\right) - u_{k+1} = \mathbf{0}$$

which is a contradiction since S is linearly independent. So by relabeling we have that $\lambda_{k+1} \neq 0$ such that

$$v_{k+1} = \frac{1}{\lambda_{k+1}}\left(u_{k+1} - \sum_{j=1}^{k} \mu_j u_j - \sum_{i=k+1}^{n} \lambda_i v_i\right)$$

so $(u_1, \ldots, u_{k+1}, v_{k+2}, \ldots, v_n\}$ spans V. So we can conclude that $m \neq n$ and $\{u_1, \ldots, u_m, v_{m+1}, \ldots, v_n\}$ spans V hence the set $T' = \{v_{m+1}, \ldots, v_n\}$ exists as claimed. □

Definition. (Nullity) For a linear map $\alpha : V \to W$ we define the *nullity* of α as

$$\mathrm{n}(\alpha) = \dim \ker \alpha.$$

Definition. (Rank) For a linear map $\alpha : V \to W$ we define the *rank* of α as

$$\mathrm{rk}(\alpha) = \dim \mathrm{im}\, \alpha.$$

Theorem. (Rank-nullity theorem) If V is a finite dimensional $\mathbb{F}$-vector space and W is a $\mathbb{F}$-vector space. Then if $\alpha : V \to W$ is linear then $\mathrm{im}\, \alpha$ is finite dimensional and

$$\dim V = \mathrm{n}(\alpha) + \mathrm{rk}(\alpha).$$

Proof. Recall the first isomorphism theorem so

$$\frac{V}{\ker \alpha} \cong \mathrm{im}\, \alpha$$

It is sufficient to prove the lemma

Lemma. For $U \leq V$,

$$\dim(V/U) = \dim V - \dim U$$

Proof. Let $B_U = \{u_1, \ldots, u_m\}$ be a basis of U. Extend to a basis $B_V = \{u_1, \ldots, u_m, v_{m+1}, \ldots, v_n\}$ of V where $m = \dim U$ and $n = \dim V$.
Set $B_{V/U} = \{v_i + U : m + 1 \leq i \leq n$. The we claim that $B_{V/U}$ is a basis for V/U of size $n - m$. To show spanning, for $v \in V$ write

$$v = \sum_i \lambda_i v_i + \sum_j \mu_j v_j$$

Then $v + U = \sum_i \lambda_i (v_i + U) \in \langle B_{V/U} \rangle$. For linear independence, suppose

$$\sum_i \lambda_i (v_i + U) = \mathbf{0} + U$$

hence

$$= \left(\sum_i \lambda_i v_i \right) + U$$

$$\sum_i \lambda_i v_i \in U$$

$$\sum_i \lambda_i v_i = \sum_j \mu_j u_j$$

since B_V is linearly independent, we have that all λ_i and μ_j are zero. Similiarly if $v_i + U = v_j + U$ with $i \neq j$ then we can write $v_i - v_j = \sum_j \mu_j u_j$ which is a contradiction. □

Remark. We can maek a direct proof without quotient spaces by rearranging some of the arguments of the proof.

Corollary. (Linear Pigeonhole principle) If $\dim V = \dim W = n$ and $\alpha : V \to W$ then the following conditions are equivalent.

(i) α is injective,
(ii) α is surjective,
(iii) α is an isomorphism.

Proof. If α injective then $\mathrm{n}(\alpha) = 0$ so by rank nullity we have that $\mathrm{rk}(\alpha) = n$ so α is surjective. If α is surjective then $\mathrm{rk}(\alpha) = n$ so by rank nullity, the dimension of the kernel is 0 hence the kernel is trivial, so α injective, hence α is an isomorphism. If α is an isomorphism, clearly it's injective, so all equivalent. □

Proposition. Suppose V is a vector space with a basis B. For any vector space W and any function $f : B \to W$ there is a unique linear map $F : V \to W$ such that $F(B) = W$.

Proof. First we'll show existance. For $v \in V$ write $v = \sum_b \lambda_b b$ for a finite sum. Then define

$$F(v) = \sum_b \lambda_b f(b).$$

This is well-defined, since B is a basis the λ_b are uniquely determined by v. For $u, v \in V$ and $\lambda \in \mathbb{F}$ we write

$$u = \sum_b \mu_b b, \quad \sum_b \lambda_b b.$$

Then

$$\begin{aligned} F(u + \lambda v) &= F(\sum_b (\mu_b + \lambda\lambda_b) f(b) \\ &= \sum_b \mu_b f(b) + \lambda \sum_b \lambda_b f(b) \\ &= F(u) + \lambda F(v). \end{aligned}$$

So F is linear. To show uniqueness $\overline{F} : V \to W$ is another linear map extending f then,

$$\overline{F}\left(\sum_b \lambda b b\right) = \sum_b \lambda_b \overline{F}(b)$$

which is the same as our definition for F hence they are the same function.

Corollary. For a vector space, V, with $\dim V = n$ with a basis $B = \{v_1, \ldots, v_n\}$ for V then there is a unique isomorphism

$$\begin{aligned} F_B : V &\to \mathbb{F}^n \\ \sum_{i=1}^n \lambda_i v_i &\to \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \end{aligned}$$

Proof. Let $E = \{e_1, \ldots, e_n\}$ be the standard basis for $\mathbb{F}^n$. Define

$$\begin{aligned} f : B &\to W \\ v_i &\to e_i \end{aligned}$$

and let F_B be the unique linear extension of f to V. We see that f defines a bijection from $B \to E$. Let $\bar{F}_B$ be the unique linear extension of $f^{-1} : E \to B$. Then $\bar{F}_B \cdot F_B$ is the composition of two linear maps, hence it's linear, moreover it is id_B. But also id_V is also a linear extension of id_B, by the proposition, they are the same map so $\bar{F}_B \cdot F_B = F_B \cdot \bar{F}_B = \mathrm{id}_B$. Hence F_B is bijective, so it is an isomorphism. □

Corollary. If V, W are finite dimensional $\mathbb{F}$-vector spaces. Then

$$V \cong W \iff \dim V = \dim W$$

Proof. Trivial from the corollary using the transitivity of the isomorphism relation. □

Definition. (Coordinate vector) $F_B(v) = [v]_B$ is the *coordinate vector* of v with respect to the basis B

For an example if $V \cong \mathbb{F}^n$ and $U \leq V$ with $U \cong \mathbb{F}^m$ then $\dim(V/U) = n - m$, so $\frac{V}{U} \cong \mathbb{F}^{n-m}$.

1.4 Direct sums

Definition. (External direct sum) For $\mathbb{F}$-vector spaces, V and W, we dnote the *external direct sum* of V and W as $V \oplus W$ with underlying set $V \times W$ with addition and scalar multiplication given in the obvious sense.

We can similarly define

$$V_1 \oplus \cdots \oplus V_n = \bigoplus_{i=1}^{n} V_i.$$

Lemma. For V, W finite dimensional vector spaces,

$$\dim(V \oplus W) = \dim V + \dim W$$

Proof.
(First Proof) Let B, C be basis for V, W respectively. Set

$$D = (B \times \{\mathbf{0}_W\}) \cup (\{\mathbf{0}_V\} \times C)$$

it is straightfoward to check that D is basis of $V \oplus W$ of the size $\dim V + \dim W$. □

(Second Proof) Suppose $V \cong \mathbb{F}^n$ and $W \cong \mathbb{F}^m$ construct an isomorphism $V \oplus W \cong \mathbb{F}^{n+m}$. □

Proposition. Let V be a vector space with $U, W \leq V$. There is a surjective linear map

$$\begin{aligned} \varphi : U \oplus W &\to U + W \\ (u, w) &\to u + w \end{aligned}$$

with $\ker \varphi \cong U \cap W$.

Proof. Surjectively and linearity are clear. Note for $(u, w) \in U \oplus W$ then $(u, w) \in \ker \varphi$ if and only if $w = -u$. Hence

$$\ker \varphi = \{(x, -x) : x \in U \cap W\}$$

the map $\psi : U \cap W \to \ker \varphi$ sending $x \to (x, -x)$ is an isomorphism.

Corollary. (Sum-Intersection Formula) If V is finite dimensional and $U, W \leq V$ then

$$\dim(U + W) = \dim U + \dim V - \dim(U \cup V)$$

Applying the rank-nullity theorem to the linear map φ in the proposition we get that

$$\begin{aligned} \dim U + \dim W &= \dim(U \oplus V) \\ &= \dim(\ker \varphi) + \dim(\operatorname{im} \varphi) \\ &= \dim(U + W) + \dim(U \cap W) \quad \square \end{aligned}$$

We can also give an explicit basis. Given a basis B for $U \cap W$, extend B to a basis B_U for U, and a basis B_W for W. Then $B_U \cap B_W$ spans $U + W$ and

$$|B_U \cup B_W| \leq |B_U| + |B_W| - |B| = \dim(U + V)$$

hence $B_U \cup B_W$ is linearly independent so it's a basis for $U + W$.

Remark. We could also check directly that $B_U \cup B_W$ is linearly independent of the size $\dim(U+V)$ without assuming the sum-intersection formula, so this alos servers as an alternative proof of the sum-intersection formula.

Definition. (Internal direct sum) Suppose $U, W \leq V$ satisify
(i) $U + W = V$,
(ii) $U \cap W = \{\mathbf{0}_V\}$.
Then

$$\varphi : U \oplus W \to V$$

is an isomorphism, and we say that V is the *internal direct sum* of U and W, and we write that $V = U \oplus W$.

Alternatively, every element $v \in V$ can be written *uniquely* as $v = u + w$ for $u \in U, w \in W$.

Definition. (Direct complement) For $U \leq V$ a *direct complement* to U in V is a subspace $W \leq V$ satisfying $V = U \oplus W$.

Proposition. If V is finite dimensional then every subspace has a direct complement.

Proof. Let $U \leq V$ and let B_U be a basis for U. Extend to a basis B_V for V. Set $W =_V \setminus B_U\rangle$. Then

$$\begin{aligned} V = \langle B_V \rangle &= \langle B_U \cup (B_V \setminus B_U) \rangle \\ &= \langle B_U \rangle + \langle B_V \setminus B_U \rangle \\ &= U + W. \end{aligned}$$

Moreover using the sum-intersection formula

$$\dim(U \cap W) = |B_V| + |B_U| - |B_V \setminus B_U| = 0.$$

Hence $U \oplus W = V$. □

More generally for $U_1, \ldots, U_n \leq V$ we say that V is the direct sum of the U_i and write that

$$V = U_1 \oplus + \cdots + \oplus V_n = \bigoplus_{i=1}^{n} V_i$$

if the map

$$\begin{aligned} \varphi : U_1 \oplus \cdots \oplus U_n &\to V \\ (u_1, \ldots, u_n) &\to u_1, \ldots, u_n \end{aligned}$$

is an isomorphism. Equivalently every $v \in V$ can be uniquely written as $v = u_1 + \cdots + u_n$ for $u_i \in U_i$.

2 Matrices and Linear Maps

2.1 Vector spaces of linear maps

Definition. For V, W $\mathbb{F}$-vector spaces we define

$$\mathcal{L}(V, W) = \{\alpha : V \to W : \alpha \text{ is linear}\}$$

which forms a $\mathbb{F}$-vector space under pointwise addition and obvious scalar multiplication.

Recall that $M_{m \times n}$ is the space of matrices over $\mathbb{F}$ with m rows and n columns. For $A \in M_{m \times n}(\mathbb{F})$ we write $A = (a_{ij})$ where $a_{ij} \in \mathbb{F}$ is the entry in the ith row and the jth column.

Let $B = \{v_1, \ldots, v_n\}, C = \{w_1, \ldots, w_m\}$ are *ordered* basis for V, W.

Let $\alpha \in \mathcal{L}(V, W)$. We can write

$$\begin{aligned} \alpha(v_1) &= a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m \\ \alpha(v_2) &= a_{12}w_1 + a_{22}w_2 + \cdots + a_{m1}w_m \\ &\vdots \\ \alpha(v_n) &= a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m \end{aligned}$$

Definition. (Matrix) The *matrix* of α with respect to the ordered basis B, C is

$$[\alpha]_C^B = (a_{ij}) \in M_{m\times n}(\mathbb{F})$$

Recall we have a linear isomorphism

$$\varepsilon_B : V \to \mathbb{F}^n$$
$$v = \sum_{i=1}^{n} \lambda_i v_i \to (\lambda_i)_i = [v]_B$$

where $[v]_B$ is the coordinate vector of v with respect to B.

Proof. Let $v \in V$ write $v = \sum_{j=1}^{n} \lambda_j v_j$. Then

$$\begin{aligned}\alpha(v) &= \sum_{j=1}^{n} \lambda_j \alpha(v_j) \\ &= \sum_{j=1}^{n} \lambda_j \sum_{i=1}^{m} a_{ij} w_i \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \lambda_j a_{ij} \right) w_i.\end{aligned}$$

So

$$\begin{aligned}[\alpha(v)]_C &= \left(\sum_{j=1}^{n} a_{ij}\lambda_j \right)_i \\ &= (a_{ij}) \cdot (\lambda_j) \\ &= [\alpha]_C^B [v]_B.\end{aligned}$$

Hence (i) is proved. For (ii), take $1 \le j \le n$, so $[v_j]_B = e_j$. Hence for $A \in M_{m\times n}(\mathbb{F})$, $A[v_j]_B$ is the jth column of A. But if $A[v_j]_B = [\alpha(v_j)]_C = [\alpha]_C^B [v_j]_B = [\alpha]_C^B e_j$, then $A[v_j]_B$ is also the jth column of $[\alpha]_C^B$. Since this holds for all j in our range, they are the same matrix.

Now for part (iii), let $\alpha, \beta \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$. Then

$$\begin{aligned}[\alpha + \lambda\beta]_C^B [v]_B &= [(\alpha + \lambda\beta)(v)]_C \\ &= [\alpha(v) + \lambda\beta(v)]_C \\ &= [\alpha(v)]_C + \lambda[\beta(v)]_C \\ &= ([\alpha]_C^B + \lambda[\beta]_C^B)[v]_B\end{aligned}$$

for all $v \in V$. Hence by (ii) we get that $[\alpha + \lambda\beta]_c^b = [\alpha]_C^B + \lambda[\beta]_C^B$ so the map is linear. Let $\alpha \in \ker(\varepsilon_C^B)$ so that $[\alpha]_C^B = 0 \in M_{m\times n}(\mathbb{F})$. Then by (i) we have that $[\alpha(v)]_C = 0$ for all $v \in V$. But $\varepsilon : w \to [w]_C$ is an isomorphism so $\alpha(v) = 0$ for all $v \in V$ hence $\alpha = 0$ and α is injective. For surjectivity let $A \in M_{m\times n}(\mathbb{F})$ and define $f : B \to W$ by $f(v_j) = \sum_{i=1}^{n} a_{ij} w_I$ and extend f to a linear map $F : V \to W$. Then $[F]_C^B = A$. So ε_C^B is an isomorphism. □

Proposition. Let V, W, X be finite-dimensional $\mathbb{F}$-vector spaces with basis B, C, D and $\alpha \in \mathcal{L}(V, W)$ and $\beta \in \mathcal{L}(W, X)$. Then

$$[\beta \circ \alpha]_D^B = [\beta]_D^C [\alpha]_C^B.$$

Proof. By the theorem $[\beta \circ \alpha]_D^B$ is the unique matrix A satisfying

$$A[v]_B = [\beta(\alpha(v))]_D, \quad \forall v \in V.$$

But $[\beta]_D^C [\alpha]_C^B [v]_B = [\beta]_D^C [\alpha(v)]_C = [\beta(\alpha(v))]_D$. So by (ii) of theorem they are equal. □

Remark. For any basis B of V,

$$[\mathrm{id}_V]_B^B = I_{\dim V}.$$

Definition. (Change of basis matrix) Let B, B' be basis for V and $\dim V = n$. The *change of basis matrix* from B to B' is given by

$$P = [\mathrm{id}_V]_{B'}^B \in M_{m \times n}(\mathbb{F})$$

Equivalently letting $B = \{v_i\}_{i=1}^n$ and $B' = \{v_i'\}_{i=1}^n$, then

$$P = (p_{ij}) \quad \text{where} \quad v_j = \sum_{i=1}^n p_{ij} v_i'$$

so the jth column of P is $[v_j]_{B'}$.

Proposition. For V, W finite-dimensional vector spaces,

(i) $[\mathrm{id}_V]_{B'}^B \in GL_n(\mathbb{F})$ with inverse $[\mathrm{id}_V]_B^{B'}$.

(ii) If $\alpha \in \mathcal{L}(V, W)$ and B, B' basis for V and C, C' basis for W, then

$$[\alpha]_{C'}^{B'} = [\mathrm{id}_W]_{C'}^C [\alpha]_C^B [\mathrm{id}_V]_B^{B'}.$$

Proof. By the remark,

$$I_n = [\mathrm{id}_V]_B^B = [\mathrm{id}_V]_B^{B'} [\mathrm{id}_V]_{B'}^B$$

and symmetrically swapping B and B'. For the second part the result is immediate from the proposition.

Definition. (Equivalent matrices) Let $A, A' \in M_{m \times n}(\mathbb{F})$. We say that A and A' are *equivalent* if $\exists P \in GL_m(\mathbb{F})$, $Q \in GL_n(\mathbb{F})$ such that $A' = PAQ$.

Remark. Certianly A is equivalent to itself by $P = I_m$ and $Q = I_n$.
If $A' = PAQ$ then $A = P^{-1} A' Q^{-1}$.
If $A'' = RA'S$ too, then $A'' = (RP)A(QS)$, so the equivalence of matrices is an equivance relation on $M_{m \times n}(\mathbb{F})$.

Theorem. Let V, W be finite-dimensional $\mathbb{F}$-vector spaces. Let $\dim V = n$, $\dim W = m$ and let $\alpha \in \mathcal{L}(V, W)$. Let $r = \mathrm{rk}(\alpha)$. Then,

(i) There exists basis B, C for V, W respectively such that

$$[\alpha]_C^B = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \in M_{m\times n}(\mathbb{F})$$

where I_r is the identity matrix of size r, and the zeros are block zero matrices.

(ii) If

$$[\alpha]_{C'}^{B'} = \begin{pmatrix} I_{r'} & 0 \\ 0 & 0 \end{pmatrix} \in M_{m\times n}(\mathbb{F})$$

for some basis B', C' of V, W respectively, then $r' = r$

Proof. By rank-nullity $\mathrm{n}(\alpha) = n - r$. Let $\{v_{r+1}, \ldots, v_n\}$ be a basis for $\ker \alpha$. Extend to a basis $B = \{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$. Then $\{\alpha(v_1), \ldots, \alpha(v_r)\}$ spans the image, and has size at most $\dim(\mathrm{im}(\alpha))$, so it's linearly independent, hence we can extend it to form a basis of W.

$$C = \{w_1 = \alpha(v_1), \ldots, w_r = \alpha(v_r), w_{r+1}, \cdots, w_m\}$$

Then

$$\alpha(v_j) = \begin{cases} w_j & 1 \le j \le r \\ \mathbf{0} & \text{otherwise} \end{cases}$$

hence we have that $[\alpha]_C^B = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$.

For the second part, if $[\alpha]_{C'}^{B'} = \begin{pmatrix} I_{r'} & 0 \\ 0 & 0 \end{pmatrix}$ then

$$\alpha(v_j') = \begin{cases} w_j' & 1 \le j \le r' \\ \mathbf{0} & \text{otherwise} \end{cases}.$$

Hence $w_1', \ldots, w_{r'}'$ span $\mathrm{im}(\alpha)$ and are linearly independent. Hence $\mathrm{rk}(\alpha) = r'$. □

Definition. (Column-space) For $A \in M_{m\times n}(\mathbb{F})$ the *column-space* $\mathrm{Col}(A)$ is the subspace of $\mathbb{F}^m$ spanned by the columns of A. The dimension of the column-space is called the *column-rank* of A.

Definition. (Row-space) For $A \in M_{m\times n}(\mathbb{F})$ the *row-space* $\mathrm{Row}(A)$ is the subspace of $\mathbb{F}^m$ spanned by the rows of A (when transposed as column vectors). The dimension of the row-space is called the *row-rank* of A.

Remark.

$$\mathrm{Row}(A) = \mathrm{Col}(A^T)$$

hence the row-rank of A is the same as the column-rank of A^T.

Remark. Given a matrix $A \in M_{m\times n}(F)$ we can define a linear map $\alpha : \mathbb{F}^n \to \mathbb{F}^m$ by $\alpha(v) = Av$. Then $\text{im}(\alpha) = \text{Col}(A)$, so the rank of α is the same as the column-rank of A. Moreover, $A = [\alpha]_{E_m}^{E_n}$ where E_k are the standard basis for $\mathbb{F}^k$.

We may write $\text{im}\, A, \ker A, \text{rk}(A), \text{n}(A)$ to refer to the corresponding concepts for α.

Theorem. Let $A, A' \in M_{m\times n}(\mathbb{F})$, then
(i) A is equivalent to

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \text{ where } r \text{ is the column-rank of } A$$

(ii) A and A' are equivalent if and only if the have the same column-rank.

Proof. We'll first prove a lemma.

Lemma. For $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times p}(\mathbb{F})$ then $\text{rk}(A \cdot B) \leq \min(\text{rk}(A), \text{rk}(B))$.

Proof. We have that $\text{im}(AB) \leq \text{im}(A)$ so $\text{rk}(AB) \leq \text{rk}(A)$. If $Bv = \mathbf{0}$ for $v \in \mathbb{F}^p$, then $ABv = \mathbf{0}$, so $\text{n}(B) \geq \text{n}(AB)$, so applying rank-nullity, we get that

$$p - \text{rk}(B) \leq p - \text{rk}(AB) \implies \text{rk}(AB) \leq \text{rk}(B) \quad \square$$

Now we'll prove the first part of the theorem. Let α the natural linear map corresponding to A, so $A = [\alpha]_{E_m}^{E_n}$. By the previous theorem, there exists matrices B, C of $\mathbb{F}^n, \mathbb{F}^m$ such that

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} = [\alpha]_C^B = [\text{id}_{\mathbb{F}^m}]_C^{E_m} [\alpha]_{E_m}^{E_n} [\text{id}_{\mathbb{F}^n}]_{E_n}^B = PAQ$$

where $r = \text{rk}(\alpha)$ which we know is equal to the column-rank of A.

If A' has column-rank r then both matrices are equivalent to $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$, so by transitivity, A and A' are equivalent. Conversely suppose that A and A' are equivalent, so $A' = PAQ$. By the lemma $\text{rk}(A') \geq \text{rk}(AQ) \geq \text{rk}(A)$ and symmetrically we get that $\text{rk}(A) \geq \text{rk}(A')$, hence $\text{rk}(A') = \text{rk}(A)$. $\square$

Theorem. For any $A \in M_{m\times n}(\mathbb{F})$, the row-rank of A is equal to the column-rank of A.

Proof. Note that if P is invertible, then so it the tranpose with inverse $(P^{-1})^T$. Let r be the column-rank of A. So there exists matrices $P \in GL_m(\mathbb{F})$ and $Q \in GL_n(\mathbb{F})$ such that $PAQ = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \in M_{m\times n}(\mathbb{F})$. Then A^T is equivalent to $Q^T A^T P^T = (PAQ)^T = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \in M_{n\times m}(\mathbb{F})$. By the previous theorem, the column-rank of A^T is r which also the row-rank of A. $\square$

Let V be a finite-dimensional vector space and B, B' be basis for V. Now let $\alpha \in \text{End}(V) = \mathcal{L}(V, V)$. Then

$$[\alpha]_{B'}^{B'} = [\text{id}_V]_{B'}^B [\alpha]_B^B [\text{id}_V]_B^{B'}$$

Definition. (Similarity) For matrices $A, A' \in M_{n\times m}(\mathbb{F})$ are *similar* if there exists $P \in GL_n(\mathbb{F})$ such that $A' = P^{-1}AP$.

Remark. We have some remarks showing the similarity and equivalence are not the same thing.

(i) Similarity is an equivalence relation on $M_{n\times n}(\mathbb{F})$.

(ii) Similar matrices are equivalent but equivalent matrices need not be similar.

For example every matrix in $GL_n(\mathbb{F})$ is equivalent to I_n but I_n forms its only single element equivalence class, when we think about similarity.

2.2 Elementary operations on matrices

Definition. (Elementary row operations) Let $r_1, \ldots, r_m$ be the rows of A. We have three types of *elementary row operations* on A

(i) Swap r_i and r_j with $i \neq j$.
(ii) Replace r_i with λr_i with $0 \neq \lambda \in \mathbb{F}$.
(iii) Replace r_i with $r_i + \lambda r_j$ with $\lambda \in \mathbb{F}$ and $i \neq j$.

Similarly there are three types of elementary column operations.

Remark. These are all reversable.

Each elementary operation has a corresponding matrix representation repesentation. All corresponding matrices are invertible.

Lemma. If E is a matrix of type (i)-(iii) then EA is obtained from A by applying the corresponding ERO to A.

Proof. Direct matrix computation.

Remark. Similarly AE is obtained by applying the corresponding ECO

Remark. EROs preserve $\text{Row}(A)$ (and ECOs preserve $\text{Col}(A)$).

So both EROs and ECOs preserve the row-rank of a matrix, and therefore also the rank of the linear map corresponding to the matrix.

Definition. (Row reduced echelon form) A matrix $A \in M_{m\times n}(\mathbb{F})$ is said to be in *row reduced echelon form* (RRE) if

(i) All non-zero rows of A appear above all zero rows.
(ii) The leftmost non-zero element of a non-zero row is 1 (called the *pivot entry*).
(iii) If row r_i, r_j are non-zero rows with $i < j$ then the index of the pivot entry of i is less than the index of the pivot entry of j.
(iv) In a column containing a pivot entry, every other entry is zero.

For an example consider

$$M = \begin{pmatrix} 1 & a & 0 & 0 & b \\ 0 & 0 & 1 & 0 & c \\ 0 & 0 & 0 & 1 & d \end{pmatrix}$$

which is in row reduced echelon form. Similarly we have column reduced echelon form, which have the exact same rules but transposed.

Lemma. If A is in row reduced echelon form then the row rank of A is the number of non-zero rows of A.

Proof. Let $r_1, \dots, r_k$ be the non-zero rows of A let $j_i = P(v_i)$ be the pivot entry. Certainly $r_1, \dots, r_k$ span $\mathrm{Row}(A)$. Suppose that

$$v = \sum_{i=1}^{k} \lambda_i r_i = 0 \quad (\lambda_i \in \mathbb{F}).$$

Then $(v)_{j_i} = \lambda_i = 0$ so the non-zero rows are linearly independent so we're done. □

Proposition. Every matrix $A \in M_{m\times n}(\mathbb{F})$ can be put into row reduced echelon form with elementary row operations.

Proof. Proceed by induction on n. Write that $A = [c_1 \mid \cdots \mid c_n]$. If $c_1 = 0$ apply induction to $[c_2 \mid \cdots \mid c_n]$, so suppose that $c_1 \neq 0$, suppose that element in $(i, 1)$ is non-zero. Applying row operations (i) we can move it to $(1, 1)$. Apply row operation (ii) to rescale it to be 1. Now we can clear the rest of the column by (iii). By induction we can use elementary row operation on rows 2-m to reduce further. This is decreasing the dimension to the process terminates, hence the matrix can be put into row reduced echelon form. □

Remark. Putting a matrix into RRE form preserves the row-space and the RRE of any matrix is unique. Also if A is a square matrix then A either has a zero row or is the identity.

Theorem. For $A \in M_{m\times n}(\mathbb{F})$ the following are equivalent:
(i) $\mathrm{rk}(A) = n$.
(ii) A is a product of elementary matrices.
(iii) A is invertible.

Proof. Let's prove that (i) $\implies$ (ii). By the proposition there exists elementary matrices E_i such that $E_1 \dots E_\ell A$ is in RRE form. By the remark this is I_n hence $A = E_\ell^{-1} \cdots E_1^{-1}$ which are also elementary. For (ii) $\implies$ (iii) elementary matrix lie in $GL_n(\mathbb{F})$ which is a group, hence closed. Finially for (iii) $\implies$ (i) suppose there exists $B \in M_{m\times n}(\mathbb{F})$ such that $AB = I_n$. Then for $v \in \mathbb{F}^n$ we have that $v = (AB)v = A(Bv)$, so $v \in \mathrm{im}\, A$.

3 Determinant and Traces

3.1 Determinant

Theorem. There exists a unique function $F : M_{m\times n} \to \mathbb{F}$ satisfying
(i) (Alternating) If $c_i = c_j$ for some $i \neq j$ then $F(A) = 0$.
(ii) (Multilinear in columns) For all $1 \leq i \leq n$ and $v_j \in \mathbb{F}^n$ the function

$$\begin{aligned}\mathbb{F}^n &\to \mathbb{F}\\ v &\to F(v_1 \mid \cdots \mid v_{j-1} \mid v \mid v_{j+1} \mid \cdots \mid v_n)\end{aligned}$$

is linear.
(iii) $F(I_n) = 1$.

Definition. (Determinant) We shall defined the F in the previous theorem as the *n-dimensional determinant*, written as $F(A) = \det(A)$. A function satisfying conditions (i) and (ii) of the theorem is called an *n--dimensional volumn form.*

Lemma. If F is an n-dimensional volumn form, $A \in M_{m\times n}(\mathbb{F})$,
(i) If A has a zero column then $F(A) = 0$,
(ii) $F(AT_{ij}) = -F(A)$,
(iii) $F(AM_{i,\lambda}) = \lambda F(A)$,
(iv) $F(AC_{i,j,\lambda}) = F(A)$.

Proof. Let $f_i : \mathbb{F}^n \to \mathbb{F}$ be given by $v \to F(c_1 \mid \cdots c_{i-1} \mid v \mid c_{i+1} \mid \cdots \mid c_n)$. So that f_i is linear. Then $f_i(c_j) = \delta_{ij}F(A)$. iF $c_i = 0$ then $F(A) = f_i(c) = f_i(0) = 0$. For (ii), let $\bar{A}$ be the matrix obtained from A by replacing both ith and jth columns of A by $c_i + c_j$. Then $0 = F(\bar{A}) = F(A) + f_i(c_j) + f_i(c_i) + F(AT_{ij})$. For (iii), $F(AM_{i,\lambda}) = f_i(\lambda c_i) = \lambda f_i(c_i) = \lambda F(A)$. Now for (iv), $F(AC_{i,j,\lambda}) = f_j(c_j + \lambda c_i) = f_i(c_j) + \lambda f_j(c_i) = F(A)$. □

Now we're ready to prove the theorem.
Proof. First we'll prove uniqueness. Let F be an n-dimensional volumn form with $F(I_n) = 1$. By the lemma, $F(T_{i,j}) = -1, F(M_{i,\lambda}) = \lambda, F(C_{i,j,\lambda}) = 1, F(AE) = F(A)F(E)$ for E elementary. Let $A \in M_{n\times n}(\mathbb{F})$, so there exists elementary matrices $E_1, \ldots, E_\ell$ such that $A' = AE_1 \cdots E_\ell$ with A' in CRE form. Then $F(A) = F(A')F(E_1)^{-1} \cdots F(E_\ell)^{-1}$, so either $A' = I_n$ so $F(A) = (F(E_1)^{-1} \cdots F(E_\ell)^{-1}$ or A' has a zero colun so by the lemma, $F(A) = F(A') = 0$.

This also proves the corollary.

Corollary. $\det A \neq 0$ if and only if A is invertible. In this case, $A = E_1 \cdots E_\ell$ then $\det A = \det(E_1) \cdots \det(E_\ell)$.

Recall that from IA Groups that sgn : $S_n \to \{\pm 1\}$ is the unique homomorphism satisfying

$\operatorname{sgn}(\tau) = -1$ for all transpositions. Now we can define the determinant.

$$\begin{aligned} \det : M_{n\times n}(\mathbb{F}) &\to \mathbb{F} \\ a &\to \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma(i)i} \end{aligned}$$

Clearly we have that $\det I_n = 1$. Each product $\prod_{i=1}^{n} a_{\sigma(i)i}$ is multilinear in columns. Hence so is $\det A$. Suppose that $c_k = c_\ell$ for $k \neq \ell$. Set $\tau = (k\ \ell)$ so that $a_{ij} = a_{i\tau(j)}$ for all i, j. Then

$$\begin{aligned} \det A &= \sum_{\sigma\in A_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma(i)i} + \sum_{\sigma\in A_n} \operatorname{sgn}(\sigma\tau) \prod_{i=1}^{n} a_{\sigma\tau(i)i} \\ &= \sum_{\sigma\in A_n} \prod_{i=1}^{n} a_{\sigma(i)i} - \sum_{\sigma\in A_n} \prod_{i=1}^{n} a_{\sigma\tau(i)i} \\ &= \sum_{\sigma} \prod_{i} a_{\sigma(i)i} - \sum_{\sigma} \prod_{i} a_{\sigma\tau(i)\tau(i)} \\ &= \sum_{\sigma} \prod_{i} a_{\sigma(i)i} - \sum_{\sigma} \prod_{j} a_{\sigma(j)j} \end{aligned}$$

□

Now we will observe some properties of the determinant.

Lemma. For $A \in M_{n\times n}(\mathbb{F})$, we have that $\det\left(A^T\right) = \det(A)$.

Proof.

$$\begin{aligned} \det\left(A^T\right) &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)} \\ &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{j} a_{\sigma(j)^{-1}j} \\ &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma^{-1}) \prod_{j} a_{\sigma(j)^{-1}j} \end{aligned}$$ □

Proposition. For all $A, B \in M_{n\times n}(\mathbb{F})$ we have that $\det(AB) = \det(A)\det(B)$.

Proof. Recall that $\operatorname{rk}(AB) \leq \min(\operatorname{rk}(A), \operatorname{rk}(b))$, so if either A or B has rank less than n then so does AB so by the corollary above $\det(AB) = 0 = \det(A)\det(B)$. If not, A and B are invertible and can be written as a product of elementary matrices, so $A = E_1 \cdots E_\ell$, $B + E_1' \cdots E_k'$. So $AB = E_1 \cdots E_\ell E_1' \cdots E_k'$, hence $\det(AB) = \det(E_1) \cdots \det(E_\ell)\det(E_1') \cdots \det(E_k') = \det(A)\det(B)$. □

Missed a lecture - 31.11.25

Definition. (Trace) For a $A \in M_{m\times n}(\mathbb{F})$ the *trace* of A is given by

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{i,i}$$

Remark. $\operatorname{tr} \in \mathcal{L}(M_{m\times n}(\mathbb{F}), \mathbb{F})$

Lemma. For all $A, B \in M_{m\times n}(\mathbb{F})$ we have that

$$\operatorname{tr}(AB) = \operatorname{tr}(BA)$$

Proof.

$$\begin{aligned}
\operatorname{tr}(AB) &= \sum_{i=1}^{n} (AB)_{i,i} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} \\
&= \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ji}a_{ij} \\
&= \sum_{j=1}^{n} (BA)_{j,j} = \operatorname{tr}(BA) \quad \square
\end{aligned}$$

But in general we don't have that $\operatorname{tr}(AB) = \operatorname{tr}(A)\operatorname{tr}(B)$.

Corollary. Similar matrices have the same trace.

Proof. For $P \in GL_n(\mathbb{F})$,

$$\operatorname{tr}\big(PAP^{-1}\big) = \operatorname{tr}\big(P^{-1}(PA)\big) = \operatorname{tr}(A)$$

Definition. For V a finite dimensional vector space and $\alpha \in \mathcal{L}(V, V)$ define the *trace* of α by

$$\operatorname{tr}(\alpha) = \operatorname{tr}\big([\alpha]_B^B\big)$$

for B a basis of V.

Proposition. This is independent of the basis B.

Proof. If B' is another basis of the vector space V then, $[\alpha]_{B'}^{B'}$ and $[\alpha]_B^B$ are similar matrices, so the result follows from the corollary.

4 Dual Spaces

Definition. (Dual space) If V is a $\mathbb{F}$-vector space, then the *dual space* of V is

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

For example we have that $\theta : \mathbb{R}^3 \to \mathbb{R}$ given by

$$\theta\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = x - 2y + 3z$$

is an element in the dual space of $\mathbb{R}^3$.

$\mathrm{tr} \in M_{n\times n}(\mathbb{F})^*$

If $V = C[0,1]$ then $\theta : V \to R$ given by $\theta(f) = \int_0^1 f(t)e^{-t}\mathrm{d}t$ is in the dual space of V^*.

An element $\theta \in V^*$ is called a *linear functional* on V.

Suppose that B is a basis for V. For $b \in B$ define $b^* \in V^*$ by

$$b^*\left(\sum_{c\in B} \lambda_c c\right) = \lambda_b$$

i.e. $b^*(c) = \delta_{bc}$. If we let $B^* = \{b^* : b \in B\}$ then:

Proposition. For B^* defined above,
(i) B^* is linearly independent;
(ii) If V is finite dimensional then B^* is a basis for V^*.

Definition. (Dual basis) If V is finite dimensional, call B^* the *dual basis* to B.

Proof. Suppose that

$$\sum_{b\in B} \lambda_b b^* = 0 \quad \text{in } V^*.$$

Then for $c \in B$,

$$0 = \left(\sum_{b\in B} \lambda_b b^*\right)(c) = \sum_{b\in B} \lambda_b b^*(c) = \lambda_c$$

so $\lambda_c = 0$ hence all coefficients are zero, so the set B^* is linearly independent.
For V, W finite dimensional we know that $\dim(\mathcal{L}(V,W)) = \dim(V)\dim(W)$ so for $\dim(V^*) = \dim(\mathcal{L}(V,\mathbb{F})) = \dim V$. We know that B^* is a linearly independent subset of V of size $\dim V$ hence it is a basis of V^*. □

We can also offer a constructive proof of (ii)

Given $\theta \in V^*$ and $b \in B$ set $\lambda_b = \theta(b) \in \mathbb{F}$ and let $\theta = \sum_{b\in B} \lambda_b b^* \in V^*$. Then $\bar{\theta} \in \langle B^* \rangle$ and for $c \in B$ we have that

$$\bar{\theta}(c) = \lambda_c = \theta(c)$$

so $\theta = \bar{\theta}$ as they agree on a basis and hence we have that $\theta \in \langle B^* \rangle$ so B^* spans V^*.

However there are vector spaces, not finite dimensional such as $P(\mathbb{R})^* = \mathbb{R}^{\mathbb{N}}$.

Corollary. For V finite dimensional $V \cong V^*$

Proof. Same dimension hence isomorphic. Note that if V is finite dimensional, $B = \{v_1, \ldots, v_n\}$ is a basis of V, for $v \in V$ and $\theta \in V^*$ we can write,

$$v = \sum_{i=1}^{n} \lambda_i v_i, \quad \theta = \sum_{j=1}^{n} \mu_j v_j^*.$$

Then $\theta(v) = \sum_{i,j} \lambda_i \mu_j v_j^*(v_i) = \sum_{i=1}^{n} \lambda_i \mu_i = ([\theta]_{B*})^T \cdot [v]_B$

Definition. (Annihilator) For V a finite dimensional $\mathbb{F}$-vector space and $S \subseteq V$, the *annihilator* of S is

$$S^0 = \{\theta \in V^* : \forall s \in S, \theta(s) = 0\} \subseteq V^*$$

Lemma. For and $S, T \subseteq V$,
(i) $S^0 \leq V^*$;
(ii) If $S \subseteq T$ then $T^0 \leq S^0$;
(iii) $S^0 = \langle S \rangle^0$;
(iv) $V^0 = \{\mathbf{0}_{V^*}\}$ and $\{\mathbf{0}_V\}^0 = V^*$.

Proof. A simple application of the subspace test proves (i). For (ii) is suffices to check that $T^0 \subseteq S^0$. For $\theta \in T^0, s \in S$ so $s \in T$ hence $\theta(s) = 0$. For (iii), $S \subseteq \langle S \rangle$ so by (ii)

$$\langle S \rangle^0 \leq S^0.$$

For the converse let $\theta \in S^0$ and $v \in \langle S \rangle$. So we can write $v = \sum_{s \in S} \lambda_s \cdot s$, so $\theta(v) = \sum_{s \in S} {}_s\theta(s) = 0$, hence $\theta \in \langle S \rangle^0$ so $S^0 = \langle S \rangle^0$. If $\theta \in V^*$ and $\forall v \in V$ we have that $\theta(v) = 0$ then θ must be the zero function, so $V^0 = \{\mathbf{0}_{V^*}\}$. Secondly for $\theta \in V^*$ we have that $\theta(\mathbf{0}_V) = 0$ so $\{\mathbf{0}_V\}^0 = V^*$. □

Proposition. For V finite dimensional with $U \leq V$, we have that

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

Proof. Suppose that $\dim V = n$ and $\dim U = k$ and let $B_U = \{v_1, \ldots, v_k\}$ be a basis for U and extend to a bassi $B_V = \{v_1, \ldots, v_n\}$ for V. Then $B_v^* = \{v_1^*, \ldots, v_n^*\}$ is a basis for V^*. Suffices to prove the following claim.

Claim. $\{v_{k+1}^*, \ldots, v_n^*\}$ forms a basis for U^0

First we show that it's a subset of U^0. For $i \leq k$ and $j \geq k+1$,

$$v_j^*(v_i) = 0$$

so

$$v_j^* \in (B_U)^0 = \langle B_u \rangle^0 = U^0.$$

Linear indepedence is obvious since it's a subspace of B_V^*. Let's check it's spanning. Let $\theta \in U^0$, so write $\theta = \sum_{j=1}^n \lambda_j v_j^*$. Then for $i \leq k$, $v_i \in U$, so $0 = \theta(v_i) = \sum_j \lambda_j v_j^*(v_i) = \lambda_i$. Hence $\theta = \sum_{j=k+1}^n \lambda_j v_j^* \in \langle v_{k+1}^*, \ldots, v_n^* \rangle$. □

Remark. If $U, W \leq V$ which are such that $V = U \oplus W$ hence $U^0 \cong W^*$ is really what's going on behind the scenes.

Proposition. If V is a $\mathbb{F}$-vector space and $U, W \leq V$ then
(i) $U^0 \cap W^0 = (U + W)^0$;
(ii) $U^0 + W^0 \leq (U \cap W)^0$;
(iii) If V is finite dimensional then we have equality in (ii).

Proof. First we prove (i). For $\theta \in V^*$ we have that $\theta \in (U+W)^0 \iff \forall u \in U, w \in W$ we have that $\theta(u + w) = 0$. This is equivalent to $\forall u, \forall w, \theta(u) = \theta(w) = 0$ so $\theta \in U^0 \cap W^0$. So we have equality.
Now for (ii) we have that

$$U \cap W \leq U, W$$

so by the previous lemma we have that

$$U^0, W^0 \leq (U \cap W)^0$$
$$U^0 + W^0 \leq (U \cap W)^0$$

Now for the finial part if we let $n = \dim V$ then we have that

$$\dim(U^0 + W^0) = \dim(U^0) + \dim(W^0) - \dim(U^0 \cap W^0)$$

which using the fact that $\dim(U^0 \cap W^0) = \dim((U+W)^0)$ we get that

$$\begin{aligned} &= (n - \dim U) + (n - \dim W) - (n - \dim(U + W)) \\ &= n - \dim(U \cap W) \quad \text{(by the Sum-Intersection formula)} \\ &= \dim((U \cap W)^0) \quad \text{(by the proposition)} \end{aligned}$$

Definition. (Dual map) If $\alpha \in \mathcal{L}(V, W)$ then the *dual map* of α is $\alpha^* : W^* \to V^*$ given by

$$\alpha^*(\theta) = \theta \circ \alpha$$

Lemma. If $\alpha, \beta \in \mathcal{L}(V, W)$ and $\gamma \in \mathcal{L}(U, V)$ and $\lambda \in \mathbb{F}$ then:
(i) α^* is linear;
(ii) $(\alpha + \lambda\beta)^* = \alpha^* + \lambda\beta^*$;
(iii) $(\alpha \circ \gamma)^* = \gamma^* \alpha^*$;
(iv) If β is an isomorphism then so is β^* and $(\beta^*)^{-1} = (\beta^{-1})^*$.

Proof. Let $\beta, \eta \in W^*$ and $\mu \in \mathbb{F}$. Then for $v \in V$ we have that

$$\begin{aligned}\alpha^*(\theta + \mu\eta(v)) &= (\theta + \mu\eta)(\alpha(v)) \\ &= \theta(\alpha(v)) + \mu\eta(\alpha(v)) \\ &= \alpha^*(\theta)(v) + \mu\alpha^*(\eta)(v) \\ &= (\alpha^*(\theta) + \mu\alpha^*(\eta))(v)\end{aligned}$$

Which holds for all v so we must have that

$$\alpha^*(\theta + \mu\eta) = \alpha^*(\theta) + \mu\alpha^*(\eta)$$

For (ii) take $\theta \in W^*$ and $v \in V$ and consider,

$$\begin{aligned}(\alpha + \lambda\beta)^*(\theta)(v) &= \theta((\alpha + \lambda\beta)(v)) \\ &= \theta(\alpha(v) + \lambda\beta(v)) \\ &= \theta(\alpha(v)) + \lambda\theta(\beta(v)) \\ &= \alpha^*(\theta)(v) + \lambda\beta^*(\theta)(v) \\ &= (\alpha^*(\theta) + \lambda\beta^*(\theta))(v).\end{aligned}$$

This is true for all v so we have that

$$(\alpha + \lambda\beta)^*(\theta) = \alpha^*(\theta) + \lambda\beta^*(\theta) = (\alpha^* + \lambda\beta^*)(\theta).$$

Which now is true for θ so that

$$(\alpha + \lambda\beta)^* = \alpha^* + \lambda\beta^*.$$

For (iii) take $\theta \in W^*$, so

$$\begin{aligned}(\alpha \circ \gamma)^*(\theta) &= \theta \circ (\alpha \circ \gamma) \\ &= \alpha^*(\theta) \circ \gamma \\ &= \gamma^*(\alpha^*(\theta)) \\ &= (\gamma^* \circ \alpha^*)(\theta).\end{aligned}$$

And again these maps both agree at all values of θ, so they are the same map.

$$(\alpha \circ \gamma)^* = \gamma^* \circ \alpha^*$$

Lastly for (iv), note that for all $\theta \in V^*$ we have,

$$(\mathrm{id}_V)^*(\theta) = \theta \circ \mathrm{id}_V = \theta,$$

so $(\mathrm{id}_V)^* = \mathrm{id}_{V^*}$. Thus $\mathrm{id}_{V^*} = (\beta^{-1} \circ \beta)^* = \beta^* \circ (\beta^{-1})^*$ and symmetrically we have that $\mathrm{id}_{W^*} = (\beta^{-1})^* \circ \beta^*$. Hence β^* is an isomorphism with our required inverse. □

Proposition. Let V, W be finite dimensional vector spaces and $\alpha \in \mathcal{L}(V, W)$. Let B, C be basis for V, W Then

$$[\alpha^*]_{B^*}^{C^*} = \left([\alpha]_C^B\right)^T.$$

Proof. Let $n = \dim V$ and $m = \dim W$. Let $B = \{b_1, \ldots, b_n\}$ and $C = \{c_1, \ldots, c_m\}$. Let $A = [\alpha]_C^B$ so

$$\alpha(b_i) = \sum_{k=1}^{m} a_{ki} c_k.$$

Now let $A' = [\alpha^*]_{B^*}^{C^*}$, so

$$\alpha^*(c_j^*) = \sum_{\ell=1}^{n} a'_{\ell i} b_\ell^*.$$

Then

$$\begin{aligned} a'_{i,j} &= \alpha^*(c_j^*)(b_i) \\ &= c_j^*(\alpha(b_i) \\ &= c_j^* \left(\sum_k a_{kc} c_k \right) = a_{j,i}. \quad \square \end{aligned}$$

For an example let B, C be basis of V finite dimensional. The setting $P[\mathrm{id}_V]_C^B$ and $[\mathrm{id}_{V^*}]_{B^*}^{C^*} = P^T$ so we have that

$$[\mathrm{id}_{V^*}]_{C^*}^{B^*} = (P^T)^{-1}$$

Corollary. Let V, W and α be as in the previous proposition. Then

$$\mathrm{rk}(\alpha^*) = \mathrm{rk}(\alpha)$$

Proof. Set $A = [\alpha]_C^B$, and recall that the column rank of a matrix is the same as the rank of the corresponding linear transformation. Hence

$$\begin{aligned} \text{c-rk}(A) &= \text{r-rk}(A) \\ &= \text{c-rk}(A^T) \\ &= \mathrm{rk}(\alpha^*) \quad \square \end{aligned}$$

Proposition. Let V, W be $\mathbb{F}$-vector spaces. Then if $\alpha \in \mathcal{L}(V, W)$,
(i) $\ker(\alpha^*) = \mathrm{im}(\alpha)^0$;
(ii) $\mathrm{im}(\alpha^*) \leq (\ker(\alpha))^0$;
(iii) If V, W are finite dimensional, then we have equality in (ii).

Proof. For $\theta \in W^*$, then

$$\begin{aligned} \theta \in \ker(\alpha^*) &\iff \theta \circ \alpha = \mathbf{0} \text{ in } V^* \\ &\iff \forall v \in V,\ \theta(\alpha(v)) = 0 \\ &\iff \theta \in \mathrm{im}(\alpha)^0. \end{aligned}$$

And for (ii), let $\eta \in V^*$,

$$\begin{aligned} \eta \in \mathrm{im}(\alpha^*) &\implies \exists \theta \in W^* \text{ s.t. } \eta = \theta \circ \alpha \\ &\implies \forall v \in \ker(\alpha),\ \eta(v) = \theta(\alpha(v)) = \theta(\mathbf{0}) = 0 \\ &\implies \eta \in (\ker(\alpha))^0. \end{aligned}$$

Now if we assume that V, W are finite dimensional, by the corollary we have that

$$\begin{aligned}\operatorname{rk}(\alpha^*) = \operatorname{rk}(\alpha) &= \dim V - \dim(\ker \alpha) \\ &= \dim(\ker(\alpha)^0) \quad \square\end{aligned}$$

4.1 The double dual

We denote the *double dual* as $V^{**} = (V^*)^*$. If V is finite dimensional then $V^{**} \cong V$ so they have the same dimension. But we can construct a much nicer isomorphism.

Theorem. If V is a $\mathbb{F}$-vector space then there is a linear map $\mathcal{E} : V \to V^{**}$ given by

$$\mathcal{E}(v)(\theta) = \theta(v) \quad \text{for } v \in V, \theta \in V^*$$

where if V is finite dimensional then ε is an isomorphism.

Proof. First we prove linearity. Take $v, w \in V;\ \lambda \in \mathbb{F};\ \theta \in V^*$. Then

$$\begin{aligned}\mathcal{E}(v + \lambda w)(\theta) = \theta(v + \lambda w) &= \theta(v) + \lambda\theta(w) \\ &= \mathcal{E}(v)(\theta) + \lambda\mathcal{E}(w)(\theta) \\ &= (\mathcal{E}(v) + \lambda\mathcal{E}(w))(\theta)\end{aligned}$$

This is true for all θ so $\mathcal{E}(v + \lambda w) = \mathcal{E}(v) + \lambda\mathcal{E}(w)$.

Now suppose that V is finite dimensional. Now we prove injectivity. Take $\mathbf{0} \neq v \in V$ with $\mathcal{E}(v) = \mathbf{0}$. i.e. $\forall\theta \in V^*,\ \mathcal{E}(v)(\theta) = 0$. Extend to a basis $\{v_1 = v, v_2, \ldots, v_n\}$ for V. Let $B^* = \{v_1^*, \ldots, v_n^*\}$ be a dual basis. Set $\theta = v_1^*$, then $0 = \theta(v) = v_1^*(v_1) = 1$, which is a contradiction. Surjectivity follows from the linear pigeonhole principle.

Definition. We call $\mathcal{E} : V \to V^{**}$ the *evaluation map* or the *natural isomorphism*.

Remark. For V finite dimensional we also have that $V \cong V^*$, but any isomorphism requires a change of basis.

Note that if we let V, W be finite dimensional with basis B, C respectively, and let $\alpha \in \mathcal{L}(V, W)$.

$$\begin{aligned}\alpha^*(\theta)(v) &= ([\alpha^*(\theta)]_{B^*})^* \cdot [v]_B \\ &= (A^T[\theta]_{C^*})^T \cdot [v]_B \quad \text{where } A = [\alpha]_C^B \\ &= ([\theta]_{C*})^T(A[v]_B) \\ &= ([\theta]_{C^*})^T[\alpha(v)]_C\end{aligned}$$

For the rest of the chapter we assume that V is a finite dimensional $\mathbb{F}$-vector space

Proposition. Every basis C for V^* is the dual basis to some basis of V.

Proof. Let $C = \{\theta_1, \ldots, \theta_n\}$ and let $C^* = \{\theta_1^*, \ldots, \theta_n^*\} \subseteq V^{**}$ be the dual basis. Let $\mathcal{E} : V \to V^{**}$ be the natural isomorphism. Set $v_i = \mathcal{E}^{-1}(\theta_i^*) \in V$. Then $B = \{v_1, \ldots, v_n\}$ is a basis of V since

the image of the basis C^* is a basis under an isomorphism. Then for all i, j

$$\begin{aligned}\theta_i(v_j) = \mathcal{E}(v_j)(\theta_i) &= \theta_j^*(\theta_i)\\ &= \delta_{ij}.\end{aligned}$$

Hence C is the dual basis to B. □

Proposition. For $U \leq V$,

$$\mathcal{E}(U) = (U^0)^0 \leq V^{**}$$

Proof. For $u \in U$ and $\theta \in U^0$

$$\mathcal{E}(u)(\theta) = \theta(u) = 0$$

True for all such θ, hence $\mathcal{E}(U) \leq (U^0)^0$. But $\dim((U^0)^0) = \dim V^* - \dim U^0 = \dim(V) - (\dim V - \dim U) = \dim(\mathcal{E}(U))$ hence we have that $\mathcal{E}(u) = (U^0)^0$. □

Remark. It is common to identify V with V^{**} under $\mathcal{E}$. Under this identification for $U \leq V$, $U = U^{00}$.
For $X \subset V^*$, $X^0 \leq V^{**}$ is identified with

$$X_0 = \{v \in V : \theta(v) = 0 \ \forall \theta \in X\} = \bigcap_{\theta \in X} \ker \theta \leq V$$

Then $\dim(X_0) = n - \dim(\langle X \rangle)$ and every $U \leq V$ is X_0 for some $X \leq V^*$, namely $X = U^0$.

4.2 Polynomials

Let $\mathbb{F}$ be a field.

Definition. (Polynomial) A *polynomial* f over $\mathbb{F}$ is a formal expression:

$$f(t) = \sum_{i=0}^{n} a_i t^i \quad n \in \mathbb{Z}_{\geq 0},\ a_i \in \mathbb{F}$$

Then we say that $\mathbb{F}[t]$ is the $\mathbb{F}$ vector space of all polynomials, with a basis $\{1, t, t^2, \dots\}$.

Definition. (Degree) The *degree* of f, written, $\deg f$ is the largest i such that $a_i \neq 0$. We also say that $\deg 0 = -\infty$.

We say that a_i is the *leading coefficent* of the polynomial and if $a_i = 1$, we say that the polynomial is monic.

We can perform addition and multiplication of polynomials in the usual sense. We have a multiplicative and additive identity (1 and 0 respectively) and additive inverses. We can also distribute over these operations, which means that $\mathbb{F}[t]$ forms a *ring* (See IB Groups, Rings and Modules).

Note that $\deg(f + g) \leq \max(\deg(f), \deg(g))$ and $\deg(fg) = \deg f + \deg g$. We can write $f \mid g$ if $\exists h \in \mathbb{F}[t]$ such that $g = fh$.

For $\lambda \in \mathbb{F}$ write $f(\lambda) = \sum_{i=0}^{n} a_i \lambda^i \in \mathbb{F}$. We can see that evaluation resepctives addition and multiplication. We distinguish between $\mathbb{F}[t]$ and the space of polynomial maps $\mathbb{F} \to \mathbb{F}$ since if $\mathbb{F}$ is finite, then $\mathbb{F}[t]$ is not finite dimensional, but the space of polynomial maps is a subspace of $\mathbb{F}^{\mathbb{F}}$ which is a finite dimensional vector space, so they're not even isomorphic spaces. For example if we're in the field $\mathbb{F}^4$ we can construct the polynomial

$$t(t-1)(t-2)(t-3) \in \mathbb{F}[t]$$

which is not zero in $\mathbb{F}[t]$, but when viewed as a function from $\mathbb{F}^4 \to \mathbb{F}^4$, induces the zero map.

Proposition. (Euclidean algorithm for polynomials in X) Let K be a field and $f, g \in K[X]$. Then there exists polynomials $r, q \in K[X]$ such that $f = gq + r$ with $\deg(r) < \deg(g)$.

(From IB Groups, Rings and Modules)

Proof. Let n be the degree of f. So $f = \sum_{i=0}^{n} a_i X^i$ with $a_i \in K, a_n \neq 0$. Similarly $g = \sum_{i=0}^{m} b_i X^i$ with $b_i \in K$ and $b_m \neq 0$.

If $n < m$ set $q = 0$ and $r = f$ so we're finished.

If instead $n \geq m$, proceed by induction on the degree. Let $f_1 = f - a_n b_m^{-1} X^{n-m} g$. Observe that $\deg(f_1) < n$. If $n = m$ then $\deg(f_1) < n = m$. So write $f = (a_{b_m^{-1}} X^{n-m})g + f_1$, so we're done. Otherwise if $n > m$, then because $\deg(f_1) < n$, by induction we can write $f_1 = gq_1 + r_1$ where $\deg(r_1) < \deg(g) = m$. Then $f = (a_n b_m^{-1}) X^{n-m} g + q_1 g + r_1 = (a_n b_m^{-1} X^{n-m} + q_1)g + r_1$ □

Corollary. (Bezout's Lemma) If $f_1, \ldots, f_n \in \mathbb{F}[t]$ have no common divisor of degree ≥ 1 (i.e the gcd is a unit) then $\exists g_1, \ldots, g_n \in \mathbb{F}[t]$ such that

$$\sum_{i=1}^{n} f_i g_i = 1 \quad \in \mathbb{F}[t].$$

Proof. Same as in $\mathbb{Z}$.

Lemma. For $\lambda \in \mathbb{F}$,

$$f(\lambda) = 0 \iff (t - \lambda) \mid f(t).$$

Proof. Apply the Euclidian algorithm to $f(t)$ and $g(t) = t - \lambda$.

Definition. (Root) $\lambda \in \mathbb{F}$ is a *root* of $f \in \mathbb{F}[t]$ of *multiplicity* greater than e if

$$(t - \lambda)^e \mid f(t).$$

Corollary. If $\deg f = n \geq 0$, then f has at most n roots counted with multiplicity.

Corollary. If $\deg f, \deg g < n$, and there exists $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ distinct such that $f(\lambda_i) = g(\lambda_i)$ for $1 \leq i \leq n$, then $f = g$.

Theorem. (Fundamental Theorem of Algebra) Every $f \in \mathbb{C}[t]$ of $\deg f \geq 1$ has exactly n roots counting multiplicity. i.e. f is a product of polynomials of degree 1.

Proof. IB Complex Analysis

5 Eigenspaces

Definition. (Diagonalisable) Let V be a finite dimensional vector space, and $\alpha \in \mathcal{L}(V, V)$. Then α is *diagonalisable* if there exists a basis B of V such that $[\alpha]_B^B$ is a diagonal matrix.

Definition. (Triangularisable) Let V be a finite dimensional vector space, and $\alpha \in \mathcal{L}(V, V)$. Then α is *triangularisable* if there exists a basis B of V such that $[\alpha]_B^B$ is an upper-triangular matrix.

A matrix $A \in M_{n\times n}(\mathbb{F})$ is diagonalisable or triangularisable if A is similar to a diagonal or upper triangular matrix respectively.

Remark. By change of basis, α is diagonalisable or triangularisable if and only if for any basis B of V, $[\alpha]_B^B$ is a diagonalisable or respectively triangularisable matrix.

(i) A triangularisable matrix has a very easy to compute determinate; we can just take the product of the entries in the leading diagonal.

(ii) A diagonalisable matrix makes it easier to understand its similarity class.

Remark. If $B = \{v_1, v_2, \ldots, v_n\}$ is such that

$$[\alpha]_B^B = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

then $\alpha(v_i) = \lambda_i v_i$ for all $1 \leq i \leq n$.

Definition. (Eigenvalues, Eigenvectors, and Eigensapces) Let V be a $\mathbb{F}$-vector space and $\alpha \in \mathcal{L}(v, v)$. An element $\lambda \in \mathbb{F}$ is an *eigenvalue* of α if there exists some $v \in$ non-zero such that $\alpha(v) = \lambda v$. Such a vector is a λ-*eigenvector* of α.

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

is the λ-*eigenspace* of α.

Remark. We can make some remarks from this definition.

(i) $V_\lambda = \ker(\alpha - v\mathrm{id}_V) \leq V$;

(ii) For V finite dimensional, α is diagonalisable if and only if V has a basis of eigenvectors of α.

(iii) If $v \in V_\mu$ for some $\lambda \neq \mu$,

$$(\alpha - \lambda \mathrm{id}_V)(v) = (\mu - \lambda)(v).$$

Thus $(\alpha - \lambda \mathrm{id}_V)$ preserves V_μ and $(\alpha - \lambda \mathrm{id}_V)|_{V\mu}$ is invertible with inverse $v \to (\mu - \lambda)^{-1} v$.

Lemma. If $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ are distinct and $\mathbf{0} \neq v_i \in V_{\lambda_i}$ then $\{v_1, \ldots, v_k\}$ is linearly independent.

Proof. Suppose that $\{v_1, \ldots, v_k\}$ is linearly dependent. Let

$$\sum_{i=0}^{k} \mu_i v_i = \mathbf{0} \quad \text{in } V$$

and *wlog* we have that $\mu_1 \neq 0$. Let

$$\beta = \prod_{i=2}^{k} (\alpha - \mu_i \mathrm{id}_V).$$

Then by remark (iii) we have that

$$\beta(v_j) = \left(\prod_{i=2}^{k} (\lambda_j - \lambda_i) \right) \cdot v_j \neq \mathbf{0}$$

if and only if $j = 1$. Thus,

$$\mathbf{0} = \beta \left(\sum_{i=1}^{n} \mu_i v_i \right) = \left(\mu_1 \prod_{i=2}^{k} (\lambda_1 - \lambda_i) \right) v_1$$

so $\mu_1 = 0$, which is a contradiction. □

Corollary. If V is a finite dimensional vector space, then every $\alpha \in \mathcal{L}(V, V)$ has only finitely many eigenvalues.

Proposition. Let V be a finite dimensional $\mathbb{F}$-vector space, and let $\alpha \in \mathcal{L}(V, V)$. Let $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ be all the eigenvalues of α. Then,
(i) $\langle V_{\lambda_1} \cup \cdots \cup V_{\lambda_k} \rangle = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$;
(ii) α is diagonalisable if and only if $V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$.

Proof.

(i) If not, then $\exists v_j \in V_{\lambda_j}$ not all zero such that

$$\sum_j v_j = \mathbf{0}$$

which is a contradiction from the lemma.

(ii) Suppose that α is diagonalisable. By the remark V has a basis of eigenvectors, so the V_{λ_i} spans V. Conversely, suppose that $V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$. Let B_i be a basis of V_{λ_i}. Then by Example Sheet 1 we have that

$$\bigcup_{i=1}^{k} B_i = B$$

is a basis for V, hence V has a basis of eigenvectors, so α is diagonalisable by the remark. $\square$

Recall that if V is a finite dimensional vector space, then any $\beta \in \mathcal{L}(V, V)$ we have that

$$\det(B) = 0 \iff \ker(B) \neq \{\mathbf{0}_V\}.$$

Thus $\lambda \in \mathbb{F}$ is an eigenvalue of $\alpha \in \mathcal{L}(v, v)$ if and only if

$$\det(\alpha - \lambda \mathrm{id}_V) = 0.$$

Definition. (Characteristic polynomial) For $A \in M_{n\times n}(\mathbb{F})$, the *characteristic polynomial* of A is

$$\chi_A(t) = \det(tI_n - A) \in \mathbb{F}[t].$$

Similarly for $\alpha \in \mathcal{L}(V, V)$ with V finite dimensional we can define

$$\chi_\alpha(t) = \det(t \cdot \mathrm{id}_V - \alpha)$$

as the characteristic polynomial of α.

Lemma. For V a finite dimensional vector space, and $\alpha \in \mathcal{L}(, V)$, we have that the set of roots of $\chi_A(t)$ are exactly the set eigenvalues of α.

Remark. We can make some remarks about the characteristic polynomial.

(i) χ_α is monic of degree $\dim V$.

(ii) Similar matrices have the same characteristic polynomial.

(iii) By the Leibniz formula, if $\chi_\alpha(t) = t^n + C_{n-1}t^{n-1} + \cdots + C_1 t + C_0$ we have that

(a) $C_{n-1} = \mathrm{tr}(\alpha)$, $C_0 = (-1)^n \det \alpha = \chi_\alpha(0)$.

Proposition. If α is triangularisable then $\chi_\alpha(t)$ can be written as aproduct of linear factors.

Proof. Let B be a basis for V such that

$$A = [\alpha]_B^B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix}.$$

Then

$$\chi_\alpha(t) = \prod_{i=1}^{n} (t - a_i).$$

For example the matrix,

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

Then $\chi_A(t) = t^2 - (2\cos\theta)t + 1$ is irreducible for $\theta \in (0, \pi)$, hence A is not triangularisable in $M_{2\times 2}(\mathbb{R})$.

Theorem. Every $A \in M_{n\times n}(\mathbb{C})$ is triangularisable.‘

Proof. By the fundamental theorem of algebra, $\chi_A(t)$ has a root $\lambda \in \mathbb{C}$, so there exists a v_1, non-zero in $\mathbb{C}^n$ such that $Av_1 = \lambda v_1$. Extend to a basis $B = \{v_1, \ldots, v_n\}$ of $\mathbb{C}^n$. So up to a change of basis, we can assume that

$$A = \begin{pmatrix} \lambda_1 & a_{12} & \cdots & a_{1n} \\ 0 & & & \\ \vdots & & C & \\ 0 & & & \end{pmatrix}$$

where C is a $(n-1) \times (n-1)$ matrix. By induction there exists a $\bar{P} \in GL_{n-1}(\mathbb{C})$ such that $\bar{P}C\bar{P}^{-1}$ is upper triangular. So set

$$P = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \bar{P} & \\ 0 & & & \end{pmatrix}.$$

Then PAP^{-1} is upper-triangular.

Corollary. Let V be a finite dimensional $\mathbb{C}$-vector space. Then every $\alpha \in \mathcal{L}(V, V)$ is triangularisable.

Remark. The same proof yields the result for any field $\mathbb{F}$, $A \in M_{n\times n}(\mathbb{F})$ is triangularisable if and only if $\chi_A(t)$ is a product of linear factors.

From now on we'll assume all vector spaces are finite dimensional and $\alpha \in \mathcal{L}(V, V)$.

Let's introduce some notation. For a polynomial

$$p(t) = \sum_{k=0}^{n} \mu_i t^i$$

we can introduce the linear map

$$p(\alpha) = \sum_{i=0}^{n} \mu_i \alpha^i \in \mathcal{L}(V, V).$$

Similarly for $A \in M_{k\times k}(\mathbb{F})$ we have that $p(A) = \sum_{i=0}^{n} \mu_i A^i$.

Remark. We have some remarks about these new linear maps.

(i) If B is a basis for V then
$$[p(\alpha)]_B^B = p([\alpha]_B^B).$$

(ii) If $\dim V = n$ then since we know that $\dim(\mathcal{L}(V,V)) = n^2$, so $\{\mathrm{id}_V, \alpha^1, \alpha^2, \ldots, \alpha^{n^2}\}$ is linearly dependent. So there exists $\mu_i \in \mathbb{F}$ such that,
$$\sum_{i=0}^{n^2} \mu_i \alpha^i = 0 = p(\alpha)$$
where $p(t) = \sum_{i=0}^{n^2} \mu_i t^i \neq 0$ in $\mathbb{F}[t]$.

Let
$$I_\alpha = \{p(t) \in \mathbb{F}[t] : p(\alpha) = 0\} \lhd \mathbb{F}[t].$$
So by the remark we know that $I_\alpha \neq \{0\}$. Let $m_\alpha \in I_\alpha$ be a non-zero element of minimal degree. Rescaling we can assume that m_α is monic.

Definition. (Minimal polynomial) We call m_α the *minimial polynomial*.

Proposition. The minimial polynomial is unique.

First we have to prove a lemma.

Lemma. For any $f \in I_\alpha$, we have that $m\alpha \mid f$.

Proof. Apply the Euclidean algorithm to $\mathbb{F}[t]$. So there exists $r, q \in \mathbb{F}[t]$ such that $\deg r < \deg m_\alpha$, and $f = qm_\alpha + r$. Then $r(\alpha) + f(\alpha) - q(\alpha)m_\alpha(\alpha) = 0$. Hence $r \in I_\alpha$, so since it has a degree less than m_α we must have that $r = 0$. Or we could prove the statement since $\mathbb{F}[t]$ is a principal ideal domain hence we can show that $I = (m_\alpha)$. □

Now we can prove the proposition

Proof. Let $\bar{m}_\alpha$ be another non-zero element of minimial degree in I_α, monic. By the lemma we have that $m_\alpha \mid \bar{m}_\alpha$, but $\deg(m_\alpha) = \deg(\bar{m}_\alpha)$ so there exists a $\lambda \in \mathbb{F}$ such that $\bar{m}_\alpha = \lambda m_\alpha$. Hence $\lambda = 1$, so they are the same polynomial. □

Theorem. (Cayley-Hamilton Theorem) For $\alpha \in \mathcal{L}(V,V)$ we have that
$$\chi_\alpha(\alpha) = 0.$$

Let's give a proof valid for $\mathbb{F} = \mathbb{C}$. The general proof for an arbitrary proof is a bonus exercise on Example Sheet 3.

Proof. We know that there is a basis B for V such that when α is written as a matrix in B it is upper-triangular. So we have that
$$\chi_\alpha(t) = \prod_{i=1}^{n} (t - \lambda_i)$$

where λ_i is the element in the ith row of the leading diagonal. Set $U_j = \langle v_1, \ldots, v_j \rangle$, so we have that $U_0 = \{\mathbf{0}_V\}$, and $U_n = V$.
Then for $1 \leq j \leq n$ we have that $(\alpha_{\lambda j}\mathrm{id}_V)(U_j) \leq U_{j-1}$ thus

$$\begin{aligned}\chi_\alpha(\alpha)(V) &= (\alpha - \lambda_i \mathrm{id}_V) \circ \cdots \circ (\alpha_{\lambda n}\mathrm{id}_V)(U_n) \\ &= U_0 = \{\mathbf{0}_V\}.\end{aligned}$$

so $\chi_\alpha(\alpha)$. □

Our proof is valid for α triangularisable over every field.

Corollary. $m_\alpha \mid \chi_\alpha$.

Definition. (Algebraic multiplicity) Let $\lambda \in \mathbb{F}$ be an eigenvalue of α. The *algebraic multiplicity* a_λ of α is the multiplicity of λ as a root of $\chi_\alpha(t)$.

Definition. (Geometric multiplicity) Let $\lambda \in \mathbb{F}$ be an eigenvalue of α. The *geometric multiplicity* g_λ of λ is $\dim(V_\lambda)$.

Proposition. Let $\lambda \in \mathbb{F}$ be an eigenvalue of α then we have that

$$g_\lambda \leq a_\lambda.$$

Proof. Let $\{v_1, \ldots, v_k\}$ be a basis for V_λ. Extend to a basis $B = \{v_1, \ldots, v_n\}$ for V. Then,

$$[\alpha]_B^B = \begin{pmatrix} \lambda I_k & B \\ 0 & C \end{pmatrix}$$

is a block triangular matrix, so $\chi_\alpha(t) = (t-\lambda)^k \chi_t(t)$. Hence we have that $a_\lambda \geq k = g_\lambda$. □

Lemma. Let the eigenvalues of α be $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$.

(i)
$$\sum_{i=1}^{k} g_{\lambda_i} \leq \dim V$$
with equality if and only if α is diagonalisable.

(ii)
$$\sum_{i=1}^{k} a_{\lambda_i} \leq \dim V$$
with equality if and only if α is triangularisable.

Proof.

(i) We have that $V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k} \leq V$ with equality if and only if α is diagonalisable. hence the result is clear.

(ii) We write $\chi_\alpha(t) = f(t)\prod_{i=1}^{k}(t-\lambda_i)^{a_{\lambda_i}}$. Then $f(t)$ has no linear factors, since a root of χ_α is an eigenvalue. Moreover

$$\begin{aligned}\dim(V) &= \dim(\chi_\alpha)\\ &= \deg(f) + \sum_{i=1}^{k} a_{\lambda_i}\\ &\leq \sum_{i=1}^{k} a_{\lambda_i}.\end{aligned}$$

with equality if and only if χ_α is a product of linear factors if and only if α is triangularisable. □

Proposition. For $\mathbb{F} = \mathbb{C}$, α is diagonalisable if and only if $a_\lambda = g_\lambda$ for every eigenvalue λ of α.

Proof. Since we're working over $\mathbb{C}$ we know that α is triangularisable. Hence

$$\sum_{i=1}^{k} a_{\lambda_i} = \dim V.$$

By the above proposition we have that

$$\sum_{i=1}^{k} g_{\lambda_i} \leq \sum_{i=1}^{k} a_{\lambda_i} = \dim V.$$

with equality if and only if $a_{\lambda_i} = g_{\lambda_i}$ for all i. But this occurs if and only if α is diagonalisable. □

Remark. For $p(t) = \sum_{i=1}^{n} \mu_i t^i$, with $v \in V_\lambda$.

$$p(\alpha)(v) = \sum_{i=1}^{n} \mu_i \alpha^i(v) = \sum_{i=1}^{n} \mu_i \lambda^i v = p(\lambda)v.$$

Lemma. For any $\lambda \in \mathbb{F}$, λ is a root of m_α is and only if it is a root of χ_α.

Proof. For the forward direction if λ is a root of $m_\alpha(t)$, by Cayley-Hamilton,

$$(t-\lambda) \mid m_\alpha(t) \mid \chi_\alpha(t).$$

Conversely if $\chi_\alpha(\lambda) = 0$ then λ is an eigenvalue of α, so let $\mathbf{0} \neq v \in V_\lambda$. Then $m_\alpha(\alpha)(v) = m_\alpha(\lambda)\cdot v$. But $m_\alpha(\alpha) = 0$, so since $v \neq \mathbf{0}$, we must have that $m_\alpha(\lambda) = 0$. □

We will introduce the notation that C_λ is the multiplicity of λ as a root of $m_\alpha(t)$.

Remark. By Cayley-Hamilton, if λ is an eigenvalue, then $1 \leq C_\lambda \leq a_\lambda$.

However there is no useful relationship between C_λ and g_λ. Let's see this in two examples.

(i) Let $A = \lambda I_n$. Then $\chi_A(t) = (t-\lambda)^n$, $V_k = \mathbb{F}^n$, so $m_\alpha(t) = t - \lambda$. Hence $C_\lambda = 1$, $g_\lambda = n$, $a_\lambda = n$.

(ii) Let

$$A = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix}.$$

Then $\chi_A(t) = (t-\lambda)^n$, so

$$V_\lambda = \langle \mathbf{e}_1 \rangle,$$

so

$$(A - \lambda I_n)^{n-1} \neq 0$$

hence $a_\lambda = n$, $g_\lambda = 1$, $C_\lambda = n$, as $m_\alpha = \chi_\alpha$.

Theorem. If there exists a $p(t) \in \mathbb{F}[t]$ non-zero such that p is a product of *distinct* linear factors, such that $p(\alpha) = 0$ then α is diagonalisable.

Proof. We can assume that p is monic. Write

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

Set

$$p_i(t) = \prod_{j \neq i} (t - \mu_j)$$

Then by Bezout, there exists $q_i(t) \in \mathbb{F}[t]$ such that

$$1 = \sum_{i=1}^{k} p_i(t) q_i(t) \quad \text{in } \mathbb{F}[t].$$

Set $\pi_i = p_i(\alpha) \circ q_i(\alpha) \in \mathcal{L}(V, V)$. Then

$$\mathrm{id}_V = \sum_{i=1}^{k} \pi_i$$

So $v = \mathrm{id}_V(v) = \sum_{i=1}^{k} \pi_i(v)$. Hence $V = \mathrm{im}(\pi_1) + \cdots + \mathrm{im}(\pi_k)$. Moreover $(t - \mu_i)p_i(t) = p(t)$, so $(\alpha - \mu_i \mathrm{id}_V) \circ p_i(\alpha) = p(\alpha) = 0$. Hence $(\alpha - \mu_i \mathrm{id}_V) \circ \pi_i = 0$. So $\mathrm{im}(\pi_i) \leq V_{\mu_i}$. Hence

$$V = \sum_{i=1}^{k} \mathrm{im}(\pi_i) \leq \bigoplus_{i=1}^{k} V_{\mu_i} \leq V.$$

So we have

$$V = \bigoplus_{i=1}^{k} V_{\mu_i}$$

so α is diagonalisable. □

Let's look an example
Take $A \in GL_n(\mathbb{C})$ with finite order. Then A is diagonalisable. Why? let m be the order of A. So $A^m = I_n$, so A satisfies the polynomial $p(t) = t^m - 1 = \prod_{i=1}^{m}(t - \omega^i)$ where $\omega = \exp\left(\frac{2\pi i}{m}\right)$. Hence by the theorem A is diagonalisable.

We'll summarise some of the theorems about diagonalibility in one theorem.

Theorem. The following statements are equivalent.
(i) α is diagonalisable;
(ii) V has a basis consisting of the eigenvectors of α;
(iii) There exists a $p(t) \in \mathbb{F}[t]$ non-zero which is a product of distinct linear factors such that $p(\alpha) = 0$;
(iv) $m_\alpha(t)$ is a product of distinct linear factors.
Moreover if $\mathbb{F} = \mathbb{C}$ these are also equivalent to

$$a_\lambda = g_\lambda$$

for all $\lambda \in \mathbb{C}$.

Proof. (i) $\iff$ (ii), (i) $\iff$ (v) already proved. (iii) $\implies$ (iv) follows from $m_\alpha \mid p$. (iv) $\implies$ (iii) by setting $p = m_\alpha$. (iii) $\implies$ (i) was the previous theorem. Hence we are only left to show that (i) $\implies$ (iii). Let B be a basis for V such that $[\alpha]_B^B$ is diagonal. Rearranging B, we can assume that

$$[\alpha]_B^B = \begin{pmatrix} \lambda_1 I_{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k I_{n_k} \end{pmatrix}.$$

Set $p(t) = \prod_{i=1}^{k}(t - \lambda_i)$. Then $[p(\alpha)]_B^B = p([\alpha]_B^B) = 0$. Hence $p(\alpha) = 0$. □

Definition. (Simultaneously diagonalisable) Let $\alpha, \beta \in \mathcal{L}(V, V)$. We say that α and β are *simultaneously diagonalisable* if there exists a basis B for V such that $[\alpha]_B^B$ and $[\beta]_B^B$ are both diagonal.

Theorem. Suppose that α and β are diagonal. Then they are simultaneously diagonalisable if and only if they commute.

Proof. Suppose that α and β are simultaneously diagonalisable. Given some B in the definition, we know that

$$\begin{aligned} [\alpha \circ \beta]_B^B &= [\alpha]_B^B [\beta]_B^B \\ &= [\beta]_B^B [\alpha]_B^B = [\beta \circ \alpha]_B^B \end{aligned}$$

since diagonal matrices commute. Hence $\alpha \circ \beta = \beta \circ \alpha$.

For the converse we need a lemma first.

Lemma. If α and β commute and V_λ is the λ-eigenspace of α, then $\beta(V_\lambda) \leq V_\lambda$.

Proof. For $v \in V_\lambda$,

$$\alpha(\beta(V)) = \beta(\alpha(v)) = \beta(\lambda v) = \lambda\beta(v).$$

Hence $\beta(v) \in V_\lambda$. □

Now we assume that we have α and β both diagonalisable which commute. Let $V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$. Let $p(t)$ be a non-zero polynomial which is a product of distinct linear factors such that $p(\beta) = 0$. By the lemma we can define $\beta|_{V_{\lambda_i}} : V_{\lambda_i} \to V_{\lambda_i}$. Then $p(\beta|_{V_{\lambda_i}}) = p(\beta)|_{V_{\lambda_i}} = 0$. Hence $\beta|_{V_{\lambda_i}}$ is diagonalisable. Let B_i be a basis of V_{λ_i}, consisting of the eigenvectors of β. Then by Example Sheet 1,

$$B = \bigcup_{i=1}^{k} B_i$$

is a basis for V, and consists of vectors which are eigenvectors for both α and β. Thus α and β are both diagonal when written in this basis. □

From now on we'll focus on $\mathbb{F} = \mathbb{C}$.

Definition. (Jordan matrix) For $\lambda \in \mathbb{C}$, the $(n \times n)$-*Jordan matrix* is

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix} \in M_{n\times n}(\mathbb{C}).$$

Remark. We found that the characteristic polynomial for $J_n(\lambda)$ is $\chi_{J_n(\lambda)}(t) = (t-\lambda)^n = m_{J_n(\lambda)(t)}$.

Definition. (Jordan normal form) A matrix $A \in M_{n\times n}(\mathbb{C})$ is in *Jordan normal form* if

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

for some $n_i \in \mathbb{N}$ such that $\sum_{i=1}^{k} n_i = n$ and some $\lambda_i \in \mathbb{C}$ not necessarily distinct.

Theorem. Every $A \in M_{n\times n}(\mathbb{C})$ can be written in Jordan normal form uniquely up to a reordering of the Jordan blocks.

Proof. See IB Groups, Rings and Modules for existance. We will prove uniqueness later.

Let's look an example with $\lambda_i \in \mathbb{C}$ distinct.
For $n = 2$ the possible Jordan normal forms are

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

These have minimial polynomials $(t-\lambda_1)^2,\ (t-\lambda_1)(t-\lambda_2),\ (t-\lambda_1)$ respectively.

For $n=3$ the possible Jordan normal forms are the three diagonal matrices (one all same, one two same, one all different), the matrix with a single Jordan normal block, and the two matrices with a two Jordan blocks, one with the two eigenvalues matching, the other with the two eigenvalues of the block different.

Remark. For $n \leq 3$, the Jordan normal form of A is uniquely determinated by the minimial polynomial and the characteristic polynomial.

However this is not true for $n > 3$ which we can see on Example Sheet 3.

Remark. If we have that

$$A = \begin{pmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_k \end{pmatrix} \in M_{m\times m}(\mathbb{C}),$$

with $A_i \in M_{n_i \times n_i}(\mathbb{C})$. Then for $p(t) \in \mathbb{C}[t]$,

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

so $m_a(t) = (m_{A_1}(t), \ldots, m_{A_k}(t))$, hence $C_\lambda(A) = \max_i(C_\lambda(A_i))$.

We also have that

$$\chi(A)(t) = \prod_{i=1}^{k} \chi_{A_i}(t)$$

hence $a_\lambda(A) = \sum_{i=1}^{k} a_\lambda(A_i)$.

Furthermore

$$\mathrm{rk}(A) = \sum_{i=1}^{k} \mathrm{rk}(A_i),$$

and

$$A - \lambda I_n = \begin{pmatrix} (A_1 - \lambda I_n) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (A_k - \lambda I_n) \end{pmatrix}$$

hence

$$\begin{aligned} g_\lambda(A) &= n - \mathrm{rk}(A - \lambda I_n) \\ &= \sum_{i=1}^{k} (n_i - \mathrm{rk}(A_i - \lambda I_{n_i})) \\ &= \sum_{i=1}^{k} g_\lambda(A_i). \end{aligned}$$

Proposition. For $A \in M_{n\times n}(\mathbb{C})$ in Jordan normal form,
(i) $a_\lambda(A)$ is the sum of the sizes of λ-Jordan blocks in A;
(ii) $g_\lambda(A)$ is the number of λ-Jordan blocks;
(iii) $C_\lambda(A)$ is the size of the largest λ-Jordan block.

Proof. Immediate from above calculations and a_λ, C_λ, g_λ for $J_n(\lambda)$ from the above remark. □

Let's see an example of putting a matrix in Jordan normal form.

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

so we see a matrix $P \in GL_3(\mathbb{C})$ such that PAP^{-1} is in Jordan normal form. We see that $\chi_A(t) = (t-1)^2(t-2)$ so the JNF is either

(i)
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

(ii)
$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

.

But

$$A - I_3 = \begin{pmatrix} 2 & -2 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

which has rank 2, namely $\ker(A - I_3) = \langle \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rangle$, so we're in case (ii). Set $v_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Similarly

$$\ker(A_2 I_3) = \langle v_1 \rangle = \langle \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} \rangle.$$

Finially solve $Av_2 = v_1 + v_2$, solving for v_2 gives that $v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.

So we can set $P = [v_1 \mid v_2 \mid v_3]$.

Now we will prove uniqueness of Jordan normal form (up to rearrangment of the blocks).

Proof. For $A, \overline{A} \in M_{n\times n}(\mathbb{C})$ with $A \sim \overline{A}$ anf $\overline{A}$ in JNF, define

$$R_{\lambda,r}(\overline{A}) = \Big(\text{Number of } \lambda\text{-J-blocks in } \overline{A} \text{ of size } \geq r\Big).$$

Then the $R_{\lambda_r}(\overline{A})$ determine $\overline{A}$ uniquely up to rearranging of the blocks. Hence uniquness follows from the lemma, only dependent on A.

Lemma.
$$R_{\lambda,r}(\overline{A}) = \mathrm{rk}((A_\lambda - I_n)^{r-1}) - \mathrm{rk}((A - \lambda I_n)^r)$$

Proof. For $P \in GL_n(\mathbb{C})$,
$$\begin{aligned} P(A-\lambda I_n)^k P^{-1} &= (P(A-\lambda I_n)P^{-1})^k \\ &= (PAP^{-1} - \lambda I_n)^k \end{aligned}$$
Thus $rk((\overline{A} - \lambda I_n)^k) = \mathrm{rk}((A-\lambda I_n)^k)$, hence we can assume that $A = \overline{A}$ is in JNF.
$$J_\ell(\lambda) - \lambda I_\ell = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$
and raising to the power n moves the diagonal of ones to the right. Thus
$$\mathrm{rk}((J_\ell(\lambda) - \lambda I_\ell)^k) = \begin{cases} \ell - k & k \leq \ell \\ 0 & k \geq \ell \end{cases}.$$
For $\mu \neq \lambda$, $J_\ell(\mu) - \lambda I_\ell$ is invertible hence
$$\begin{aligned} \mathrm{rk}((A-\lambda I_n)^k) = R &- k \cdot (\text{Number of } \lambda\text{-J-blocks in } \overline{A} \text{ of size } \geq r) \\ &- \sum (\text{Number of } \lambda\text{-J-blocks in } \overline{A} \text{ of size } < r) \end{aligned}$$
which we then can use to get the required lemma.

Hence the JNF is unique. □

Definition. (Generalised eigenspace) Let V be a finite dimensional $\mathbb{C}$-vector space and $\alpha \in \mathcal{L}(V,V)$. Write
$$m_\alpha(t) = (t-\lambda_1)^{C_1} \cdots (t-\lambda_k)^{C_k}$$
with $\lambda_i \in \mathbb{C}$ all distinct and $C_i \geq 1$. The *generalised λ_i-eigenspace* of α is
$$V_i = \ker((\alpha - \lambda_i \mathrm{id}_V)^{C_i}) \leq V.$$

Theorem. (Generalised eigenspace decomposition) Let V be as in the above definition, then
$$V = V_1 \oplus \cdots \oplus V_k.$$

Proof. Example Sheet 3. □

Lemma. If $\mathbf{0} \neq v_i \in V_i$, then $\{v_1, \ldots, v_k\}$ are linearly independent.

Proof. Let $0 \leq d_i < C_i$ be such that $\mathbf{0} \neq w_i = (\alpha - \lambda_i \mathrm{id}_V)^{d_i}(v_i)$. But $(\alpha - \lambda_i \mathrm{id}_V)^{d_i+1}(v_i) = \mathbf{0}$. So w_i is a λ_i-eigenvector of α, so $\{w_1, \ldots, w_k\}$ is linearly independent. Set
$$\beta = \prod_{j=1}^{k} (\alpha - \lambda_j \mathrm{id}V)^{d_j} \in \mathcal{L}(V,V).$$

As the $(\alpha - \lambda_j \mathrm{id}V)$ commute, we get,

$$\begin{aligned}\beta(v_i) &= \left(\prod_{j\neq i}(\alpha - \lambda_j \mathrm{id}_V)^{d_j}\right)(w_i) \\ &\left(\prod_{i\neq j}(\lambda_i - \lambda_j)^{d_j}\right) w_i \\ &= \mu_i w_i,\end{aligned}$$

with $\mu_i \neq 0$. Suppose

$$\mathbf{0} = \sum_{i=1}^{k} \nu_i v_i$$

for $\nu_i \in \mathbb{C}$. Then

$$\begin{aligned}\mathbf{0} &= \left(\sum_{i=1}^{k} \nu_i v_i\right) \\ &= \sum_{i=1}^{k} \mu_i \nu_i w_i\end{aligned}$$

so $\nu_i = 0$. □

We can give a sketch proof of the existance of Jordan normal form.

A commutes with $A - \lambda_i I_n$ so A preserves V_i. Thus we can write A in block diagonal form with each block acting on each generalised eigenspace. Each block has minimial polynomial $(t - \lambda_i)^{c_i}$ so it suffices to show each block has Jordan normal form i.e. we only have to show existance of JNF for matrices with a single eigenvalue. We can also replace our matrix A_i with $A_i - \lambda_i I$, so we can assume this eigenvalue is zero. So $A^m = 0$ i.e. that A is nilpotent. So we can reduce to the following theorem.

Theorem. Let V be a $\mathbb{C}$-vector space with $\dim V = n$, and $\alpha \in \mathcal{L}(V, V)$ nilpotent. Then there exists a basis $B = \{v_1, \ldots, v_n\}$ for V such that for all $1 \leq i \leq n$, $\alpha(v_i) \in \{v_{i-1}, \mathbf{0}\}$.

Proof. Induct on n. Let $W = \mathrm{im}(\alpha)$. Then $\dim W < n$ and $\alpha(W) \leq W$, so we can define $\alpha\mid_W \colon W \to W$ which is nilpotent since α is nipotent. By the inductive hypothesis we have a basis

$$B' = \{\alpha^{\ell_1}(w_1), \alpha^{\ell_1 - 1}(w_1), \cdots, \alpha(w_1), w_1, \alpha^{\ell_2(w_2)}, \cdots, \alpha(w_2), w_2, \cdots, w_k\}$$

with $\alpha^{\ell_i + 1}(w_i) = \mathbf{0}$. Write $w_i = \alpha(v_I)$ for some $v_i \in V$ and extend $\{\alpha^{\ell_1}(w_1), \cdots, \alpha^{\ell_k}(w_k)\} \subseteq \ker \alpha$ by $\{u_1, \ldots, u_m\}$ to a basis of $\ker \alpha$.

Check that $B = B' \cup \{u_1, \cdots, u_m\} \cup \{v_1, \cdots, v_k\}$ is linearly independent. By rank-nullity we have that $|B| = \dim V$, so up to reordering B is as desired. □

6 Bilinear forms

Let U, V be finite dimensional $\mathbb{F}$-vector spaces.

Definition. (Bilinear form) A function $\varphi : U \times V \to \mathbb{F}$ is a *bilinear form* if for each fixed $u_0 \in U, v_0 \in V$, we have that

$$v \to \varphi(u_0, v) \quad u \to \varphi(u, v_0)$$

are linear.

Remark. Be careful! Bilinear forms are almost *never* linear maps (from the external product of U and V.

Let's look at some examples

(i) $U = V = \mathbb{F}^n$, $\varphi(x, y) = \sum_{i=1}^n x_i y_i$.

(ii) For $A \in M_{m\times n}(\mathbb{F})$ we can define

$$\begin{aligned} \varphi_A : \mathbb{F}^m \times \mathbb{F}^n &\to \mathbb{F} \\ (x, y) &:\to x^T A y \end{aligned}$$

which is bilinear.

(iii) If $U = V = C[0, 1]$ we can define

$$\varphi(f, g) = \int_0^1 f(t)g(t)\mathrm{d}t.$$

(iv) $\varphi : V \times V^* \to \mathbb{F}$ given by $\varphi(v, \theta) \to \theta(v)$.

Definition. (Matrix representation) If U, V are finite dimensional and $B = \{b_1, \dots, b_m\}$, $C = \{c_1, \dots, c_n\}$ are basis for U, V respectively and $\varphi : U \times V \to \mathbb{F}$ is bilinear, then the *matrix* of φ with respect to B and C is

$$[\varphi]_{B,C} = (\varphi(b_i, c_j))_{i,j} \in M_{m\times n}(\mathbb{F}).$$

Proposition. $[\varphi]_{B,C}$ satisfies

$$([u])_B^T [\varphi]_{B,C} [v]_c = \varphi(u, v) \tag{$\star$}$$

for all $u \in U$, $v \in V$ and $[\varphi]_{B,C}$ is the *unique* matrix satisfying $(\star)$ for all u, v.

Proof. Let

$$u = \sum_{i=1}^m \lambda_i b_i, \quad v = \sum_{j=1}^n \mu_j c_j,$$

so that $[u]_B = \lambda$, $[v]_C = \mu$. Then

$$\begin{aligned} \varphi(u, v) &= \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j \varphi(b_i, c_j) \\ &= \lambda^T \cdot (\varphi(b_i, c_j))_{i,j} \mu. \end{aligned}$$

Now for uniqueness, if $A \in M_{m\times n}(\mathbb{F})$ satifies $(\star)$ for all u, v

$$\begin{aligned}\varphi(b_i, c_j) = \big([b_i]_B\big)^T A[c_j]_C &= e_i^T A e_j \\ &= a_{i,j}. \quad \square\end{aligned}$$

Corollary. If B, B' are basis for U and C, C' are basis for V, then

$$[\varphi]_{B',C'} = ([\mathrm{id}_U]_B^{B'})^T [\varphi]_{B,C} [\mathrm{id}_V]_C^{C'}$$

Proof. For $u \in U, v \in V$,

$$\begin{aligned}([u]_{B'})^T ([\mathrm{id}_U]_B^{B'})^T [\varphi]_{B,C} [\mathrm{id}_V]_C^{C'} [v]_{C'} &= ([\mathrm{id}_U]_B^{B'} [u]_{B'})^T [\varphi]_{B,C} [v]_C \\ &= ([u]_B)^T [\varphi]_{B,C} [v]_C \\ &= \varphi(u, v). \quad \square\end{aligned}$$

Definition. (Rank) The *rank* of φ is the rank of $[\varphi]_{B,C}$ for any basis B, C for U, V.

Remark. Let's make some remarks.

(i) Equivalent matrices have the same rank, so by the corollary this is well defined, as changing the matrix representation keeps the matrix in the equivalence class, hence they have the same rank still.

(ii) For $U = V$, B, B' basis for V there exists some $P \in GL_n(\mathbb{F})$ such that $[\varphi]_{B',B'} = P^T [\varphi]_{B,B} P$.

Definition. (Congurency) $A, A' \in M_{m\times n}(\mathbb{F})$ are *congruent* if there exists some $P \in GL_n(\mathbb{F})$ such that $A' = P^T A P$.

Proposition. Congruence is an equivalence relation.

Proof. Exercise.

Definition. Let $\varphi : U \times V \to \mathbb{F}$ be a bilinear form. For $u \in U, v \in V$ let

$$\varphi_L(u) \in V^*, \quad \varphi_R(v) \in U^*$$

be given by

$$\varphi_L(u)[v] = \varphi(u, v) = \varphi_R(v)[u].$$

Lemma. $\varphi_L : U \to V^*$, $\varphi_R : V \to U^*$ are linear.

Proof. Exercise.

Definition. (Left/Right-Kernel) Let φ_L, φ_R be as above. Then we define $\ker(\varphi_L) \leq U$ to be the *left-kernel* of φ and $\ker(\varphi_R) \leq V$ to be the *right-kernel* of φ.

Proposition. Let U, V be finite dimensional and B, C be basis for U, V and B^*, C^* be dual to the basis B, C respectively. Then
(i) $[\varphi_L]^B_{C^*} = [\varphi]_{B,C}$.
(ii) $[\varphi_R]^C_{B^*} = ([\varphi]_{B,C})^T$.

Proof. We'll just prove (i). Let $B = \{b_1, \cdots, b_m\}$, $C = \{c_1, \ldots, c_n\}$. Set

$$\varphi_L(b_i) = \sum_{j=1}^{n} a'_{i,j} c_j^*.$$

Write $A = [\varphi]_{B,C}$. Then $a'_{i,j} = \varphi_L(b_i)[c_j] = \varphi(b_i, c_j) = a_{i,j}$. The proof for (ii) is similar. □

Corollary. $\mathrm{rk}(\varphi_L) = \mathrm{rk}(\varphi_R) = \mathrm{rk}(\varphi)$.

Proof. By the proposition and the fact that row-rank and column-rank are the same, we get the equality.

Definition. For $S \subseteq U$ and $T \subseteq V$, we define

$$S^\perp = \{v \in V : \forall s \in S, \varphi(s, v) = 0\}$$
$${}^\perp T = \{u \in U : \forall t \in T, \varphi(u, t) = 0\}.$$

Remark. We can see that

(i) $S^\perp \leq V$ and ${}^\perp T \leq U$;

(ii) If $S_1 \subset S_2 \subset U$ then $S_2^\perp \leq S_1^\perp \leq V$. Similarly if $T_1 \subseteq T_2 \subseteq V$ then ${}^\perp T_2 \leq {}^\perp T_1$.

(iii) $U^\perp = \ker(\varphi_R)$ and ${}^\perp V = \ker(\varphi_L)$.

Definition. (Degeneracy) We say that φ is *degenerate* if $U^\perp$ or ${}^\perp V$ is nontrivial.

Proposition. For U, V finite dimensional and B, C basis for U, V then φ is non-degenerate if and only if $\dim U = \dim V$ and $[\varphi]_{B,C}$ is invertible.

Proof. Non-degeneracy $\iff$ $\ker(\varphi_L), \ker(\varphi_R)$ trivial. So this is equivalent to $\dim U = \mathrm{rk}(\varphi_L) = \mathrm{rk}(\varphi_R) = \mathrm{rk}(\varphi) = \dim V$ which occurs if and only if $[\varphi]_{B,C}$ is square of full rank.

Henceforth we'll consider bilinear forms with $U = V$. If V is finite dimensional, write $[\varphi]_B$ for $[\varphi]_{B,B}$.

Definition. (Symmetric bilinear form) A bilinear form $\varphi : V \times V \to \mathbb{F}$ is *symmetric* if $\forall u, v \in V$,

$$\varphi(u, v) = \varphi(v, u).$$

Remark. Some consequences,

(i) Recall that $A \in M_{n\times n}(\mathbb{F})$ is symmetric if $A^T = A$. Fir V finite dimensional, φ is symmetric if and only if for some (or equivalently any) basis B for V we have that $[\varphi]_B$ is symmetric.

(ii) If φ is symmetric, then for any $S \subseteq V$, ${}^S = S^\perp$.

(iii) In particular, the left kernel is the same as the right kernel which will just call the kernel of φ.

Definition. (Quadratic form) A *quadratic form* on V is a function $Q : V \to \mathbb{F}$ such that there exists a bilinear form $\varphi : V \times V \to \mathbb{F}$ such that for all $v \in V$, $Q(v) = \varphi(v, v)$.

Let's see an example. Set $V = \mathbb{R}^2$. For $A \in M_{2\times 2}(\mathbb{F})$, define $Q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = a_{11}x^2 + (a_{12} + a_{21})xy + a_{22}y^2$. So the corresponding bilinear form is $\varphi(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T A\mathbf{y}$.
Note that $Q_A = Q_{\frac{1}{2}(A+A^T)}$.

Proposition. (Polarisation Identity). Suppose that $\mathbb{F}$ is a field of characteristic greater than 2. Let $Q : V \to \mathbb{F}$ be quadratic form. Then there is unique symmetric bilinear form $\psi : V \times V \to \mathbb{F}$ such that,

$$Q(v) = \psi(v, v) \quad \forall v \in V. \tag{$\star$}$$

And ψ is given by

$$\psi(u, v) = \frac{1}{2}(Q(u + v) - Q(u) - Q(v)). \tag{$\dagger$}$$

Proof. First we'll prove existance . Let φ be bilinear as in the definition of a quadratic form. Set $\psi(u, v) = \frac{1}{2}(\varphi(u, v) + \varphi(v, u))$, so ψ is bilinear and symmetric by definition.

For uniqueness, suppose ψ is a symmetric bilinear form satisfying ($\star$). Then

$$\begin{aligned} Q(u + v) &= \psi(u + v, u + v) \\ &= \psi(u, u) + \psi(u, v) + \psi(v, u) + \psi(v, v) \\ &= Q(u) + Q(v) + 2\psi(u, v), \end{aligned}$$

so ψ satisfies ($\dagger$). □

Theorem. Suppose $\mathbb{F}$ is a field of characteristic greater than 2 and V is a finite dimensional $\mathbb{F}$-vector space. Let $\varphi : V \times V \to \mathbb{F}$ be a symmetric bilinear form. Then there is a basis B for V such that $[\varphi]_B$ is diagonal.

Proof. Induct on the dimension, $n = \dim V$. If $\forall v \in V$, $\varphi(v, v) = 0$, then by polarisation, $\varphi = 0$ which is diagonal so we're done. Otherwise there exists some $\mathbf{0} \neq v \in V$ such that $Q(v_1, v_1) \neq 0$. Set $U = \langle v_1\rangle^\perp = \ker(\varphi_L(v_1) : V \to \mathbb{F})$. Then $v_1 \notin U$, so by rank-nullity, $\dim U = n - 1$ and $V = \langle v_1\rangle \oplus U$. Define $\varphi\,|_U \colon U \times U \to \mathbb{F}$ from φ, so it is symmetric and bilinear, so by induction

there exists a basis $B_U = \{v_2, \ldots, v_n\}$ such that $[\varphi\,|_U]_{B_U}$ is diagonal. Then $B = \{v_1, \ldots, v_n\}$ is a basis for V and

$$[\varphi]_B = \begin{pmatrix} \varphi(v_1, v_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \varphi(v_n, v_n) \end{pmatrix}.$$

Hence we're done. □

Remark. The assumption that φ is symmetric is necessary in the theorem. If $P^T AP = D$ is diagonal, since $D = D^T$ we have that $(P^T AP)^T = P^T A^T P$, so $A = A^T$, since P is invertible.

In practice, shortcuts are available.

Corollary. Suppose $\mathbb{F}$ is a field of characteristic greater than 2. Then every symmetric matrix is congruent to a diagonal matrix.

Corollary. Suppose

(i) V is a finite dimensional $\mathbb{C}$-vector space, and φ is as in the theorem. Let $Q(v) = \varphi(v, v)$. Then there is a basis $B = \{v_1, \ldots, v_n\}$ for V such that

$$[\varphi]_B = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}, \quad r = r\mathrm{rk}(\varphi).$$

(ii) If instead V is a $\mathbb{R}$-vector space,

$$[\varphi]_B = \begin{pmatrix} I_p & 0 & 0 \\ 0 & I_q & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad p + q = \mathrm{rk}(\varphi).$$

Proof. Pick basis $C = \{c_1, \ldots, c_n\}$ such that,

$$[\varphi]_C = \begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{pmatrix}.$$

Reorder indices such that

- $a_i \neq 0$ if and only if $i \leq r = \mathrm{rk}(\varphi)$.
- If $\mathbb{F} = \mathbb{R}$, $r = p + q$ and $a_i > 0$ for $i \leq p$, and $a_i < 0$ for $p + 1 \leq i \leq p + q$.

Then,

(i) If $\mathbb{F} = \mathbb{C}$, set

$$v_i = \begin{cases} \frac{1}{\sqrt{a_i}} c_i & i \leq r \\ c_i & i \geq r + 1. \end{cases}$$

(ii) If $\mathbb{F} = \mathbb{R}$, set

$$v_i = \begin{cases} \frac{1}{\sqrt{a_i}} c_i & i \leq p \\ \frac{1}{\sqrt{-a_i}} c_i & p + 1 \leq i \leq p + q \\ c_i & i \geq p + q + 1 \end{cases}$$

Henceforth $\mathbb{F} = \mathbb{R}$.

Definition. (Positive definite) Let $\varphi : V \times V \to \mathbb{R}$ be a symmetrical bilinear form. φ is said to be *positive definite* if $\varphi(v, v) > 0$ for all $\mathbf{0} \neq v \in V$. If we replace the $>$ with a $\geq$ and φ satisfies the inequality for all v we say that φ is *positive semi-definite*.

Definition. (Negative definite) Let $\varphi : V \times V \to \mathbb{R}$ be a symmetrical bilinear form. φ is said to be *negative definite* if $\varphi(v, v) < 0$ for all $\mathbf{0} \neq v \in V$. If we replace the $<$ with a $\leq$, and φ satisfies the inequality for all v we say that φ is *negative semi-definite*.

Definition. (Indefinite) If $\varphi : V \times V \to \mathbb{R}$ symmetrical is not positive semi-definite or negative semi-definite, we say that φ is *indefinite*.

Remark. Similar definitions exist for quadratic forms.

Let's see an example. Let V be finite dimensional and B be a basis such that

$$[\varphi]_B = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix}.$$

Then φ is positive semidefinite and positive definite if and only if $p = \dim V$.

Theorem. (Sylvester's Law of Inertia) Let V be finite dimensional and φ be a symmetric bilinear form on V. If B, B' are basis for V such that

$$[\varphi]_B = \begin{pmatrix} I_p & 0 & 0 \\ 0 & I_q & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad [\varphi]_{B'} = \begin{pmatrix} I_{p'} & 0 & 0 \\ 0 & I_{q'} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

then $p = p'$ and $q = q'$.

Definition. (Signature) We say that $\sigma(\varphi) = p - q$ is the *signature* of φ.

Corollary. $\sigma(\varphi)$ is well-defined.