LINEAR ALGEBRA

by

RYAN RUEGER

Some insights from my tutorial sessions given for the course *Linear Algebra*, held by PROF. DR. M. EINSIEDLER and PROF. DR. P. BIRAN at The Federal Technical Institute of Technology in Zurich, Spring 2023.

Unfortunately I had a horrible bicycle accident during the summer, pushing back the release of these notes by 4 weeks. I also had exams to prepare for myself, making it difficult to catch up and find the time to finalise these notes. I wished there to be many more examples and calculations to make some of the more abstract content more reachable. Alas. Here are 50 something pages on some of the things I thought about when teaching.

I apologise for incomplete proofs and trailing sentences. If you find any glaring errors, send me an email, and I will try to fix them. Of course questions are always welcome.

Contents

1

Injectivity, Surjectivity and composition

Definition (Image of a map)**.** Let $f\colon X \to Y$ be a map. The *image* of f is defined to be

$$\operatorname{im}(f) = \{y \in Y \mid \exists x \in X : f(x) = y\} = \{f(x) \in Y \mid x \in X\} \subseteq Y.$$

Definition (Preimage)**.** Let $f\colon X \to Y$ be a map and S a subset of Y. We define *preimage of S under f* to be

$$f^{-1}(S) = \{x \in X \mid f(x) \in S\} \subseteq X$$

Remark. Do not confuse the preimage with the inverse of a map, although the notation looks similar. An important point to remember, is that the preimage of a set always exists, but an inverse must not.

Definition (Injective, Into, One-to-one)**.** Let $f\colon X \to Y$ be a map. We call a map *injective* or *into* or *one-to-one* if one (and therefore all) of the following equivalent statements hold true

- For all $f(y), f(x)$ in $\operatorname{im}(f)$, such that $f(x) = f(y)$ we have that $x = y$.
- For all y in Y we have $\left|f^{-1}(\{y\})\right| \leq 1$.

Definition (Surjective, Onto)**.** Let $f\colon X \to Y$ be a map. We call a map *surjective* or *onto* if one (and therefore all) of the following equivalent statements hold true

- For every y in Y there is an x in X such that $f(x) = y$.
- For every y in Y we have $\left|f^{-1}(\{y\})\right| \geq 1$.
- $\operatorname{im}(f) = Y$.

Definition (Bijective, one-to-one correspondence)**.** Let $f\colon X \to Y$ be a map. We call a map *bijective* or a *one-to-one correspondence* if one (and therefore all) of the following equivalent statements hold true

- For every y in Y there is exactly one x in X such that $f(x) = y$.
- For every y in Y we have $\left|f^{-1}(\{y\})\right| = 1$.

We also call a bijective map simply a *bijection*.

Remark. The terms "into, onto, one-to-one, one-to-one correspondence" are older terms that are not used in modern literature that often. This is especially true for "one-to-one" and "one-to-one correspondence", whereas "onto" and "into" are still used in algebraic fields. Do not confuse the terms "one-to-one" and "one-to-one correspondence"

Lemma (Composition lemma)**.**

1. *The composition of injective maps is injective*
2. *The composition of surjective maps is surjective*
3. *The composition of bijective maps is bijective*

Now let $f\colon X \to Y$ *and* $g\colon Y \to Z$ *be maps.*

4. *If* $g \circ f$ *is surjective then* g *must be surjective (but not necessarily* f*).*
5. *If* $g \circ f$ *is injective, then* f *must be injective (but not necessarily* g*).*
6. *If* $g \circ f$ *is surjective and* g *injective (and therefore bijective by point 4), then* f *is surjective.*
7. *If* $g \circ f$ *is injective and* f *surjective (and therefore bijective by point 5), then* g *is injective.*

2

Hierarchy of Sets, Groups, Fields, ...

Definition ((Binary) operation, (Binäre) Verknüpfung)**.** Let X be a set. A *binary operation* (ger. *binäre Verknüpfung*) on X is a map $\mathrm{op}\colon X \times X \to X$. To obtain meaningful structures, almost all of the binary operations that we will study are *associative*, that is, for all x, y, z in X we have that $\mathrm{op}(\mathrm{op}(x,y),z) = \mathrm{op}(x,\mathrm{op}(y,z))$. Common symbols used for associative binary operations are $\cdot_X, \cdot, \odot, +, +_X, \oplus$. In these cases, we write $x \cdot y$ instead of $\cdot(x,y)$. When a binary operation $\cdot$ is associative, it becomes meaningful to write $x \cdot y \cdot z$ and we speak of an *operation* (dropping the term "binary"). Often we will study operations that are *commutative*, that is, for all x, y in X, $\mathrm{op}(x,y) = \mathrm{op}(y,x)$. When a (binary) operation is commutative (and associative), we conventionally use the notations $+, +_X$ or $\oplus$ and when the operation is only associative we conventionally use the notations $\cdot, \cdot_X$ or $\odot$.

Definition (Closed under a binary operation)**.** Let X be a set equipped with the binary operation op and Y a subset of X. We say that Y is *closed under the binary operation* op if for all y, y' in Y we have that $\mathrm{op}(y,y')$ also lies in Y.

Definition (A semigroup)**.** A (commutative) *semigroup* is a tuple $(S, \cdot)$ comprising a set X equipped with an associative (commutative) binary operation $\cdot$.

Definition (Neutral element of a binary operation)**.** Let X be a set with a binary operation $\mathrm{op}\colon X \times X \to X$. We say that e in X is a *neutral element of the binary operation* op if for all x in X we have that $\mathrm{op}(x,e) = \mathrm{op}(e,x) = x$. This is often shortened to "e is a neutral element".

Proposition (Neutral elements are unique)**.** *Any two neutral elements of an operation are equal.*

Proof. Let X be a set and $\mathrm{op}\colon X \times X \to X$ a binary operation. Suppose e, g are both neutral elements of op. Then $\mathrm{op}(e,g) = e$ by neutrality of g. On the other hand $\mathrm{op}(e,g) = g$ by neutrality of e. Hence $e = \mathrm{op}(e,g) = g$. □

Remark. When we furnish a set X with a binary operation op, we often say "X is associative/commutative" when we mean that the binary operation op equipped to X is associative/commutative.

Corollary. *Since neutral elements are unique, we may speak of* the *neutral element of an operation.*

Definition (Monoid)**.** A (commutative) *monoid* is a tuple $(M, \cdot, e)$ comprising a (commutative) semigroup $(M, \cdot)$ together with a distinguished element e which is the neutral element.

Definition (Inverse with respect to the operation)**.** Let $(M, \cdot, e)$ be a monoid and x, y elements of M. y is said to be an *inverse* of x if $x \cdot y = y \cdot x = e$. If x has an inverse, then x is said to be *invertible*.

Proposition (Inverse elements are unique)**.** *Let x be an element of a monoid $(M, \cdot, e)$ and y, w both inverses of x. Then $w = (x \cdot y) \cdot w = (y \cdot x) \cdot w = y \cdot (x \cdot w) = y$.*

Corollary. *Since inverse elements are unique, we may speak of* the *inverse of an element x.*

Definition (Group)**.** A (commutative) *group* is a (commutative) monoid $(G, \cdot, e)$ such that every element in G has an inverse.

Definition (Subgroup)**.** Let $(G, \cdot_G, e_G)$ be a group. A subset H of G is a *subgroup* of G if it satisfies the following three conditions

(i) H contains e_G;

(ii) H is closed under $\cdot_G$, that is for all h, h' in H, we have that $h \cdot h'$ lies in H;

(iii) and for every h in H, h^{-1} also lies in H.

Important consequences of this definition: If H is a subgroup of G, it is closed under $\cdot_G$, so we can restrict $\cdot_G \colon G \times G \to G$ to $\cdot_H \colon H \times H \to H; (h, h') \mapsto h \cdot_G h'$. Also, since e_G is the neutral element of $\cdot_G$ it is also the neutral element of our newly defined $\cdot_H$, so we write $e_H = e_G$. Consequently $(H, \cdot_H, e_H)$ is a group.

Definition (Ring, Non-unitary, Rng)**.** A (commutative) *ring* is a tuple $(R, +, 0, \cdot)$ so that $(R, +, 0)$ is a commutative group, $(R, \cdot)$ is a (commutative) semigroup and the operations $+, \cdot$ distribute as follows: for all a, b, c in R we have

$$a \cdot (b + c) = a \cdot b + b \cdot c \quad \text{and} \quad (a + b) \cdot c = a \cdot c + b \cdot c.$$

We call the commutative operation $+$ *addition* and the not necessarily commutative operation $\cdot$ *multiplication*. Often, we omit the $\cdot$ in multiplications and just write $a \cdot b = ab$.

Some authors describe also describe this as an *Non-unitary ring* or a *Rng* (“ring” without the multiplicative “i”dentity).

Remark. Note that the additive operation of a ring is always commutative. A non-commutative ring still has commutative addition, but the multiplication is not commutative. A good example is the set of $n \times n$ matrices.

Definition (Unitary Ring)**.** A (commutative) *unitary ring* is a tuple $(R, +, 0, \cdot, 1)$ so that $(R, +, 0, \cdot)$ is a (commutative) ring and $(R, \cdot, 1)$ is a (commutative) monoid.

Example (A Rng that is not unitary)**.** *The even integers constitute a rng, but not a unitary ring.*

Definition* (Module (Generalisation of a vector space))**.** Let $(R, +_R, 0_R, \cdot_R, 1)$ be a unitary ring. A *left R-module* is a tuple $(M, +_M, 0_M, \cdot_{M,l})$ for which $(M, +_M, 0_M)$ is a commutative group together with a binary operation $\cdot_{M,l} \colon R \times M \to M$ which we call *left scalar multiplication* that is associative and distributive. That is, for all r, s in R and x, y in M we have

$$\begin{aligned} r \cdot_{M,l} (s \cdot_{M,l} x) &= (r \cdot_R s) \cdot_{M,l} x \\ r \cdot_{M,l} (x +_M y) &= r \cdot_{M,l} x +_M r \cdot_{M,l} y \\ (r +_R s) \cdot_{M,l} x &= r \cdot_{M,l} x +_M s \cdot_{M,l} x \\ 1 \cdot_{M,l} x &= x. \end{aligned}$$

The multiplication $r \cdot_R s$ happens in the ring R. A *right R-module* is a also a commutative group $(M, +_M, 0_M)$ but with a binary operation $\cdot_{M,r} \colon M \times R \to M$ that is associative and distributive analogously. For a left module, multiplication from the left is permitted, for a right module multiplication from the right is permitted. If there is a multiplication from both sides, we call M simply an R-module, however we note that in general $s \cdot_{M,l} x \neq x \cdot_{M,r} s$.

Definition (Units of a ring)**.** An element u of a ring R is said to be a *unit* of R if u is invertible. That is, there exists a v in R such that $uv = vu = 1$. We denote the units of a ring by $R^\times$.

Definition (Field)**.** A *field* is a commutative unitary ring $(k, +, 0, \cdot, 1)$ for which one (and therefore all) of the following equivalent conditions is true

- $(k \setminus \{0\}, \cdot, 1)$ is a group.
- $k^\times = k \setminus \{0\}$.
- Every non-zero element in k is invertible.

The second condition is why some authors define $k^\times \overset{\text{def.}}{=} k \setminus \{0\}$.

Definition (Vector space)**.** Let k be a field. A *k-vector space* is a tuple $(V, +_V, 0_V, \cdot_V)$ such that $(V, +_V, 0_V)$ is a commutative group together with two associative binary operations $\cdot_l \colon k \times V \to V$ *scalar multiplication from the left* and $\cdot_r \colon V \times k \to V$ *scalar multiplication from the right* such that for all λ in k and v in V we have $\lambda \cdot_l v = v \cdot_r \lambda$. Due to symmetry, we call both of these operations *scalar multiplication* and simply denote them by $\cdot$. Moreover, this scalar multiplication is distributive. That is, the following conditions are satisfied for all λ, μ in k and v, u in V

$$\begin{aligned}\lambda \cdot (\mu \cdot v) &= (\lambda\mu) \cdot v\\ \lambda \cdot (v +_V u) &= \lambda \cdot v +_V \lambda \cdot u\\ (\lambda + \mu) \cdot v &= \lambda \cdot v +_V \mu \cdot v\\ 1 \cdot v &= v.\end{aligned}$$

In a shorter form, we say that $(V, +, 0, \cdot)$ is a vector space if $(V, +, 0, \cdot)$ is a k-module.

Definition (Linear subspace)**.** Let k be a field and $(V, +, 0, \cdot)$ a k-vector space. A subset U of V is a *linear subspace* (German *Untervektorraum*) of V if U is a subgroup of $(V, +, 0)$ and U is closed under scalar multiplication. The distributivity properties are given, since V is already a vector space. Consequently, U is also a vector space. Equivalently, U is a linear subspace of V if all of the following properties are satisfied

(i) 0_V lies in U;

(ii) for every u, v in U we have that $u + v$ lies in U;

(iii) for every λ in k and u in U, λu also lies in U.

3
Gaussian Elimination

(i) Row operations do not change the solutions of the system of equations.

(ii) Row operations do not represent base-change operations.

4

Vector Spaces

Example (Operations on linear subspaces).

(i) *The union of linear subspaces is not necessarily a linear subspace.*

(ii) *The intersection of linear subspaces is a linear subspace.*

(iii) *The sum $V + W = \{v + w \mid v \in V, w \in W\}$ of two sub vector-spaces is again a linear subspace.*

Proposition (Span)**.** *Let S be a subset of a k-vector space V. Let W be the intersection of all linear subspaces of V that contain S. Then*

$$W = \left\{ \sum_{i=1}^{n} \lambda_i s_i \mid n \in \mathbb{N}_0, \lambda_i \in k, s_i \in S \right\}.$$

We call W the span *of S and denote it* $\mathrm{Span}(S)$ *or* $\langle S \rangle$. *We also say that S* spans *or* generates *W or that S is a* spanning *or* generating *set of W. Since* $\mathrm{Span}(S)$ *is the intersection of linear subspaces it again is a linear subspace.*

Proposition. *A subset W of a vector space V is a linear subspace if and only if* $\mathrm{Span}(W) = W$.

5
Equivalences of linear (in)dependence

Proposition (Linear (in)dependence)**.** *Let* $(V, +, 0_V, \cdot)$ *be a* k*-vector space,* S *a subset thereof and* $W = \mathrm{Span}(S)$*. We say that* S *is* linearly independent *if one (and therefore all) of the following equivalent conditions holds:*

(i) There is no non-trivial linear combination of vectors in S *to form* 0_V*.*

(ii) No vector in S *can be written as a linear combination of other vectors in* S*.*

(iii) Every vector in W *can be uniquely written as a linear combination of vectors in* S*.*

(iv) S *is minimal amongst* W*-generating sets: that is, there is no proper subset* S' *of* S *so that* S' *generates* W*.*

Conversely, we say that S *is* linearly dependent *if one (and therefore all) of the following equivalent conditions holds:*

(i) There is a non-trivial linear combination of vectors in S *that forms* 0_V*.*

(ii) There is a vector s *in* S *that can be written as a linear combination of vectors in* $S \setminus \{s\}$*.*

(iii) There is a vector v *in* V *that can be written as two different linear combinations of vectors in* S*.*

(iv) There is a proper subset S' *of* S *that generates* W*.*

6
Bases

Definition (Basis)**.** Let V be a vector space. A subset B of V is called a *basis* of V if it is linearly independent and spans V.

It is conventional to say that the basis of the trivial vector space $\{0\}$ is the empty set. The only other natural candidate $\mathcal{B} = \{0\}$ does not work, because $\{0\}$ is not linearly independent as a set. This convention is also consistent with the convention that an empty sum equals 0. Indeed, we have seen that the span of any set $S \subseteq V$ is the set of elements formed by (finite) linear combinations of elements in S. If we try this with the empty set $S = \emptyset$, then we only obtain empty linear combinations (sums) which, by convention are 0.

A vector space may have many different bases. In general, we can simply multiply any basis vector by a non-zero scalar to get a new basis that is different. Of course this does not work for the trivial vector space $\{0\}$ (whose basis is the empty set) and a vector space over $\mathbb{F}_2$ (because the only-non zero scalar is 1, which would not change the basis). Indeed, let $\mathcal{B}$ be a basis of the k-vector space V. If V is non-trivial, then $\mathcal{B}$ is non-empty and we can pick a b in $\mathcal{B}$. Moreover, if $k \neq \mathbb{F}_2$, then there must be some non-zero scalar $\lambda \neq 1$ in k. Hence we can construct a new basis $\mathcal{B}' = \mathcal{B} \setminus (b) \cup \lambda b$. That is, we just replace the vector b in $\mathcal{B}$ with the scalar multiple λb.

A basis balances the trade-off between being large (so that they span the whole space), and small (so that they are still linearly independent).

Lemma (Steiniz Exchange Lemma)**.** *Let $\{v_1, \dots, v_n\}$ be a subset of a vector space V. If $w_1, \dots, w_m$ is a linearly independent subset of* $\mathrm{Span}(\{v_1, \dots, v_n\})$*, then $m \leq n$*

Corollary. *Let V be a vector space, that can be generated by a finite set of vectors. That is, there exists a finite set $S \subseteq V$ so that* $\mathrm{Span}(S) = V$*. Then any two bases of V have the same cardinality. We call this cardinality the* dimension *of V.*

Theorem 6.1. *Every vector space has a basis.*

Lemma (Dimension formula)**.** *Let V be a vector space and U, W subspaces thereof. Then*

$$\dim(U + V) = \dim(U) + \dim(V) - \dim(U \cap V)$$

Lemma 6.2 (Basis completion lemma)**.** *Every linearly independent subset of a vector space can be extended to a basis. That is, if $\{v_1, \dots, v_k\}$ is a set of linearly* *independent vectors in a vector space V, there exist vectors $v_{k+1}, \dots, v_n$ so that $\{v_1, \dots, v_n\}$ is a basis of V.*

Corollary. *Let V be a (finite dimensional) vector space of dimension n and U a sub vector space thereof.* $\dim(U) = \dim(V)$ *if and only if $U = V$.*

Corollary. *Let U be a linear subspace of a finite dimensional vector space V. Then* $\dim(U) \leq \dim(V)$*. Moreover,*

- *$U = V$ if and only if* $\dim(U) = \dim(V)$*.*
- *$U \subsetneq V$ if and only if* $\dim(U) < \dim(V)$*.*

In fact, these results give us an equivalent definition of what a basis is

Proposition 6.3. *Let V be a finite-dimensional vector space. We call a strictly inclusion-increasing sequence $0 \subsetneq U_1 \subsetneq U_2 \subseteq \cdots \subsetneq U_k \subseteq V$ of linear subspaces U_i a* chain*. The* length *of this chain is the index n. Then the supremum over the lengths of all possible chains in V is equal to* $\dim(V)$*.*

It is a nice exercise to prove this with the results we have gathered so far.

This definition is a little strange in the context of vector spaces, but is the most natural way the concept of dimension is defined for many other objects

(i) Let X be a topological space, we define its dimension $\dim(X)$ to be the supremum of the lengths of all possible chains $\emptyset \subsetneq X_1 \subsetneq X_2 \subsetneq \cdots \subsetneq X_n \subseteq X$ of irreducible[1] subspaces $X_i \subseteq X$.

(ii) Let R be a ring. We define the dimension $\dim(R)$ to be the supremum of the lengths of all possible chains $0 \subsetneq I_1 \subsetneq I_2 \subsetneq \cdots \subsetneq I_n \subseteq R$ of prime ideals[2] in R.

[1] A topological space is *irreducible* if it cannot be written as the union of two proper closed sets.

[2] We did not define this in this course, no worries.

7

Linear maps

Definition (Linear maps, Homomorphisms)**.** A map $T\colon V \to W$ of k-vector spaces V, W is said to be *linear* if it satisfies the additivity and homogeneity conditions. That is,

- (additivity) for every two vectors v, v' in V, we have that $T(v+v') = T(v)+T(v')$, and
- (homogeneity) for any vector v in V and any scalar λ in k, we have $T(\lambda v) = \lambda T(v)$.

The two conditions of additivity and homogeneity can be succinctly combined into one condition:

- T is linear if for all scalars λ in k and vectors v, v' in V we have $T(\lambda v + v') = \lambda T(v) + T(v')$

We denote the set of all linear maps $V \to W$ by $\mathrm{Hom}_k(V, W)$. Here, Hom is short for *homomorphism*, meaning to transform (morph) in a structure preserving (homo) way.

Lemma. *Linear maps are uniquely characterised by their behaviour on bases. That is, if $\mathcal{B}$ is a basis of V and $T, T'\colon V \to W$ linear maps such that $T(b) = T'(b)$ for all b in $\mathcal{B}$, then $T = T'$ on all of V.*

Proof. Let v be a vector in V. Then there exist (unique) coefficients $v_1, \dots, v_n$ in k and unique $b_1, \dots, b_n$ in $\mathcal{B}$ so that $v = v_1 b_1 + \dots + v_n b_n$. Then

$$\begin{aligned} T(v) &= T(v_1 b_1 + \dots + v_n b_n) \\ &= v_1 T(b_1) + \dots + v_n T(b_n) \\ &= v_1 T'(b_1) + \dots + v_n T'(b_n) \\ &= T'(v_1 b_1 + \dots + v_n b_n) \\ &= T'(v). \end{aligned}$$

Since the choice of v was arbitrary, we have shown that for all v in V, $T(v) = T'(v)$. Hence $T = T'$. □

Remark. This is a profoundly important result. It underpins the notion that linear maps preserve the structure of a vector space.

Definition (Kernel)**.** The *kernel* $\ker(T)$ of a linear map $T\colon V \to W$ is the preimage of 0_W under T. That is

$$\ker(T) = T^{-1}(\{0_W\}) = \{v \in V \mid T(v) = 0_W\} \subseteq V.$$

Lemma 7.1. *The kernel and image of a linear map $T\colon V \to W$ are subspaces of V and W respectively. In particular, if $\mathcal{B}$ is a basis of V, then $\mathrm{im}(T) = \mathrm{Span}(\{T(b) \mid b \in \mathcal{B}\})$.*

Proof. First we will show that the kernel is a subspace. By the homogeneity property of linearity, we know that for any vector v in V and λ in k, we have that $T(\lambda v) = \lambda T(v)$. In particular this is true when $\lambda = 0$ and we get $T(0_V) = T(0 \cdot v) = 0 \cdot T(v) = 0_W$. Hence 0_V lies in $\ker(T)$. Now let v, v' lie in $\ker(T)$, then $T(v+v') = T(v)+T(v') = 0_V+0_V = 0_V$, hence $v + v'$ lies in $\ker(T)$. Finally, we see that if v is in $\ker(T)$, then λv also lies in $\ker(T)$ for every λ in k. Indeed, $T(\lambda v) = \lambda T(v) = \lambda \cdot 0_V = 0_V$.

Now we will show that the image is a subspace of W. It suffices to show that for a choice of basis $\mathcal{B}$, we have that $\mathrm{im}(T) = \mathrm{Span}(\{T(b) \mid b \in \mathcal{B}\})$, since $\mathrm{Span}(\cdot)$ is always a vector (sub)space by definition (Proposition 4). To that end, let w lie in $\mathrm{im}(T)$. Then there exists some v in V such that $T(v) = w$. However, since $\mathcal{B}$ is a basis of V, we can write $v = v_1 b_1 + \cdots v_k b_k$ for unique coefficients v_i in k and vectors b_i in $\mathcal{B}$. Then $T(v) = T(v_1 b_1 + \cdots v_k b_k) = v_1 T(b_1) + \cdots + v_k T(b_k)$. This clearly lies in $\mathrm{Span}(\{T(b) \mid b \in \mathcal{B}\})$ and we have shown $\mathrm{im}(T) \subseteq \mathrm{Span}(\{T(b) \mid b \in \mathcal{B}\})$. Conversely, let w lie in $\mathrm{Span}(\{T(b) \mid b \in \mathcal{B}\})$. Then $w = w_1 T(c_1) + \cdots + w_k T(c_k)$ for a some c_i in $\mathcal{B}$ and scalars w_i in k. By linearity of T, we see that $w = T(w_1 c_1 + \cdots + w_k c_k)$. Thus w lies in $\mathrm{im}(T)$ and we have shown $\mathrm{Span}(\{T(b) \mid b \in \mathcal{B}\}) \subseteq \mathrm{im}(T)$. □

Lemma. *A linear map is injective if and only if it has a trivial kernel. That is, the map $T\colon V \to W$ is injective if and only if $\ker(T) = \{0_V\}$.*

Proof. Let $T\colon V \to W$ be an injective linear map and let v lie in its kernel. Then $T(v) = T(0_V) = 0_V$, and $v = 0_V$ by injectivity. Therefore the only element in the kernel of T is 0_V.

Conversely, suppose $\ker(T) = \{0_V\}$. If $T(v) = T(w)$, then by linearity $0 = T(w) - T(v) = T(w - v)$. Hence $w - v$ lies in $\ker(T) = \{0_V\}$, so $w - v = 0$ and $w = v$. We conclude that T is injective. □

Lemma 7.2. *Let V, W be k-vector spaces of arbitrary dimension and $T\colon V \to W$ linear. Then*

1. *If T is injective and $S \subseteq V$ linearly independent, then $T(S) = \{T(s) \mid s \in S\}$ is linearly independent.*

2. *If T is surjective and $S \subseteq V$ spans V ($\mathrm{Span}(S) = V$), then* $\mathrm{Span}(T(S)) = \mathrm{Span}(\{T(s) \mid s \in S\}) = W$.

We can succinctly summarise this result as follows

1. *Injective linear maps map linearly independent sets to linearly independent sets.*

2. *Surjective linear maps map spanning sets to spanning sets.*

Proof. Let $T\colon V \to W$ be an injective linear map and $S \subseteq V$ linearly independent. Consider a trivial linear combination $0_W = \lambda_1 T(s_i) + \cdots + \lambda_k T(s_k)$ of k elements in $T(S) = \{T(s) \mid s \in S\}$. Then $0_W = T(\lambda_1 s_1 + \cdots + \lambda_k s_k)$, so $\lambda_1 s_1 + \cdots + \lambda_k s_k$ lies in the kernel of T. Since T is injective, $\ker(T) = \{0_V\}$ and $0_V = \lambda_1 s_1 + \cdots + \lambda_k s_k$. However, by assumption S is linearly independent, so $\lambda_1 = \ldots = \lambda_k = 0$. We can conclude that $T(S)$ is linearly independent, since every trivial linear combination of vectors was shown to have coefficients $\lambda_1 = \ldots = \lambda_k = 0$.

The second statement with surjective T is an immediate consequence of Lemma 7.1. □

Definition (Rank, Nullity)**.** The dimension of the image of a linear map $T\colon V \to W$ is called the *rank* of T and denoted $\mathrm{Rank}(T) = \dim(\mathrm{im}(T))$. The dimension of the kernel of a linear map is called the *nullity* of T and denoted $\dim(\ker(T)) = \mathrm{Null}(T)$.

Lemma (Rangsatz)**.** *Let $T\colon V \to W$ be linear and V finite-dimensional then*

$$\mathrm{Rank}(T) + \mathrm{Null}(T) = \dim(V).$$

Proof. Let $\dim(V) = n$. We know that $\ker(T)$ is a (finite-dimensional) subspace of V (Lemma 7.1). Therefore it has a dimension k and a basis $b_1, \ldots, b_k$ (because it is a vector space, Lemma 6.1). By the basis completion lemma (Lemma 6.2) we can extend $b_1, \ldots, b_k$ to a basis $b_1, \ldots, b_k, b_{k+1}, \ldots, b_n$ of V. Now, by Lemma 7.1

$$\mathrm{Rank}(T) \overset{\text{def.}}{=} \dim(\mathrm{im}(T)) = \dim(\mathrm{Span}(\{T(b_1), \ldots, T(b_k), T(b_{k+1}), \ldots, T(b_n)\})).$$

Since $b_1, \ldots, b_k$ span the kernel, $T(b_1) = \ldots = T(b_k) = 0$ and

$$\mathrm{Rank}(T) = \dim(\mathrm{Span}(\{T(b_{k+1}), \ldots, T(b_n)\})).$$

However, by Lemma 7.2, we know that $\{T(b_{k+1}), \ldots, T(b_n)\}$ is linearly independent.

We conclude that $\mathrm{Rank}(T) + \mathrm{Null}(T) = (n-k) + k = n = \dim(V)$. □

Corollary 7.3. *Let $T\colon V \to W$ be injective linear. Then*

$$\dim(V) = \mathrm{Rank}(T) \leq \dim(W).$$

Corollary 7.4. *Let $T\colon V \to W$ be surjective linear. Then*

$$\dim(V) \geq \operatorname{Rank}(T) = \dim(W).$$

Corollary 7.5. *If $\dim(V) = \dim(W)$ and $T\colon V \to W$ linear. Then T is injective if and only if it is surjective.*

Proof. If T is surjective, then $\operatorname{Rank}(T) = \dim(W) = \dim(V)$. So $\operatorname{Null}(T) = 0$. This can only be the case if $\ker(T) = \{0\}$. Hence T is injective.

If T is injective, then $\operatorname{Null}(T) = 0$ and $\operatorname{Rank}(T) = \dim(V) = \dim(W)$. Therefore $\operatorname{im}(T) = W$ and T is surjective. □

8

Row Echelon Form

In this section, we highlight the value of the row-echelon form of a matrix.

Lemma. *Left multiplication by a matrix is a linear map. That is, if M is a $n \times m$ matrix over k, then the map $L_M \colon \mathbb{R}^m \to \mathbb{R}^n; x \mapsto Mx$ is linear.*

Lemma. *The elements of the kernel of L_M are exactly the solutions to the system of equations $Mx = 0$.*

Lemma 8.1 (Matrix Product Formula)**.** *Let A be a $m \times n$ matrix with entries in k, and B a $n \times k$ matrix with entries in k. The product AB is a $m \times k$ matrix. We have the following formula for the (i, j)-th entry of AB*

$$(AB)_{i,j} = \sum_{l=1}^{n} A_{i,l} B_{l,j}$$

In particular, if A is a $m \times n$ matrix, and $B = v$ an n-row column vector ($n \times 1$ matrix), then Av is a m-row column vector ($m \times 1$ matrix) whereby the i-th entry is

$$(Av)_i = \sum_{l=1}^{n} A_{i,l} v_l$$

Lemma 8.2. *The pivot elements of a matrix M in row-echelon form form a basis of the image of L_M.*

Proof. If M is a $n \times m$ matrix, we know that $L_M \colon \mathbb{R}^m \to \mathbb{R}^n; x \mapsto Mx$ is a linear map. If e_i is the i-th standard basis vector of R^m, then $L_m(e_i) = Me_i = M^{(i)}$ is the i-th column of M. Indeed, writing out the product using the matrix product formula (Lemma 8.1) we obtain

$$(Me_i)_j = \sum_{l=1}^{m} M_{j,l} (e_i)_l = M_{j,i} \qquad \text{so} \qquad Me_i = \begin{pmatrix} M_{1,i} \\ \vdots \\ M_{j,i} \\ \vdots \\ M_{n,i} \end{pmatrix} = M^{(i)}$$

Since the e_i generate $\mathbb{R}^m$, we know that Me_i generate $\operatorname{im}(T)$ (Lemma 7.2). Therefore, any linearly independent subset of $\{Me_i \mid e_i\}$ is a basis of the image of L_M. From the row-echelon form, it is clear that the pivot-columns are linearly independent. Therefore, the pivot columns of M generate the image of L_M. □

9

The quotient space

9.1 Equivalence relations

We will also be talking a lot about how different elements can be *represented* by different objects. The most elementary example of this is seen in the rational numbers. Both $1/2$ and $2/4$ represent the same quantity in $\mathbb{Q}$.

In fact, this can be generalised. Given any set X and an equivalence relation $\sim$ we can define $X/\sim = \left\{[x]_\sim \mid x \in X\right\}$. This is a new set where any two elements $[x]_\sim, [y]_\sim$ are said to be the same if (and only if) $x \sim y$. We call the elements $[x]_\sim$ in $X/\sim$ *(equivalence) classes* of $X/\sim$. So when $[x]_\sim = [y]_\sim$, we say that x, y *represent* the same class in $X/\sim$. For simpler notation we often drop the $\sim$ and just write $[x]$ for $[x]_\sim$.

As an example, take $X = \mathbb{Z}$, set n to be some fixed integer and define $x \sim y$ if and only if n divides $(x - y)$. We write this as $n \mid (x - y)$. This is an equivalence relation (check). For example $[a] = [a + n]$ for all a in $\mathbb{Z}$ because $a - (a + n) = -n$ is always divisible by n. So we would say that $[a]$ and $[a + n]$ represent the same class in $X/\sim$. We can now write down all the classes of $X/\sim$ as follows

$$X/\sim = \{[m] \mid m \in \mathbb{Z}\} = \{[0], [1], [2], \dots, [n-1]\}$$

We know that $X/\sim$ really only has these n elements because if we choose any $k > n$ then there exists some l in $\mathbb{Z}$ so that $0 \leq k' = k - ln \leq (n-1)$ and then $[k] = [k']$ is in our set above. Of course this is not the only way to write down all the classes in $X/\sim$. It is just as correct to say

$$\begin{aligned} X/\sim &= \{[1], [2], \dots, [n-1], [n]\} \\ &\overset{\text{or}}{=} \{[n+1], [2+n], \dots, [2n-1], [2n]\} \\ &\overset{\text{or}}{=} \{[-3], [-2], [-1], [0], , \dots, [n-5], [n-4]\} \end{aligned}$$

What is special about our $X = \mathbb{Z}$ example is that $\mathbb{Z}$ also has a group structure: addition. This carries over to $X/\sim$, or, in more formal language, $X/\sim$ *inherits* the group structure of $\mathbb{Z}$.

Before we show this, we go back to our example of $\mathbb{Q}$. Here the equivalence relation is given by $a/b \sim a'/b' \iff ab' - a'b = 0$. If we define any binary operation on $\mathbb{Q}$ we must ensure it does not depend on the representatives. For example, suppose we try to define $a/b \oplus a'/b' = a + a'$. This is *not* well-defined. Indeed, take $1/2$ and $2/4$ as before. Then

$$1/2 \oplus 1/2 = 1 + 1 = 2 \quad \text{but} \quad 2/4 \oplus 2/4 = 2 + 2 = 4.$$

This is bad because $1/2$ and $2/4$ are supposed to be the same element in $\mathbb{Q}$ so adding them to themselves should yield the same result. Using the correct definition $a/b + a'/b' = (ab' + a'b)/bb'$ yields an operation that does not depend on the representatives. Indeed, suppose $c/d = a/b$ and $c'/d' = a'/b'$, then by definition $cb - ad = 0$ and $c'b' - a'd' = 0$. Calculating

$$\frac{c}{d} + \frac{c'}{d'} = \frac{cd' + dc'}{dd'} = \frac{bb'cd' + bb'dc'}{bb'dd'} = \frac{b'd'ad + bda'd'}{bb'dd'} = \frac{b'a + ba'}{bb'} = \frac{a}{b} + \frac{a'}{b'}$$

we see that the addition is well-defined.

Returning to our $\mathbb{Z} = X$ example we define define $[a] \oplus [b] = [a + b]$. To verify that this is well-defined, we take two different representatives $[a'], [b']$ of $[a], [b]$ and calculate

$$[a] \oplus [b] = [a + b] = [a' + b'] = [a'] \oplus [b'].$$

Here, $[a + b] = [a' + b']$ because $[a] = [a']$ and $[b] = [b']$, So we know $n \mid (a - a')$ and $n \mid (b - b')$ and thus $n \mid (a - a') + (b - b') = (a + b) - (a' + b')$.

The attentive reader will have noticed that this example is exactly that of modular arithmetic. Here we often use the notation "$x \pmod{n}$" to mean the class of x modulo n in $\mathbb{Z}$.

I would recommend really understanding this example thoroughly before moving on. It might also help to see the general construction of a quotient of commutative group.

Construction 9.1 (Quotient group). *Let $(G, +, 0)$ be a commutative group and H a subgroup. That is, 0 is in H; for all h, h' in H, we have that $h + h'$ is again in H; and that for every element h in H, $-h$ is also in H. Define the equivalence relation $g \sim g' \iff (g - g') \in H$, and write as a set $G/_H = \{[g] \mid g \in G\}$. The inherited group structure on $G/_H$ is defined by $[g] \oplus [g'] = [g + g']$.*

If one wants to generalise this to generic (non-commutative) groups, one needs to introduce the notion of a *normal* subgroup. After this, the construction is exactly the same.

9.2 Construction of the quotient space

We have spoken about algebraic and geometric aspects of vector spaces. Initially we defined a vector space as an abstract algebraic object, but have seen that with the introduction of the scalar product that (inner product) vector spaces can encapsulate geometric notions.

For now, forget all geometric notions and go back to the plain algebraic definition of a vector space: a vector space V over a field k is a tuple $(V, +, 0, \cdot)$ so that $(V, +, 0)$ is a commutative group together with a scalar multiplication $\cdot$ which distributes with $+$.

Now we can construct the quotient of a vector space and a linear subspace.

Construction 9.2 (Quotient vector space)**.** *Let $(V, +, 0, \cdot)$ be a vector space and U a linear subspace of V.*

The set. *For every vector v in V we can write down the symbol $[v]_U$. We call this a symbol because at the moment it is only a* notation *and has no mathematical meaning. We could also write $Q_v^{\mathcal{U}}$ for every v in V instead. We call the set of all of these symbols*

$$V/U = \left\{ [v]_U \mid v \in V \right\}.$$

Now we define the equivalence relation $[v]_U \sim [w]_U \iff (v - w) \in U$. This is reflexive, since $v - v = 0$ lies in every linear subspace. It is symmetric because every linear subspace is closed under negation (taking the additive inverse). That is, if $v - w$ is in U, then $w - v = -(v - w)$ also lies in U. Finally, it is transitive, because if $v - w$ and $w - z$ lie in U, then $v - z = (v - w) + (w - z)$ lies in U.

The additive group structure. *Now we an addition on V/U as follows $[v]_U \oplus [v']_U \overset{\text{def.}}{=} [v + v']_U$. This is well-defined. Indeed, take two representatives $[w]_U, [w']_U$ of $[v]_U, [v']_U$ respectively and add*

$$[v]_U \oplus [v']_U = [v + v']_U = [w + w']_U = [w]_U \oplus [w']_U.$$

(As in the example of modular arithmetic) we see that $[v + v']_U = [w + w']_U$ because $(v + v') - (w + w') = (v - w) + (v' - w')$ lies in U.

We claim that $[0]_U$ is the neutral element of this group with respect to $\oplus$. Indeed, for all $[v]_U$ in V/U we have $[v]_U + [0]_U = [v + 0]_U = [v]_U$. Every element also has an inverse with respect to $\oplus$. For all $[v]_U$ in V/U we take $[-v]_U$ and add $[v]_U + [-v]_U = [v - v]_U = [0]_U$.

So we have a group $(V/U, \oplus, [0]_U)$. To obtain a vector space be must also define a scalar multiplication. We do this in the obvious way. For every λ in k and $[v]_U$ in V/U, we define $\lambda \odot [v]_U = [\lambda v]_U$. We must now ensure that this distributes with

$\oplus$ in the correct way. Let $[v]_U, [w]_U$ lie in V/U and λ, μ in k, then

$$\begin{aligned}(\lambda+\mu)\odot[v]_U &= [(\lambda+\mu)\cdot v]_U\\ &= [\lambda\cdot v+\mu\cdot v]_U\\ &= [\lambda\cdot v]_U\oplus[\mu\cdot v]_U\\ &= [\lambda\cdot v]_U\oplus[\mu\cdot v]_U\\ &= \lambda\odot[v]_U\oplus\mu\odot[v]_U\end{aligned}$$

moreover

$$\begin{aligned}\lambda\odot([v]_U\oplus[w]_U) &= \lambda\odot[v+w]_U\\ &= [\lambda\cdot(v+w)]_U\\ &= [\lambda\cdot v+\lambda\cdot w)]_U\\ &= [\lambda\cdot v]_U\oplus[\lambda\cdot w]_U\\ &= \lambda\odot[v]_U\oplus\lambda\odot[w]_U.\end{aligned}$$

In other words, all of the properties are simply inherited from V.

So we conclude that $(V/U, \oplus, [0]_U, \odot)$ forms a vector space. For notational simplicity we will only write $(V/U, +, 0, \cdot)$. When the subspace U is clear, we may write $\overline{v}$ instead of $[v]_U$.

Finally, we note the map $V \to V/U; v \mapsto [v]_U$ is called the quotient map. *It is linear, surjective and usually denoted by π_U or π when the subspace is clear. Notably, the kernel of π is exactly U. Indeed, suppose $\pi(v) = [0]_U$ then $v - 0 = v$ lies in U, so $\ker(\pi) \subseteq U$. Conversely, for every element in u in U, the element $u - 0 = u$ is also in U (no surprise), therefore $\pi(u) = [u]_U = [0]_U$ and we conclude that $U \subseteq \ker(\pi)$.*

Remark (A remark on notation). Another popular notation for $[v]_U$ is $v + U$. This is inspired by the fact that $\pi^{-1}(\overline{v}) = \{v + u \mid u \in U\} \overset{\text{notation}}{=} v + U$.

Altogether, we have the following notations for elements in V/U

$$\overline{v} = \pi_U(v) = \pi(v) = [v]_U = [v] = v + U$$

I will mostly use $\overline{v}$ or $\pi(v)$ depending on the context, and use the subscript where there is ambiguity.

There are two trivial examples of quotients. Firstly, quotienting by the zero-space $\{0\}$. Then we claim that $V \cong V/\{0\}$ via $\pi\colon v \mapsto \overline{v}$. We know that π is surjective. So let us verify injectivity. Suppose $\overline{v} = \overline{w}$ then $v - w$ is in $\{0\}$ so $v = w$. The second trivial example is quotienting by the entire space. We claim that $V/V \cong \{0\}$. Here it is enough to count the number of elements of $V/_V$. If it only has one element, it must be 0 because $V/_U$ is always a vector space. Take two elements $\overline{v}, \overline{w}$ in $V/_V$. Then $\overline{v} = \overline{w}$ because $v - w$ is in V.

Lemma 9.3 (Bases of quotient spaces (Exercise 5.4.9(a)))**.** *Let V be a n-dimensional vector space and U a subspace. Let $\mathcal{U} = (b_1, \dots, b_k)$ be a basis of U and $\mathcal{B} = (b_1, \dots, b_k, b_{k+1}, \dots, b_n)$ an extension of this basis to a basis of V. Then $\mathcal{Q} = \left([b_{k+1}]_U, \dots, [b_n]_U\right)$ is a basis of the quotient $V/_U$.*

Proof. First we note that $[b_1]_U = \cdots = [b_k]_U = [0]_U$ since b_i are in U for $i = 1, \dots, k$.

Now let $[v]_U$ be an element of the quotient $V/_U$. Since $\mathcal{B}$ is a basis, we may write $v = \alpha_1 b_1 + \cdots + \alpha_n b_n$ for some scalars α_i in k. Then

$$\begin{aligned}
[v]_U &= [\alpha_1 b_1 + \cdots + \alpha_n b_n]_U \\
&= \alpha_1 [b_1]_U + \cdots + \alpha_n [b_n]_U \\
&= \underbrace{\alpha_1 [b_1]_U + \cdots + \alpha_k [b_k]_U}_{=[0]_U} + \alpha_{k+1}[b_{k+1}]_U + \cdots + \alpha_n [b_n]_U \\
&= \alpha_{k+1}[b_{k+1}]_U + \cdots + \alpha_n [b_n]_U
\end{aligned}$$

therefore $\mathcal{Q}$ surely spans $V/_U$.

Now suppose

$$[0]_U = \beta_{k+1}[b_{k+1}]_U + \cdots + \beta_n [b_n]_U = [\beta_{k+1} b_{k+1} + \cdots + \beta_n b_n]_U$$

for some scalars β_i in k. Then $u = \beta_{k+1} b_{k+1} + \cdots + \beta_n b_n$ lies in U. This can only be if $u = 0$ since $b_{k+1}, \dots, b_n$ form a basis of the complement $U^\perp$ of U. Another way of seeing this, is that u can be written as $u = \beta_1 b_1 + \cdots + \beta_k b_k$. Then

$$0 = \underbrace{\beta_1 b_1 + \cdots + \beta_k b_k}_{=u} - \underbrace{(\beta_{k+1} b_{k+1} + \cdots + \beta_n b_n)}_{\text{also } =u}$$

and since the b_i are linearly independent, all $\beta_i = 0$. □

Example 9.4 (Geometry of the quotient space)**.** *We can now think about the geometry of a quotient space. The common go-to example here is to consider $\mathbb{R}^2$ and a 1-dimensional subspace U. This 1-dimensional subspace geometrically is a line through the origin. If we now form the quotient $\mathbb{R}^2/_U$ and take an element $\overline{v}$ from there. We note that by definition all representatives of $\overline{v}$ are vectors w in $\mathbb{R}^2$ so that $w - v$ is in U. In other words, there exists a u in U so that $w = u + v$. Consequently the set of representatives for $\overline{v}$ are all the vectors in the line (not necessarily through the origin) parallel to U that goes through v.*

There are three further interesting results on quotient spaces.

Lemma 9.5. *There is a bijection between the linear subspaces of V that contain U and the linear subspaces of ${}^{V}\!/_{U}$. This bijection is induced by π. That is*

$$\varPi\colon \{W \subseteq V \text{ a linear subspace} \mid U \subseteq W \subseteq V\} \xrightarrow{\sim} \left\{\overline{W} \subseteq {}^{V}\!/_{U} \text{ a linear subspace}\right\}$$
$$W \mapsto \pi(W)$$

is a bijection

Proof. Note that these sets are not vector spaces, so we firstly did not say that they are isomorphic, and secondly we did not claim that $\varPi$ is linear — although π itself is linear.

Also note that $\varPi(W) = \pi(W)$ for every subspace W of V that contains U. This is a syntactic distinction because π is a map $V \mapsto {}^{V}\!/_{U}$ and $\varPi$ is as above.

We must only verify bijectivity. Suppose $\pi(W) = \pi(W')$ for two subspaces W of V and let w lie in W. Then there exists a w' in W' so that $\pi(w) = \pi(w')$. In other words there exists a u in U so that $w - w' = u$. Since $U \subseteq W'$ we see that $w = w' + u$ lies in W' and we have shown that $W \subseteq W'$. By symmetry we see that $W' \subseteq W$ and $W = W'$ so $\varPi$ is injective. Conversely, suppose $\overline{W}$ is a subspace of ${}^{V}\!/_{U}$, then $\pi^{-1}(\overline{W})$ is also a subspace of V (the preimage of a linear subspace under a linear map is again a linear subspace). Further, since $\overline{0}$ lies in $\overline{W}$ we see that $U \subseteq \pi^{-1}(\overline{W})$. Therefore there exists a linear subspace $\pi^{-1}(\overline{W}) \subseteq V$ which contains U and is mapped to $\overline{W}$ under π.

We conclude that $\varPi$ is indeed bijective. □

Corollary 9.6 (of the proof.)**.** *Let U be a linear subspace of V and W, W' linear subspaces that both contain U. Then $W \subseteq W'$ if and only if $\pi_U(W) \subseteq \pi_U(W')$.*

Proof. That π is inclusion preserving is clear. This is the direction $W \subseteq W'$ implies $\pi_U(W) \subseteq \pi_U(W')$. This is true of all maps. The converse was seen in the proof. Namely that when $\pi_U(W) \subseteq \pi_U(W')$ we know that for all w in W, there exists a u in U so that $w = w' + u$ and since $U \subseteq W'$ we see that w lies in W'. □

Theorem 9.7 (Isomorphism theorem of vector spaces)**.** *Let $f\colon V \to W$ be a linear map. Then*

$$\overline{f}\colon {}^{V}\!/_{\ker(f)} \xrightarrow{\sim} \operatorname{im}(f); \quad \overline{v} \mapsto f(v)$$

is a linear isomorphism.

This theorem is used everywhere in algebra. The idea behind both the statement and the proof can be thought of in two steps. First you restrict $f\colon V \to W$ to the image, $\tilde{f}\colon V \to \operatorname{im}(f)$. This map is clearly surjective by construction. If we now also quotient out the kernel $\overline{f}\colon V/_{\ker(f)} \to \operatorname{im}(f)$, we are removing any non-trivial vectors which could be sent to 0, making the map injective.

Proof. First we note that $\ker(f)$ is always a linear subspace, so $V/_{\ker(f)}$ is a well-defined object.

Secondly we must verify that this is a well-defined map. Indeed, suppose v, w are two representatives of the same class in $V/_{\ker(f)}$, that is $\overline{v} = \overline{w}$. Then $f(v) - f(w) = f(v - w) = 0$, since $v - w$ is in $\ker(f)$ because they represent the same class. Therefore $\overline{f}(\overline{w}) = f(w) = f(v) = \overline{f}(\overline{v})$.

Thirdly, we must check that $\overline{f}$ is linear. Indeed, for all v, w in V and λ, μ in k we have $f(\lambda\overline{v} + \mu\overline{w}) = f(\lambda v + \mu w) = \lambda f(v) + \mu f(w) = \lambda f(\overline{v}) + \mu f(\overline{w})$.

Finally, we must verify that $\overline{f}$ is bijective. It is clearly surjective: if z is in $\operatorname{im}(f)$, then there exists a v in V so that $f(v) = z$, therefore $\overline{f}(\overline{v}) = f(v) = z$. For injectivity, suppose $\overline{f}(\overline{v}) = \overline{f}(\overline{w})$. Then $f(v) - f(w) = 0$ so $v - w$ is in $\ker(f)$ and $\overline{v} = \overline{w}$. □

Finally, we have the final interesting result on quotient spaces. Namely

Lemma 9.8. *Let V be a finite dimensional vector space. Let U be a linear subspace. Then*

$$\pi|_{U^\perp}\colon U^\perp \xrightarrow{\sim} V/_U; \quad v^\perp \mapsto \pi(v^\perp) = \overline{v^\perp}$$

is an isomorphism. Equivalently, $\dim\left(V/_U\right) = \dim(V) - \dim(U)$.

Note that $U^\perp$ has nothing to do with orthogonality in this case, it is simply the linear complement of U.

Direct proof. This map is a restriction of $\pi\colon V \to V/_U$ to $\pi|_{U^\perp}\colon U^\perp \to V/_U$. Therefore we know that it is linear. We simply prove that this map is bijective.

For injectivity, suppose $\pi|_{U^\perp}(v) = \pi|_{U^\perp}(v')$, then $v - v'$ lies in U. However, since v, v' also lie in $U^\perp$ we see that $v - v'$ is also in $U^\perp$ and $v - v' = 0$ because $U \cap U^\perp = \{0\}$. So $v = v'$ and $\pi|_{U^\perp}$ is injective. For surjectivity, recall that π is surjective. So for every $\overline{v}$ in $V/_U$, there exists a v in V so that $\pi(v) = \overline{v}$ (this is almost tautological, because we define $\overline{v} = \pi(v)$). Moreover, by the decomposition $V = U \oplus U^\perp$ we know that $v = u + u^\perp$ for some u in U and $u^\perp$ in $U^\perp$. Then $\overline{v} = \pi(v) = \pi(u) + \pi(u^\perp) = \pi(u^\perp) = \pi|_{U^\perp}(u^\perp)$. Therefore $\pi|_{U^\perp}$ is surjective. □

9.3 Induced maps on quotient spaces

Theorem 9.9 (General result on induced maps)**.** *Let $f\colon V \to W$ be a linear map, and $U \subseteq V$ a linear subspace. Then f induces a well-defined map*

$$\overline{f}\colon V/U \to W/f(U); \quad [v]_U \mapsto [f(v)]_{f(U)}$$

such that the following diagram

$$\begin{tikzcd}
V \arrow[r, "f"] \arrow[d, "\pi_U"'] & W \arrow[d, "\pi_{f(U)}"] \\
V/U \arrow[r, dashed, "\exists! \overline{f}"'] & W/f(U)
\end{tikzcd}$$

commutes. In other words, $\overline{f} \circ \pi_U = \pi_{f(U)} \circ f$.

Here we have used the $[v]_U$ and $[f(v)]_{f(U)}$ notation because we are dealing with quotients of two different linear subspaces $U, f(U)$. Of course it should be clear from context, what is meant when $\overline{v}$ and $\overline{f(v)}$ is written, but this illustrates that using the bracket-notation can be helpful.

Proof. First we note that $\operatorname{im}(f) = f(U)$ is a linear subspace of W and so the quotient $W/f(U)$ is well-defined.

For well-definedness, we must verify that two representatives of the same class $[v]_U = [v']_U$ are sent to the same element. We quickly recall that $[v]_U = [v']_U$ is true if and only if $v - v'$ is in U. Now calculate $[f(v)]_{f(U)} - [f(v')]_{f(U)} = [f(v) - f(v')]_{f(U)} = [f(v - v')]_{f(U)} = [0]_{f(U)}$. The first equality is true by definition of addition (subtraction) in a quotient vector space; the second is owing to linearity of f; and the third is because $v - v'$ is in U, so $f(v - v')$ is in $f(U)$ because U is f-invariant. □

We see this is a generalisation of our isomorphism theorem. We will not often use this result in its full generality, but restrict ourselves to endomorphisms.

Corollary 9.10. *Let f be an endomorphism of V and U a linear subspace. Then f induces a well-defined map $\overline{f}\colon V/U \to V/f(U)$.*

In fact, we will restrict ourselves to endomorphisms and invariant subspaces.

Corollary 9.11 (Induced endomorphisms)**.** *Let f be an endomorphism of V and U an f-invariant linear subspace. Then f induces a well-defined endomorphism $\overline{f}$ on*

V/U via $\overline{v} \mapsto \overline{f(v)}$. Note that in this case $\overline{f} \circ \pi_U = \pi_U \circ f$. Moreover, $\overline{f}^k = \overline{f^k}$ for all $k \geq 1$.

Proof. The proof is exactly the same. Take two representatives $\overline{v} = \overline{v'}$ of the same class. Then $\overline{f(v)} - \overline{f(v')} = \overline{f(v) - f(v')} = \overline{f(v - v')} = \overline{0}$. The difference is that we know that $f(v - v')$ lies in U because $v - v'$ lies in U and U is f-invariant.

Finally, to show that $\overline{f}^k = \overline{f^k}$, let $\overline{v}$ lie in V/U. Then

$$\overline{f^k}(\overline{v}) \overset{\text{def.}}{=} \overline{f^k(v)} = (\pi \circ f^k)(v) = (\pi \circ f \circ f^{k-1})(v) = (\overline{f} \circ \pi \circ f^{k-1})(v) = \overline{f}^k \circ \pi(v) = \overline{f}^k(\overline{v}).$$

□

Corollary 9.12 (Representations of induced endomorphisms (Exercise 5.4.9(b))). *Let V be a vector space, T an endomorphism of V and U a T-invariant subspace. Let $\mathcal{U} = (b_1, \dots, b_k)$ be a basis of U and $\mathcal{B} = (b_1, \dots, b_k, b_{k+1}, \dots, b_n)$ an extension of this basis to V. Then*

$$[T]_{\mathcal{B}} = \left(\begin{array}{c|c} [T|_U]_{\mathcal{U}} & * \\ \hline 0 & \left[T_{V/U}\right]_{\mathcal{Q}} \end{array}\right)$$

where $\mathcal{Q} = (\overline{b_{k+1}}, \dots, \overline{b_n})$ is a basis of V/U and $T_{V/U}$ is the T-induced endomorphism on V/U. (Here we are using the notation $T_{V/U}$ instead of $\overline{T}$ because the notation got a little ugly).

Proof. First we recall that for any linear map $L\colon A \to B$ and basis $\mathcal{S} = (v_1, \dots, v_n)$ of V, the matrix $[L]_{\mathcal{S}}$ is the unique matrix so that $Lv_i = \sum_{j=1}^n ([L]_{\mathcal{S}})_{ij} v_j$.

We note that $[T]_{\mathcal{B}}$ can be written as a block-matrix

$$[T]_{\mathcal{B}} = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$$

with A in $\mathrm{Mat}_{k \times k}(k)$, B in $\mathrm{Mat}_{k \times (n-k)}(k)$. We now compute for $i = 1, \dots, k$

$$T(b_i) = \sum_{j=1}^{k} A_{ij} b_j + \sum_{j=k+1}^{n} C_{ij} b_j$$

however, since U is T-invariant, and $\mathcal{U} = (b_1, \dots, b_k)$ is a basis of U, we know that $C_{ij} = 0$.

We continue computing for $i = (k+1), \dots, n$

$$\begin{aligned}\sum_{j=k+1}^{n} ([T_{V/U}]_Q)_{ij}\overline{b_j} &= T_{V/U}(\overline{b_i}) \\ &= \overline{T(b_i)} \\ &= \overline{\sum_{j=1}^{k} B_{ij}b_j + \sum_{j=k+1}^{n} D_{ij}b_j} \\ &= \sum_{j=1}^{k} B_{ij} \underbrace{\overline{b_j}}_{=\overline{0} \text{ in } V/U} + \sum_{j=k+1}^{n} D_{ij}\overline{b_j} \\ &= \sum_{j=k+1}^{n} D_{ij}\overline{b_j}.\end{aligned}$$

This is true for all i, j therefore $[T_{V/U}]_Q = D$. □

10

Representations of endomorphisms and their invariant subspaces

We recall that an *endomorphism* is a linear map on a vector space to itself. That is, if V is a vector space, the set of all linear maps $T\colon V \to V$ is denoted $\operatorname{End}(V)$ and the maps are called endomorphisms. Endomorphisms that are also isomorphisms are called *automorphisms*. The set of all automorphisms of a vector space V is denoted $\operatorname{Aut}(V)$.

Since linear maps are uniquely defined by their behaviour on bases, and a linear map is an isomorphism if and only if it sends a basis to another basis, we see a very simple way to construct some automorphisms of a vector space V: we pick an ordered basis $(b_1, b_2, b_3, \dots)$ of V and simply send this to a permutation $(b_{\sigma(1)}, b_{\sigma(2)}, b_{\sigma(3)}, \dots)$. This construction works for infinite-dimensional spaces, but is a little easier to conceptualise for finite-dimensional spaces.

For example, the reflection along the $x = y$ axis in $\mathbb{R}^2$ is simply the map that permutes the x-axis and the y-axis. More formally, the linear map

$$R\colon \mathbb{R}^2 \to \mathbb{R}^2; \quad (x, y) \mapsto (y, x)$$

is the reflection at the $x = y$ axis and is characterised by sending the ordered basis $((1, 0), (0, 1))$ to $((0, 1), (1, 0))$. This shows that for an n-dimensional k-vector space V, there must be at least $|S_n| = n!$ automorphisms. However, in reality there are many more. Consider scaling by a λ in k. Clearly the scale-by-λ map $\lambda \operatorname{id}_V$ is an automorphism, but it does not permute basis vectors.

An example of a non-automorphism endomorphism is a *projection*. An endomorphism P of V is called a projection if $P \circ P = P$. More generally, any map $f\colon X \to X$ which satisfies $f \circ f = f$ is called *idempotent* (does not need to be linear). This is not to be confused with an *nilpotent* endomorphism, which is a linear map N that satisfies $N^k = N \circ \dots \circ N = 0$ for some $k > 0$.

The example that we already know is the projection onto some coordinates. For example $X\colon \mathbb{R}^2 \to \mathbb{R}^2; (x, y) \mapsto (x, 0)$. This is the projection onto the x-axis. This can be generalised to any linear subspace. Let V be a vector space and U a linear subspace. Let $b_1, \dots, b_k$ be a basis of U, and $b_{k+1}, \dots, b_n$ vectors to complete this to a basis $\mathcal{B}$ of V. Just like in the x-axis projection example, we want to send

$$(v_1, \dots, v_k, v_{k_1}, \dots, v_n) \mapsto (v_1, \dots, v_k, 0, \dots, 0)$$

of course the v_i need to be the coordinates in the basis $(b_1, \dots, b_n)$. So in other words the projection P_U onto U can be written as

$$P_U = [\mathrm{id}]^{\mathcal{E}_n}_{\mathcal{B}} \begin{pmatrix} 1_{k\times k} & 0 \\ 0 & 0_{(n-k)\times(n-k)} \end{pmatrix} [\mathrm{id}]^{\mathcal{B}}_{\mathcal{E}_n}$$

Another way of writing this, is to decompose $V = U \oplus U^{\perp}$ (recall, we do not need a scalar product to do this), and then define $P_U|_U = \mathrm{id}\,|_U$ and $P_U|_{U^\perp} = 0$.

Now, when given an endomorphism T it is very natural to ask oneself which "parts" of V are invariant under T. Of course, the "parts" we are looking for are linear subspaces. We saw in the projection example, that U is invariant under P_U. In fact, here $P_U|_U = \mathrm{id}\,|_U$. This is a much stronger assertion than just invariance. Recall that for a subspace U of V, we say that U is *T-invariant* if $T(U) \subseteq U$.

In general, for any endomorphism T in $\mathrm{End}(V)$, we already know of two invariant subspaces. Namely $\ker(T)$ and $\mathrm{im}(T)$.

A further natural object of study is set of vectors on which under a given linear map. More precisely, if $T\colon V \to V$ is an endomorphism of the vector space V, we can study the set of vectors for which $T(v) = v$. Another perspective is that before, we looked at linear subspaces that are invariant $T(U) \subseteq U$, now we are looking at vectors v for which $T(\mathrm{Span}(v)) \subseteq \mathrm{Span}(v)$.

Both of these cases can be generalised by studying *eigenspaces* of a vector space. For a k-vector space V, T an endomorphism of V, and λ a scalar in k, we define an eigenspace by

$$\mathrm{Eig}_T(\lambda) = \ker(T - \lambda\,\mathrm{id}_V)$$

We see that this is a generalisation of the previous two cases. The kernel of an endomorphism T is the eigenspace for the scalar $\lambda = 0$. The set of vectors invariant under T is the eigenspace for the scalar $\lambda = 1$. Indeed, if v lies in $\mathrm{Eig}_1(T) = \ker(T - \mathrm{id}_V)$, then $(T - \mathrm{id}_V)(v) = T(v) - \mathrm{id}_V(v) = T(v) - v = 0$, so $T(v) = v$.

Proposition 10.1. *Let T be an endomorphism of the finite dimensional vector space V. There exists a basis $\mathcal{B}$ of V so that $[T]_{\mathcal{B}}$ is a block-diagonal matrix if and only if there exist T-invariant subspaces V_i of V so that $V = V_1 \oplus \cdots \oplus V_k$. If $\mathcal{B}_i \subseteq \mathcal{B}$ is a basis of V_i, then the i-th block in the block-diagonal representation $[T]_{\mathcal{B}}$ is exactly $[T|_{V_i}]_{\mathcal{B}_i}$.*

Let T be an endomorphism of V. We call a linear subspace U of V *T-indecomposable* if any decomposition $U = U_1 \oplus U_2$ of T-invariant linear subspaces U_1, U_2 implies that either $U_1 = \{0\}$ or $U_2 = \{0\}$.

Lemma 10.2. *Let T be an endomorphism of V. There exist a T-indecomposable* *linear subspaces* V_i *such that* $V = V_1 \oplus \cdots \oplus V_k$. *We call this a* T-indecomposable decomposition *of* V.

11
Factorising polynomials

A degree n polynomial $p(X)$ in $k[X]$ is said to *split into linear factors* if it can be written as the product $p(X) = (X - r_1) \cdots (X - r_n)$ where the r_i lie in k. Clearly the r_i are roots of p. When two roots coincide $r_i = r_j$ we may group these together and rewrite the product as

$$p(X) = (X - r_1) \cdots (X - r_n) = \prod_{i=1}^{r} (X - r_i)^{m_1}$$

where r is the number of distinct roots. In this form, we call m_i the *order* or *arithmetic multiplicity* of the root r_i.

Example 11.1. *Not all polynomials split into linear factors. Indeed, $p(X) = X^2+1$ in $\mathbb{R}[X]$ cannot be written as $p(X) = (X - \alpha)(X - \beta)$ with α, β in $\mathbb{R}$. This can be proven in many ways. Here is one way.*

Suppose α, β exist in $\mathbb{R}$ so that $p(X) = X^2+1 = (X-\alpha)(X-\beta) = X^2-(\alpha+\beta)X+1$. Then by comparing coefficients, $\alpha\beta = 1$ and $\alpha + \beta = 0$ must hold true. Clearly $\alpha = 0$ cannot be true then, so let us assume that $\alpha > 0$ without loss of generality. Then $\beta < 0$ since $\alpha + \beta = 0$. However, then $\alpha\beta = 1$ cannot hold, since $\alpha\beta < 0$.

It is of note, that polynomials over $\mathbb{C}$ (that is, polynomials in $\mathbb{C}[X]$) always split into linear factors.

We call fields k for which all polynomials $p(X)$ in $k[X]$ split into linear factors *algebraically closed*. That is to say, $\mathbb{C}$ is an algebraically closed field.

11.1 A short detour on algebraically closed fields

The notions underlying algebraically closed fields is explained in the following. This section is additional content and not super relevant to our course. However, it aims to explain why we call a field "algebraically closed" and give examples.

Let L be a field, and $k \subseteq L$ a subset thereof. If k is also a field under the addition and multiplication of L, then k is said to be a *subfield* of L and L is said to be a *field extension* of k.

Examples of this are $\mathbb{Q} \subseteq \mathbb{R}$, or $\mathbb{R} \subseteq \mathbb{C}$ or $\mathbb{Q} \subseteq \mathbb{C}$. Although we often think of $F_p = \mathbb{Z}/p\mathbb{Z}$ as $\{0, 1, \dots, (p-1)\}$ we note that F_p is *not* a subfield of $\mathbb{Q}$. Indeed, in $\mathbb{Q}$ we know that $1 +_{\mathbb{Q}} 1 +_{\mathbb{Q}} 1 +_{\mathbb{Q}} \cdots +_{\mathbb{Q}} 1 = p$ however in F_p we have $1 +_{F_p} 1 +_{F_p} 1 +_{F_p} \cdots +_{F_p} 1 +_{F_p} 1 = 0$.

Here we see that even though F_p may be thought of as a subset of $\mathbb{Q}$ it is certainly not a subfield, because the operations are not the same ($+_{F_p}$ v.s. $+_{\mathbb{Q}}$).

An element α in L is said to be *algebraic* over k if it is the root of a non-trivial polynomial in $k[X]$. All elements in k are algebraic over k. Indeed, $X-\alpha$ is a polynomial in $k[X]$ and has α as a root.

We call the set of all elements α in L that are algebraic over k the *algebraic closure of* k in L and denote it $\overline{k}$.

A subfield k of L is said to be *algebraically closed in* L if the algebraic closure of k is k itself. When $k = L$ we simply say k is *algebraically closed.*

Time for some examples

Example 11.2 ($\mathbb{Q}$ is *not* algebraically closed in $\mathbb{Q}$). *The polynomial $p(X) = X^2 - 2$ cannot be written as $(X-\alpha)(X-\beta)$ for α, β in $\mathbb{Q}$. Essentially this is because $\sqrt{2}$ is not a rational number, i.e. does not lie in $\mathbb{Q}$. However we will give a short proof that forgoes any mention of the square root and uses prime factorisation of the integers instead.*

If $p(X)$ were to split into linear factors $(X-\alpha)(X-\beta)$ with α, β in $\mathbb{Q}$, then $\alpha\beta = -2$. However, since any solution λ of $p(\lambda) = 0$ also has $-\lambda$ as a solution, we see that $\alpha = -\beta$. So $\alpha = \beta = -2$ implies $\alpha^2 = 2$ and there would exist coprime[3] *integers n, m with $\alpha = n/m$ such that $n^2 = 2m^2$.*

Now clearly n^2 must be even, since $n^2/2 = m^2$ is an integer. However, since 2 is a prime number, 2 dividing n^2 implies 2 divides n. This fact is true for all primes p dividing a product ab of integers a, b, then p must divide a or b. So n is even, and n^2 is divisible by 4. So there exists an integer n' so that $4n' = n^2$. Then $m = n^2/2 = 2n'$ and m is even. This contradicts that n, m are coprime.

Example 11.3 ($\mathbb{Q}$ is *not* algebraically closed in $\mathbb{R}$).

[3] A pair of integers n, m are said to be *coprime* if the greatest common divisor is 1.

12

Diagonalisability

Theorem 12.1 (Main characterisations of Diagonalisability)**.** *Let T be an endomorphism of the finite dimensional k-vector space V. The following statements are equivalent*

(i) T is diagonalisable.

(ii) The characteristic polynomial of T splits into linear factors and the arithmetic multiplicity of each eigenvalue is equal to the geometric multiplicity of each eigenvalue.

(iii) There exists an eigenspace decomposition *of V. That is*

$$V = \operatorname{Eig}_T(\lambda_1) \oplus \cdots \oplus \operatorname{Eig}_T(\lambda_k)$$

where λ_i are distinct eigenvalues of T.

There are some less deep, yet nonetheless equivalent characterisations.

Corollary 12.2 (Further characterisations of Diagonalisability)**.** *Let T be an endomorphism of the n-dimensional vector space V. The following statements are equivalent*

(i) T is diagonalisable.

(ii) There are n linearly-independent eigenvectors of T.

(iii) The sum of the dimensions of the eigenspaces is $\dim(V)$*.*

Lemma 12.3. *Let T be an endomorphism of a finite dimensional k-vector space V. Let λ be an eigenvalue of T. Then $1 \le m_g(\lambda) \le m_a(\lambda)$. That is, the geometric multiplicity of λ is at most the algebraic multiplicity and at least 1.*

Proof. If λ is an eigenvalue, then $p_T(\lambda) = \det(T - \lambda\,\mathrm{id}) = 0$. Therefore $T - \lambda\,\mathrm{id}$ is not invertible, and as such cannot be injective and must have a non-trivial kernel. Hence $1 \le \dim(\ker(T - \lambda\,\mathrm{id})) = m_g(\lambda)$.

Suppose the geometric multiplicity of λ is k, then there are k linearly independent eigenvectors $v_1, \dots, v_k$ of T with eigenvalue λ. Using the basis extension lemma, we can extend $v_1, \dots, v_k$ to a basis $\mathcal{B} = (v_1, \dots, v_k, v_{k+1}, \dots, v_n)$ of V. Then

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda I_k & B \\ 0 & D \end{pmatrix}$$

and by Laplace expansion successively down the first column we calculate the characteristic polynomial of T to be $p_T(x) = (x-\lambda)^k p_D(x)$. Hence the algebraic multiplicity of λ is at least k, that is at least the geometric multiplicity. □

12.1 Calculating diagonalisability criteria

Let T be an $n \times n$-matrix. First we outline the steps. Then we will show each of the steps with an example.

(i) **Calculate the characteristic polynomial** using the definition

$$\mathsf{char}_T(\lambda) = \det(T - \lambda\,\mathrm{id}_n).$$

As we will see in the later example, it can be very beneficial to use row (column) operations to simplify the matrix before calculating the determinant. This will likely help with the factorisation steps later.

(ii) **Factorise the characteristic polynomial into linear factors to obtain the eigenvalues and their arithmetic multiplicities**. Here the usual method is to guess roots and then use polynomial division until there is only a quadratic polynomial remaining. Then one can use the quadratic formula.

If the polynomial *does not split* into a product of linear polynomials, then we know that T is *not* diagnosable.

Example. $p(x) = x^1 + 1$ does not split into linear factors over the reals.

If the polynomial *does split* into a product of linear factors, and the *roots are distinct*, we immediately know that T *is* diagonalisable. We recall that the roots are distinct if and only if the arithmetic multiplicity is 1 for each root. Moreover, we know that the geometric multiplicity $m_g(\lambda)$ of any eigenvalue λ satisfies $1 \leq m_g(\lambda) \leq m_a(\lambda)$. Since $m_a(\lambda) = 1$, we conclude that $m_g(\lambda) = m_a(\lambda)$ for every eigenvalue.

Example. $p(x) = x^2 - 1 = (x+1)(x-1)$ splits into linear factors over the rationals.

If the roots are not distinct, we must calculate the geometric multiplicity of the eigenvalues manually.

Example. $p(x) = x^3 - 7x + 16x - 12 = (x-2)^2(x-3)$ splits into linear factors, but the roots are not pairwise distinct: $\lambda = 2$ is a double root.

(iii) **Compute the dimension of the eigenspaces** using the definition

$$\operatorname{Eig}_T(\lambda) = \ker(T - \lambda \,\mathrm{id}_n)$$

and the Gauss Algorithm. Concretely, put the matrix in row-echelon form, and count the number of zero-rows. We do not need to calculate any elements explicitly.

At this point, we know whether the matrix is diagonalisable or not.

(iv) **Compute a basis of each of the eigenspaces**. This can usually be easily done using the Gauss algorithm.

(v) **Use the Gram-Schmidt algorithm to find orthogonal bases of the eigenspaces**. The Gram-Schmidt algorithm guarantees that the matrix will still be diagonal.

We will be following these steps with an example of

$$T = \begin{pmatrix} -4 & -3 & -1 & -7 \\ -3 & -1 & -1 & -4 \\ 6 & 4 & 3 & 8 \\ 3 & 3 & 1 & 6 \end{pmatrix}$$

(i) **Calculate the characteristic polynomial** using the definition

$$\mathrm{char}_T(\lambda) = \det(T - \lambda\,\mathrm{id}_n).$$

In our example this amounts to calculating

$$
\begin{aligned}
&\det\left(\begin{pmatrix} -4-\lambda & -3 & -1 & -7 \\ -3 & -1-\lambda & -1 & -4 \\ 6 & 4 & 3-\lambda & 8 \\ 3 & 3 & 1 & 6-\lambda \end{pmatrix}\right) \\
&\overset{\text{Laplace } R_1}{=} (-4-\lambda)\det\left(\begin{pmatrix} -1-\lambda & -1 & -4 \\ 4 & 3-\lambda & 8 \\ 3 & 1 & 6-\lambda \end{pmatrix}\right) \\
&\qquad + 3\det\left(\begin{pmatrix} -3 & -1 & -4 \\ 6 & 3-\lambda & 8 \\ 3 & 1 & 6-\lambda \end{pmatrix}\right) \\
&\qquad - 1\det\left(\begin{pmatrix} -3 & -1-\lambda & -4 \\ 6 & 4 & 8 \\ 3 & 3 & 6-\lambda \end{pmatrix}\right) \\
&\qquad + 7\det\left(\begin{pmatrix} -3 & -1-\lambda & -1 \\ 6 & 4 & 3-\lambda \\ 3 & 3 & 1 \end{pmatrix}\right) \\
&= (-4-\lambda)\Big((-1-\lambda)\big((3-\lambda)(6-\lambda)-8\big)+1\big(4(6-\lambda)-24\big)-4\big(4-(3-\lambda)3\big)\Big) \\
&\quad +3\Big(-3\big((3-\lambda)(6-\lambda)-8\big)+1\big(6(6-\lambda)-24\big)-4\big(6-(3-\lambda)3\big)\Big) \\
&\quad -\Big(-3\big(4(6-\lambda)-24\big)-(-1-\lambda)\big(6(6-\lambda)-24\big)-4\big(18-12\big)\Big) \\
&\quad +7\Big(-3\big(4-(3-\lambda)3\big)-(-1-\lambda)\big(6-(3-\lambda)3\big)-1\big(18-12\big)\Big) \\
&= \cdots \\
&= \lambda^4-4\lambda^3+3\lambda^2+4\lambda-4
\end{aligned}
$$

$$
\begin{aligned}
&\det\left(\begin{pmatrix}-4-\lambda & -3 & -1 & -7\\ -3 & -1-\lambda & -1 & -4\\ 6 & 4 & 3-\lambda & 8\\ 3 & 3 & 1 & 6-\lambda\end{pmatrix}\right)\\
&\overset{R_1\leftarrow R_1+R_4}{=} \det\left(\begin{pmatrix}-1-\lambda & 0 & 0 & -1-\lambda\\ -3 & -1-\lambda & -1 & -4\\ 6 & 4 & 3-\lambda & 8\\ 3 & 3 & 1 & 6-\lambda\end{pmatrix}\right)\\
&\overset{C_4\leftarrow C_4-C_1}{=} \det\left(\begin{pmatrix}-1-\lambda & 0 & 0 & 0\\ -3 & -1-\lambda & -1 & -1\\ 6 & 4 & 3-\lambda & 2\\ 3 & 3 & 1 & 3-\lambda\end{pmatrix}\right)\\
&\overset{\text{Laplace along } R_1}{=} (-1-\lambda)\det\left(\begin{pmatrix}-1-\lambda & -1 & -1\\ 4 & 3-\lambda & 2\\ 3 & 1 & 3-\lambda\end{pmatrix}\right)\\
&\overset{C_3\leftarrow C_3-C_1}{=} (-1-\lambda)\det\left(\begin{pmatrix}-1-\lambda & -1 & \lambda\\ 4 & 3-\lambda & -2\\ 3 & 1 & -\lambda\end{pmatrix}\right)\\
&\overset{R_1\leftarrow R_1+R_3}{=} (-1-\lambda)\det\left(\begin{pmatrix}2-\lambda & 0 & 0\\ 4 & 3-\lambda & -2\\ 3 & 1 & -\lambda\end{pmatrix}\right)\\
&\overset{\text{Laplace along } R_1}{=} (-1-\lambda)(2-\lambda)\det\left(\begin{pmatrix}3-\lambda & -2\\ 1 & -\lambda\end{pmatrix}\right)\\
&= (-1-\lambda)(2-\lambda)(-\lambda(3-\lambda)+2)\\
&= (-1-\lambda)(2-\lambda)(\lambda^2-3\lambda+2)
\end{aligned}
$$

Of course here R_i stands for the i-th row, and C_j stands for the j-th column.

(ii) **Factorise the characteristic polynomial**. Here the usual method is to guess roots and then use polynomial division until there is only a quadratic polynomial remaining. Then one can use the quadratic formula.

For example consider $p(X) = X^4 - 4X^3 + 3X^2 + 4X - 4$. Calculating $p(1) = 1 - 4 + 3 + 4 - 4 = 0$ we obtain a root $X = 1$. Using polynomial division we then see that $p(X) = (X - 1)(X^3 - 3X^2 + 4)$. Now repeat the process by, guessing a root of $q(X) = X^3 - 3X^2 + 4$. We can try $q(1) = 1 - 3 + 4 = 2$. We can try $q(-1) = -1 - 3 + 4 = 0$. Again using polynomial division we obtain $q(X) = (X + 1)(X^2 - 4X + 4)$. Finally, we use the quadratic formula to find roots of $r(X) = X^2 - 4X + 4$.

13

Inner-product spaces

Scalar products have two main interpretations: a geometric one, namely a scalar product encodes what it means for vectors to be orthogonal; and a topological one: a scalar products gives us a metric which we can use to define a topology on the vector space.

13.1 Bilinear forms

Definition 13.1 (Multi-linear map)**.** Let V, W be k-vector spaces. A *n-linear* map is a map $\tau\colon V^n \to W$ so that for all all $i = 1, \dots, n$ and $v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n$ in V, the map

$$\tau_i\colon V \to W; \quad v \mapsto \tau(v_1, \dots, v_{i-1}, v, v_{i+1}, \dots, v)$$

is linear. Maps like this are called *multi-linear*. If $W = k$, then we speak of a *multi-linear form*.

Remark. This coincides with the term *linear form*. Namely, a linear map $V \to k$. We drop the “1-” from “1-linear form” when speaking of linear forms.

Definition 13.2 (Bilinear form)**.** A *bilinear form* on a vector space V is a map $b\colon V \times V \to k$ so that for all w in V, the maps

$$b_1\colon V \to k; \quad v \to b(v, w) \quad \text{and} \quad b_2\colon V \to k; \quad v \to b(w, v)$$

are linear. We often denote the maps b_1 and b_2 as $b(\cdot, w)$ and $b(w, \cdot)$ respectively. A bilinear form is said to be *symmetric* if $b(v, w) = b(w, v)$ for all v, w in V. Finally, a bilinear form is said to be *positive definite* if $b(v, v) \geq 0$ for all v in V and $b(v, v) = 0$ if and only if $v = 0$.

Definition 13.3 (Sesquilinear form)**.** Let k be a field in which the complex conjugate is defined (e.g. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$). Let V be a k-vector space. A *sesquilinear* form on V is a map $s\colon V \times V \to k$ so that for all w in V the map $b_1\colon V \to k; v \to s(v, w)$ is linear (as before), but now for all v, w, u in V and α in k we have

$$b(v, \alpha w + u) = \overline{\alpha} b(v, w) + b(v, u).$$

Note that sometimes this is written the other way round, with a sesquilinear form being linear in the second argument. Just make sure you know which convention you are following. Here we do not speak of symmetric forms, but of *Hermitian forms*, namely forms that satisfy $b(v, w) = \overline{b(w, v)}$.

13.2 Scalar Products

Definition 13.4 (Scalar product)**.** A *scalar product* on a $\mathbb{C}$-vector space V is a Hermitian positive definite sesquilinear form. We define a scalar product in the same way on a $\mathbb{R}$-vector space. Here we note that being sesquilinear is the same as being bilinear, and Hermitian is the same as being symmetric. Instead of denoting these sesquilinear forms by $b(\cdot,\cdot)$ or $s(\cdot,\cdot)$, we often use the notation $\langle\cdot,\cdot\rangle$.

Example 13.5. *The standard scalar product. Let V be an n-dimensional $\mathbb{C}$-vector space. Then the* standard scalar product *on V is defined as*

$$V \times V \to k; \quad (v,w) \mapsto \sum_{i=1}^{n} w_i \overline{v_i}$$

where v_i, w_i are the coordinates of the vectors v, w expressed in the standard basis $\mathcal{E}_n$ of V.

This is a very important example. It gives us an easy way of defining a scalar product on any $\mathbb{C}$ or $\mathbb{R}$ vector space. This will be relevant for the spectral theorems.

Example 13.6. *Let $V = C([0,1])$ be the space of continuous functions $f\colon [0,1] \to \mathbb{C}$. We can define a scalar product on V with*

$$\langle f, g\rangle = \int_0^1 f(x)\overline{g(x)}\,\mathrm{d}\,x$$

Exercise 13.7. *Is this still an scalar product if defined on the space of square-integrable functions $\mathcal{L}^2([0,1])$ (not just the continuous functions)?*

Definition 13.8 (Inner-product space)**.** We call a vector space equipped with a scalar product an *inner-product space*. We often denote this as $(V, \langle\cdot,\cdot\rangle)$.

Lemma 13.9. *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner-product space and v, w in V. If $\langle v, u\rangle = \langle w, u\rangle$ for all u in V, then $v = w$.*

Remark. It does not suffice for $\langle v, u\rangle = \langle w, u\rangle$ for one u, it must be for all u in V. Choosing $u = 0$ would otherwise tell us that any two vectors v, w are the same, because $\langle v, 0\rangle = \langle w, 0\rangle = 0$ due to sesquilinearity.

Lemma 13.10. *Let $\langle\cdot,\cdot\rangle$ be a scalar product on an n-dimensional $\mathbb{C}$-vector space V. Then for any basis $\mathcal{B}$ of V there exists a matrix $M_{\mathcal{B}}$ in* $\mathrm{Mat}_{n\times n}(\mathbb{C})$ *so that*

$$\langle v, w\rangle = (v_{\mathcal{B}})^t M_{\mathcal{B}} w_{\mathcal{B}}.$$

Here $v_{\mathcal{B}}$ is the vector v expressed in the basis $\mathcal{B}$ and has dimensions $1 \times n$. Moreover, the matrix is Hermitian*, that is, it is equal to its conjugate transpose.*

13.3 Geometry of inner-product spaces

Definition 13.11 (Orthogonal vectors)**.** Let $(V, \langle\cdot,\cdot\rangle)$ be an inner-product space. Then we say that two vectors v, w are *orthogonal* if $\langle v, w\rangle = 0$. We often denote this by $v \perp w$.

Lemma 13.12. *Every finite dimensional vector space has an orthogonal basis.*

Proof. We know that every finite dimensional vector space has a basis. The Gram-Schmidt algorithm can be applied to this basis to find an orthogonal basis. □

13.4 Topology of inner-product spaces

Lemma 13.13. *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner-product space. Then*

$$\|x\| = \sqrt{\langle x, x\rangle}$$

defines a norm on V which in turn defines a metric on V

$$\mathrm{d}(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y\rangle}.$$

As we know, every metric induces a topology by defining open neighbourhoods via open balls. More formally, given a space X with a metric $\mathrm{d}(\cdot,\cdot)$ we define the *open ball of radius r at x* by $B_x(r) = \{y \in X \mid \mathrm{d}(x, y) < r\}$. We then say a set $U \subseteq X$ is *open* if for every point u in U there exists a $0 < r_u$ so that $B_u(r_u) \subseteq U$. Equivalently, we say that U is open if it is the union of (arbitrarily many) balls $B_u(r_u)$.

Lemma. *The standard topology on $\mathbb{R}^n$ is induced by the standard scalar product.*

However, not every norm on V is induced by a scalar product. We have the following lemma to decide this

Lemma 13.14. *Let V be an $\mathbb{R}$-vector space with a norm $\|\cdot\|$. The norm $\|\cdot\|$ is induced by a scalar product if and only if the* parallelogram equality *holds for all v, w in V*

$$\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2$$

13.5 Scalar Products and the dual space

Let $(V, \langle\cdot,\cdot\rangle)$ be an inner-product k-vector space. Then for every u in V

$$\varphi_u : V \longrightarrow k; v \mapsto \langle v, u\rangle$$

is a linear map. (This follows the convention, that scalar products are linear in their first argument. If your convention states that scalar products are linear in the second argument, then you must switch the order of the elements in the scalar product above).

Therefore φ_u lies in the dual V^*. Is the opposite true? This is answered by

Theorem 13.15 (Riesz). *Let $(V, \langle\cdot,\cdot\rangle)$ be a finite dimensional inner-product space. Then $\Phi\colon V \to V^*; v \mapsto \varphi_v(\cdot) = \langle\cdot, u\rangle$ is a bijection. If V is a real vector space, then Φ is a (bijective) linear map, and therefore an isomorphism. In fact, it is a canonical isomorphism.*

Remark. We note that Φ is not an isomorphism when V is a $\mathbb{C}$-vector space. It is not linear, because $\langle\cdot,\cdot\rangle$ is *sesqui*linear in the second argument, and not linear.

Remark. A short note on notation. $\Phi\colon V \to V^*$ is a map into the dual V^*. This means by definition, that for every v in V, $\Phi(v)$ is a map $V \to k$. There are many ways of writing “a map that sends an element to a new map”. Here are three ways

$$\begin{aligned}
\Phi&\colon V \to V^*; u \mapsto (v \mapsto \langle v, u\rangle)\\
\Phi&\colon V \to V^*; u \mapsto \langle\cdot, u\rangle\\
\Phi&\colon V \to V^*; u \mapsto \varphi_u(\cdot)
\end{aligned}$$

Proof. We note that in the real case, the map Φ is clearly linear. It only remains to show that it is a bijection.

Injectivity of Φ follows from Lemma 13.9. Indeed, if $\varphi_u(\cdot) = \varphi_{u'}(\cdot)$, then by definition $\langle v, u\rangle = \langle v, u'\rangle$ for all v in V. Then by Lemma 13.9 we know that $u = u'$.

To show surjectivity, we choose a linear form f in V^* and choose an orthonormal basis $\mathcal{B} = (b_1, \dots, b_n)$ of V. Now, for any vector v in V we may write $v = \langle v, b_1\rangle\, b_1 + \dots + \langle v, b_n\rangle\, b_n$. Then

$$f(v) = f\left(\sum_{i=1}^{n} \langle v, b_i\rangle\, b_i\right) = \sum_{i=1}^{n} \langle v, b_i\rangle\, f(b_i) = \sum_{i=1}^{n} \left\langle v, \overline{f(b_i)} b_i\right\rangle = \left\langle v, \sum_{i=1}^{n} \overline{f(b_i)} b_i\right\rangle.$$

Now $u \overset{\text{def.}}{=} \sum_{i=1}^{n} \overline{f(b_i)} b_i$ does not depend on v: v does not appear in the expression anywhere. Therefore $f = \langle\cdot, u\rangle = \Phi(u)$.

We also note that u does not depend on the choice of basis $\mathcal{B}$. Let $\mathcal{C} = (c_1, \dots, c_n)$ be another orthonormal basis of V. Then by the same computation, we have that

$$f(v) = \left\langle v, \sum_{i=1}^{n} \overline{f(c_i)} c_i\right\rangle = \left\langle v, \sum_{i=1}^{n} \overline{f(b_i)} b_i\right\rangle.$$

This equality holds for all v in V, therefore by Lemma 13.9 we have $\sum_{i=1}^{n} \overline{f(b_i)} b_i = \sum_{i=1}^{n} \overline{f(c_i)} c_i$. This makes the isomorphism canonical in the real case. □

Remark. We showed that in the real case $(V, \langle\cdot,\cdot\rangle)$ is canonically isomorphic to V^*. We have previously mentioned though, that in general V is not canonically isomorphic to V^*. We need the scalar product to find this canonical isomorphism. On the other hand, we have already previously mentioned that any (finite dimensional) vector space can be given the standard scalar product (Example 13.5).

It sounds like we can give a vector space (which is *not* canonically isomorphic to its dual) the standard scalar product, and then it *is* canonically isomorphic to its dual. What gives?

When endowing a (finite dimensional) vector space with the standard scalar product, we are making a choice of a basis.

14

The spectral theorems

14.1 The adjoint and its properties

Consider a linear map of inner-product spaces $T\colon (V, \langle \cdot, \cdot \rangle_V) \to (W, \langle \cdot, \cdot \rangle_W)$. Then, for every w in W we obtain a linear map

$$\varphi_w\colon V \to k; v \mapsto \langle T(v), w \rangle_W$$

So φ_w lies in V^*. The theorem of Riesz 13.15 tells us that $\varphi_w(\cdot) = \langle \cdot, u \rangle$ for some u in V. Let us define $T^*(w) = u$. Then

$$\langle T(v), w \rangle_W = \langle v, T^*(w) \rangle_V$$

for all v in V and w in W.

If V, W are $\mathbb{R}$-vector spaces we can look back at our proof of the theorem of Riesz, and note that $T^* = \Phi^{-1}$ and is therefore linear.

For $\mathbb{C}$-vector spaces we simply compute for all w, w' in W

$$\begin{aligned} \langle v, T^*(w + w') \rangle &= \langle T(v), w + w' \rangle \\ &= \langle T(v), w \rangle + \langle T(v), w' \rangle \\ &= \langle v, T^*(w) \rangle + \langle v, T^*(w') \rangle \\ &= \langle v, T^*(w) + T^*(w') \rangle . \end{aligned}$$

Since this is true for all v in V, we see that $T^*(w + w') = T^*(w) + T^*(w')$. Similarly, we compute for all α in $\mathbb{C}$ and w in W

$$\langle v, T^*(\alpha w) \rangle = \langle T(v), \alpha w \rangle = \overline{\alpha} \langle T(v), w \rangle = \overline{\alpha} \langle v, T^*(w) \rangle = \langle v, \alpha T^*(w) \rangle .$$

Again, since this holds for all v in V, we see that $T^*(\alpha w) = \alpha T^*(w)$. We summarise this in a

Lemma 14.1. *For every linear map of inner-product spaces $T\colon (V, \langle \cdot, \cdot \rangle_V) \to (W, \langle \cdot, \cdot \rangle_W)$ there exists a unique map $T^*\colon W \to V$ so that for all v in V and w in W we have*

$$\langle T(v), w \rangle_W = \langle v, T^*(v) \rangle_V$$

We call this map the adjoint *of T.*

Proposition 14.2 (Proposition 7.1.6)**.** *Let $T\colon V \to W$ be a linear map between inner-products spaces. Let $\mathcal{V} = (v_1, \dots, v_n), \mathcal{W} = (w_1, \dots, w_n)$ be orthonormal bases of V, W*

respectively. Then $[T^*]_{\mathcal{V}}^{\mathcal{W}} = {[T]_{\mathcal{W}}^{\mathcal{V}}}^*$*. In particular, if* T *is an endomorphism, then* $[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}^*$ *for any orthonormal basis* $\mathcal{B}$ *of* V*.*

Proof. The proof is very direct. First recall, the general fact of inner-product spaces: if Z is a finite-dimensional inner-product space with an orthonormal basis $\mathcal{Z} = (z_1, \dots, z_n)$, then the i-th component of any vector z in Z when represented under the basis $\mathcal{Z}$ is $\langle z, z_i \rangle_Z$. In symbols this means

$$z_{\mathcal{Z}} = \left(\langle z, z_1 \rangle, \dots, \langle z, z_n \rangle \right)^t.$$

Now

$$\left([T^*]_{\mathcal{V}}^{\mathcal{W}}\right)_{i,j} \overset{\text{def.}}{=} \left(T^*(w_j)_{\mathcal{V}}\right)_i = \langle T^*(w_j), v_i \rangle = \overline{\langle v_i, T^*(w_j) \rangle} = \overline{\langle T(v_i), w_j \rangle} = \overline{\left(T(v_i)_{\mathcal{W}}\right)_j} \overset{\text{def.}}{=} \overline{\left([T]_{\mathcal{W}}^{\mathcal{V}}\right)_{j,i}} = \left({[T]_{\mathcal{W}}^{\mathcal{V}}}\right)^*_{i,j}$$

and we are done. □

Definition 14.3 (Normal operators, normal matrices)**.** An endomorphism T of an inner-product space is *normal* if it commutes with its adjoint. That is, if $TT^* = T^*T$. We call a matrix A *normal* if it commutes with its adjoint.

Lemma 14.4. *Let* $(V, \langle \cdot, \cdot \rangle)$ *be a finite-dimensional inner-product space. An endomorphism* T *of* V *is normal if and only if its representations matrix* $[T]_{\mathcal{B}}$ *with respect to an orthonormal basis* $\mathcal{B}$ *of* V *is normal.*

Proof. Suppose T is normal. Then

$$[T]_{\mathcal{B}}^*[T]_{\mathcal{B}} = [T^*]_{\mathcal{B}}[T]_{\mathcal{B}} = [T^*T]_{\mathcal{B}} = [TT^*]_{\mathcal{B}} = [T]_{\mathcal{B}}[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}[T]_{\mathcal{B}}^*.$$

Here we used Proposition 14.2 twice. Notice that this requires $\mathcal{B}$ to be orthonormal.

The converse is exactly analogous. Indeed, suppose $[T]_{\mathcal{B}}$ is normal, then

$$[T^*T]_{\mathcal{B}} = [T^*]_{\mathcal{B}}[T]_{\mathcal{B}} = [T]_{\mathcal{B}}^*[T]_{\mathcal{B}} = [T]_{\mathcal{B}}[T]_{\mathcal{B}}^* = [T]_{\mathcal{B}}[T^*]_{\mathcal{B}} = [TT^*]_{\mathcal{B}}.$$

Since a representations matrix uniquely defines the linear operator, we conclude that $TT^* = T^*T$ and that T is normal. □

14.2 Trigonalisability

In this section we will prove the big

Lemma 14.5. *Let* T *be an endomorphism of a finite-dimensional* k*-vector space* V*. The characteristic polynomial* p_T *of* T *splits into linear splits into linear factors over* k *if and only if* T *is trigonalisable.*

We recall that an endomorphism T is said to be *trigonalisable* if there exists a basis $\mathcal{B}$ so that $[T]_{\mathcal{B}}$ is an upper triangle matrix. In a sense this can be seen as a precursor to diagonalising a matrix or putting it in Jordan Form.

Proof. If T is trigonalisable, it is clear that the characteristic polynomial splits into linear factors over k.

Conversely, suppose the characteristic polynomial $p_T(x)$ of T splits into linear factors. We want to show that T is trigonalisable. We will do this via induction over the dimension of the space V. For $n = 1$ the result is clear. In fact *every* linear map $T\colon \mathbb{R}^1 \to \mathbb{R}^1$ is trigonalisable.

Now let us assume that the result is true for all endomorphisms of vector spaces of dimension $k = 1, \dots, (n-1)$ and let V be n-dimensional with T an endomorphism of V for which the characteristic polynomial $p_T(x)$ splits over k.

Let λ (in k) be a root of $p_T(x)$. Then λ is an eigenvalue, and we can write down $\varLambda = \langle v_\lambda \rangle$ for any eigenvector v_λ to the value λ. Clearly $\varLambda$ is a T-invariant subspace of V. By Corollary 9.11 we know that T induces an endomorphism $\overline{T}\colon V/\varLambda \to V/\varLambda$.

Since $\dim(V/\varLambda) = \dim(V) - \dim(\varLambda) = \dim(V) - 1 < \dim(V)$ we are in a good position to use the induction hypothesis. However, we do not know yet, whether the characteristic polynomial $p_{\overline{T}}(x)$ of $\overline{T}$ splits over k.

This is not too hard, since we have done all the work in Corollary 9.12. By definition (v_λ) is a basis of $\varLambda$. Let $\mathcal{B} = (v_\lambda, b_2, \dots, b_n)$ be a basis of V and denote $\mathcal{Q} = (\overline{b_2}, \dots, \overline{b_n})$. This is a basis of $V/\varLambda$. Now the corollary tells us that

$$[T]_{\mathcal{B}} = \left(\begin{array}{c|c} [T|_{\varLambda}]_{(v_\lambda)} & * \\ \hline 0 & \left[\overline{T}\right]_{\mathcal{Q}} \end{array}\right) = \left(\begin{array}{c|c} \lambda & * \\ \hline 0 & \left[\overline{T}\right]_{\mathcal{Q}} \end{array}\right)$$

therefore $p_T(x) = p_{T|_\varLambda}(x) p_{\overline{T}}(x) = (x - \lambda) p_{\overline{T}}(x)$. At this point we recall that $\varLambda = \langle v_\lambda \rangle$ so restricting T to $\varLambda$ yields the map $T|_\varLambda = (v \mapsto \lambda v)$.

We can now conclude that $p_{\overline{T}}(x)$ splits over k

$$p_{\overline{T}}(x) = \frac{p_T(x)}{p_{T|_\varLambda}(x)} = \frac{p_T(x)}{x - \lambda} = \prod_{\substack{\mu \neq \lambda \\ \mu \text{ eigenvalue}}} (x - \mu)$$

We may apply the induction hypothesis to conclude that $\overline{T}$ is trigonalisable. Therefore there exists a basis $\mathcal{P}$ of the quotient $V/\varLambda$ so that $[\overline{T}]_{\mathcal{P}}$ is an upper-triangle matrix. Equivalently, there exists a $\overline{T}$-invariant flag $\left\{\overline{0}\right\} \subseteq \overline{W_1} \subseteq \overline{W_2} \subseteq \cdots \subseteq \overline{W_{n-1}} = V/\varLambda$.

Corollary 9.6 now tells us that $\{0\} \subseteq \Lambda \subseteq \pi^{-1}(\overline{W_1}) \subseteq \pi^{-1}(\overline{W_2}) \subseteq \pi^{-1}(\overline{W_{n-1}}) = V$ is a T-invariant flag. Hence T is trigonalisable. □

Proof alternate ending without flags. ...there exists a basis $\mathcal{P}$ of the quotient V/Λ so that $[\overline{T}]_{\mathcal{P}}$ is an upper-triangle matrix.

Lemma 9.8 tells us that $\mathcal{P}$ can be lifted to a basis $\mathcal{R}$ of $\Lambda^{\perp} \subseteq V$. Recall, this lemma told us that there is an isomorphism $\varphi \colon U^{\perp} \xrightarrow{\sim} W/U$ for all vector spaces W and subspaces U thereof. Therefore $\mathcal{R} = \varphi^{-1}(\mathcal{P}) = (r_2, , \dots , r_n)$ must be a basis of $U^{\perp}$ which is $\Lambda^{\perp}$ in our case.

Since v_λ is a basis of the 1-dimensional space $\Lambda = \langle v_\lambda \rangle$ we conclude that $\mathcal{N} = (v_\lambda, r_2, \dots , r_n)$ is a basis of V. We apply Corollary 9.12 again with the basis $\mathcal{N}$ and obtain

$$[T]_{\mathcal{N}} = \left(\begin{array}{c|c} \lambda & * \\ \hline 0 & [\overline{T}]_{\mathcal{P}} \end{array}\right)$$

which is an upper-triangle matrix. □

Theorem 14.6 (Schur). *Let T be an endomorphism of a finite-dimensional inner-product space V. T is trigonalisable if and only if T is orthogonally trigonalisable.*

Proof. If T is orthogonally trigonalisable, then T is clearly trigonalisable, so let us assume that T is merely trigonalisable.

Let $\mathcal{B} = (b_1, \dots , b_n)$ be a basis so that $[T]_{\mathcal{B}}$ is trigonal. Then, $0 \subseteq \langle b_1 \rangle \subseteq \langle b_1, b_2 \rangle \subseteq \dots \subseteq \langle b_1, b_2, \dots , b_n \rangle = V$ is a T-invariant increasing sequence of linear subspaces (a *flag*) of V. If we now perform Gram-Schmidt on $\mathcal{B}$ we obtain an orthonormal basis $\mathcal{B}'$ of V for which $\langle b_1, \dots , b_k \rangle = \left\langle b'_1, \dots , b'_k \right\rangle$ holds true. Therefore $0 \subseteq \langle b'_1 \rangle \subseteq \langle b'_1, b'_2 \rangle \subseteq \dots \subseteq \langle b'_1, b'_2, \dots , b'_n \rangle = V$ is a flag of V and $[T]_{\mathcal{B}'}$ is trigonal. Hence T is orthogonally trigonalisable. □

14.3 The spectral theorem over $\mathbb{R}$

Theorem 14.7 (Spectral theorem over $\mathbb{R}$). *Let $(V, \langle\cdot,\cdot\rangle)$ be a finite dimensional real inner-product space and T an endomorphism of V. The following are equivalent.*

(i) T is self-adjoint ($[T]_{\mathcal{B}}$ is symmetric for every orthonormal basis $\mathcal{B}$).

(ii) T is orthogonally diagonalisable.

(iii) T is normal and diagonalisable.

(iv) T *is normal and trigonalisable.*

(v) T *is normal and its characteristic polynomial* p_T *splits into linear factors over* $\mathbb{R}$*.*

Proof. We will show $(ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (v) \Rightarrow (ii)$ and $(ii) \Rightarrow (i) \Rightarrow (iv) \Rightarrow (v) \Rightarrow (ii)$, although after having shown the first cycle, the implications $(iv) \Rightarrow (v) \Rightarrow (ii)$ in the second cycle will already be done.

$(ii) \Rightarrow (iii)$ is clear. If T is orthogonally diagonalisable, there exists an orthonormal basis $\mathcal{B}$ of V so that $[T]_{\mathcal{B}}$ is diagonal. In particular $[T]_{\mathcal{B}} = [T]_{\mathcal{B}}^*$, so $[T]_{\mathcal{B}}$ is normal and therefore T is normal.

$(iii) \Rightarrow (iv)$ is clear. Diagonal matrices are trigonal.

$(iv) \Rightarrow (v)$ is clear. If T is trigonalisable, there exists a basis $\mathcal{B}$ of V so that $A = [T]_{\mathcal{B}}$ is an upper-triangle matrix. Then $\operatorname{char}_T(\lambda) = \det(A - \lambda \operatorname{id}) = (A_{1,1} - \lambda)(A_{2,2} - \lambda) \cdots (A_{n,n} - \lambda)$ splits into linear factors over $\mathbb{R}$.

$(v) \Rightarrow (ii)$ is a little more involved. Lemma 14.5 tells us that T is trigonalisable. Moreover, Theorem 14.6 tells us that T is orthogonally trigonalisable with an orthogonal basis $\mathcal{B} = (b_1, \dots, b_n)$. Now we claim that $[T]_{\mathcal{B}}$ is already diagonal.

Before we proceed, we recall the following general fact: Let $\mathcal{C} = (c_1, \dots, c_n)$ be a basis of the k-vector space V', A an $n \times n$ k-matrix and v a vector in V' with $[v]_{\mathcal{C}} = (v_1, \dots, v_n)$, then

$$\left([Av]_{\mathcal{C}}\right)_i = \sum_{j=1} A_{i,j} v_j \quad \text{in other words} \quad Av = \sum_{i=1}^{n} \left(\sum_{j=1} A_{i,j} v_j \right) c_i.$$

Now, we calculate using our recalled formula and that $(b_i)_j = \left([b_i]_{\mathcal{C}}\right)_j = \delta_{i,j}$

$$Tb_1 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} T_{i,j} (b_1)_j \right) b_i \overset{(b_i)_j = \delta_{ij}}{=} \sum_{i=1}^{n} T_{i,1} b_i \overset{\uparrow\Delta}{=} T_{1,1} b_1$$

and so

$$\|Tb_1\|^2 = \langle Tb_1, Tb_1 \rangle = \left\langle T_{1,1} b_1, T_{1,1} b_1 \right\rangle = {T_{1,1}}^2.$$

On the other hand

$$T^* b_1 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} {T^*}_{i,j} (b_1)_j \right) b_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} T_{j,i} (b_1)_j \right) b_i = \sum_{i=1}^{n} T_{1,i} b_i$$

and so

$$\|T^* b_1\|^2 = \langle T^* b_1, T^* b_1 \rangle = \left\langle \sum_{i=1}^{n} T_{1,i} b_i, \sum_{i=1}^{n} T_{1,i} b_i \right\rangle = \sum_{i=1} {T_{1,i}}^2 \sum_{j=1} {T_{1,j}}^2 \left\langle b_i, b_j \right\rangle = \sum_{i=1} {T_{1,i}}^2.$$

However, since T is normal

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle v, T^*Tv\rangle = \langle v, TT^*v\rangle = \langle TT^*v, v\rangle = \langle T^*v, T^*v\rangle = \|T^*v\|^2$$

and so

$$T_{1,1}{}^2 = \sum_{i=1}^{n} T_{1,i}{}^2 = T_{1,1}{}^2 + T_{1,2}{}^2 + \cdots + T_{1,n}{}^2 \implies T_{1,2} = T_{1,3} = \cdots = T_{1,n} = 0.$$

Therefore T has the shape

$$\begin{pmatrix} T_{1,1} & 0_{1\times(n-1)} \\ 0_{(n-1)\times 1} & *_{(n-1)\times(n-1)} \end{pmatrix}$$

We can now proceed inductively to show that T is in fact already diagonal.

This completes proving the first cycle $(ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (v) \Rightarrow (ii)$. Now we will prove $(ii) \Rightarrow (i) \Rightarrow (iv)$.

$(ii) \Rightarrow (i)$. This is clear. If T is orthogonally diagonalisable with the basis $\mathcal{B}$ then $[T]_{\mathcal{B}}$ is diagonal and therefore symmetric. Considering any other orthonormal basis $\mathcal{B}'$ we see that

$$[T]_{\mathcal{B}'} = [\mathrm{id}]^{\mathcal{B}}_{\mathcal{B}'}[T]_{\mathcal{B}}[T]^{\mathcal{B}'}_{\mathcal{B}} = ([\mathrm{id}]^{\mathcal{B}'}_{\mathcal{B}})^{-1}[T]_{\mathcal{B}}[T]^{\mathcal{B}'}_{\mathcal{B}} = ([\mathrm{id}]^{\mathcal{B}'}_{\mathcal{B}})^{T}[T]_{\mathcal{B}}[T]^{\mathcal{B}'}_{\mathcal{B}}$$

Now

$$\left([T]_{\mathcal{B}'}\right)^T = \left(([\mathrm{id}]^{\mathcal{B}'}_{\mathcal{B}})^T[T]_{\mathcal{B}}[T]^{\mathcal{B}'}_{\mathcal{B}}\right)^T = ([T]^{\mathcal{B}'}_{\mathcal{B}})^T([T]_{\mathcal{B}})^T(([\mathrm{id}]^{\mathcal{B}'}_{\mathcal{B}})^T)^T = ([T]^{\mathcal{B}'}_{\mathcal{B}})^T([T]_{\mathcal{B}})^T[\mathrm{id}]^{\mathcal{B}'}_{\mathcal{B}} = [T]_{\mathcal{B}'}$$

so $[T]_{\mathcal{B}}$ is in fact symmetric for all orthonormal bases $\mathcal{B}'$.

$(i) \Rightarrow (iv)$. If T is self-adjoint, it is clearly normal. It remains to show that T is trigonalisable. Here we use our fact again, that an endomorphism is trigonalisable if and only if the characteristic polynomial splits into linear factors (over $\mathbb{R}$ in this case).

The characteristic polynomial always splits over $\mathbb{C}$. Let λ be a (possibly complex) root, and therefore λ is an eigenvalue. If we can show that λ is in fact real, then all roots of the characteristic polynomial are real, and it splits over $\mathbb{R}$.

Consider that T and id are normal, so $T - \lambda\,\mathrm{id}$ is normal. Indeed,

$$\begin{aligned}(T-\lambda\,\mathrm{id})(T-\lambda\,\mathrm{id})^* &= (T-\lambda\,\mathrm{id})(T^* - \overline{\lambda}\,\mathrm{id}^*) \\ &= (T-\lambda\,\mathrm{id})(T^* - \overline{\lambda}\,\mathrm{id}) \\ &= TT^* - \lambda T^* - \overline{\lambda}T + \lambda^2\,\mathrm{id} \\ &= T^*T - \lambda T^* - \overline{\lambda}T + \lambda^2\,\mathrm{id} \\ &= (T-\lambda\,\mathrm{id})^*(T-\lambda\,\mathrm{id})\end{aligned}$$

Therefore $\|(T-\lambda\,\mathrm{id})v\| = \|(T-\lambda\,\mathrm{id})^*v\|$. Indeed, for any normal endomorphism S we show

$$\|Sv\| = \langle Sv, Sv\rangle = \langle v, S^*Sv\rangle = \langle v, SS^*v\rangle = \overline{\langle SS^*v, v\rangle} = \overline{\langle S^*v, S^*v\rangle} = \overline{\|S^*v\|} \overset{\text{real}}{=} \|S^*v\|.$$

Finally, we conclude that if v is an eigenvector of T to the eigenvalue λ, then v is an eigenvector of T^* for the eigenvector $\overline{\lambda}$. Indeed

$$\|(T - \lambda\,\mathrm{id})v\| = 0 = \|(T - \lambda\,\mathrm{id})^*v\| = \left\|T^* - \overline{\lambda}\,\mathrm{id}\,v\right\|$$

However, T is self-adjoint, so

$$\lambda v = T(v) = T^*(v) = \overline{\lambda}v$$

and $\lambda = \overline{\lambda}$ is real. Therefore the characteristic polynomial of T has roots only in $\mathbb{R}$, therefore it splits over $\mathbb{R}$ and hence T is trigonalisable.

This completes the proof of the real spectral theorem. □

15

The Jordan Normal Form

15.1 Matrices in Jordan Form

Let λ lie in a field k. We define a *Jordan block of size n and eigenvalue λ* as the $n \times n$ matrix

$$J_{\lambda,n} = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \lambda & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix}.$$

We immediately remark that a Jordan block $J_{\lambda,n}$ with eigenvalue λ (and size n) indeed has λ as an eigenvalue: using e_i to denote the i-th standard basis vector $e_i = (0, \dots, 1, \dots, 0)$, we see that $J_{\lambda,n}e_1 = \lambda e_1$. The characteristic polynomial of $J_{\lambda,n}$ is $p(x) = (x - \lambda)^n$ and so the algebraic multiplicity of λ is n. The minimal polynomial of $J_{\lambda,n}$ is also $m(x) = (x - \lambda)^n$. This is easily seen by calculating powers of $J_{\lambda,n} - \lambda I_n = J_{0,n}$ which is an easy calculation because most entries are zero. The k-th power of $J_{0,n}$ is zero everywhere with ones on the k-th superdiagonal[4]. The geometric multiplicity of λ is 1. Proving this requires almost no calculation because $J_{\lambda,n} - \lambda I_n$ is already in row-echelon form with one zero-column. In particular, Jordan blocks are good examples of non-diagonalisable matrices.

We then go on to say that a matrix M is in *Jordan form* if there exist Jordan blocks $J_{\lambda_1,n_1}, \dots, J_{\lambda_k,n_k}$ so that

$$M = \begin{pmatrix} J_{\lambda_1,n_1} & & \\ & \ddots & \\ & & J_{\lambda_k,n_k} \end{pmatrix}$$

Calculations on matrices in Jordan form are easy. Many properties can be read off the matrix directly

Proposition 15.1. *Let M be a matrix in Jordan form with blocks $J_{\lambda_1,n_1}, \dots, J_{\lambda_k,n_k}$. Then*

(i) Every λ_i is an eigenvalue of M.

[4] Nebendiagonale

(ii) *The geometric multiplicity of λ_i is equal to the number of Jordan blocks in M with eigenvalue λ_i.*

(iii) *The characteristic polynomial of M is*

$$p_M(x) = (x - \lambda_1)^{n_1} \cdots (x - \lambda_k)^{n_k}$$

(iv) *The minimal polynomial of M is*

$$p_M(x) = (x - \mu_1)^{m_1} \cdots (x - \mu_l)^{m_l}$$

where $\mu_1, \dots, \mu_l$ is the list of distinct eigenvalues of M and m_i is the size of the largest Jordan block λ_{λ_j, n_j} with $\lambda_i = \mu_i$.

Unfortunately, calculating powers of Jordan blocks is not as easy as calculating powers of diagonal matrices, but the following result will show that it is still manageable.

Lemma 15.2 (Powers of jordan blocks)**.** *Setting $\binom{a}{b} = 0$ when $b > a$ we obtain*

$$J_{\lambda,n}{}^k = \begin{pmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \cdots & \cdots & \binom{k}{n}\lambda^{k-n} \\ & \lambda^k & \binom{k}{1}\lambda^{k-1} & \cdots & \cdots & \binom{k}{n}\lambda^{k-n+1} \\ & & \lambda^k & \cdots & \cdots & \vdots \\ & & & \ddots & \cdots & \vdots \\ & & & & \lambda^k & \binom{k}{1}\lambda^{k-1} \\ & & & & & \lambda^k \end{pmatrix}$$

15.2 The Theorem of Jordan

Theorem 15.3 (Jordan)**.** *Let T be an endomorphism of the (finite dimensional) k-vector space V whose characteristic polynomial splits into linear factors over k. Then there exists a basis $\mathcal{B}$ of V so that $[T]_{\mathcal{B}}$ is in Jordan form. This Jordan form representation of T is unique up to permutations of the blocks.*

It is because of the uniqueness (up to permutation of the blocks) that we call this the Jordan *normal* form.

15.3 Preparations in the quotient space

Definition 15.4 (Lifetime of a vector)**.** Let N be a nilpotent endomorphism of the vector space V and let v lie in V. Since N is nilpotent, there exists a k so that $N^k(v) = 0$. We call the largest value l for which $N^l(v) \neq 0$ the *lifetime*[5] of the vector v and denote it by $l_N(v)$.

[5] Lebensdauer

Lemma 15.5. *Let N be a nilpotent endomorphism of V. Let v lie in V and set $l = l_N(v)$. Then the vectors $N^l(v), N^{l-1}(v), \dots, N(v), v$ are linearly independent.*

Proof. Consider $\alpha_l N^l(v) + \alpha_{l-1}N^{l-1}(v), \dots, \alpha_1 N(v) + \alpha_0 v = 0$. Applying N^l we obtain $\alpha_l N^{2l}(v) + \alpha_{l-1}N^{2l-1}(v), \dots, \alpha_1 N^{l+1}(v) + \alpha_0 N^l(v) = \alpha_0 N^l(v) = 0$. So $\alpha_0 = 0$. By repeatedly applying N^{l-i} for $i = 1, \dots, l$ we conclude that all $\alpha_i = 0$. □

The following lemma is quite technical and can be used as a black box result.

Lemma 15.6 (Lifting lemma)**.** *Let N be a nilpotent endomorphism of V. Let w be a vector in V of maximal lifetime and write $W = \mathrm{Span}(N^{l_N(w)}(w), N^{l_N(w)-1}(w), \dots, N(w), w)$. Let $\overline{V} = V/W$. For every vector $\overline{u}$ in $\overline{V}$ there exists a vector v in V so that $\overline{v} = \overline{u}$ and that $l_N(v) = l_{\overline{N}}(\overline{u})$.*

In other words, in the set of vectors $\pi^{-1}(\overline{u}) = \{v \in V \mid \overline{v} = \overline{u}\}$ one them must have the same lifetime as $\overline{u}$. Note that although u is clearly in $\pi^{-1}(\overline{u})$, the lifetime $l_N(u)$ may not equal the lifetime $l_{\overline{N}}(\overline{u})$.

We call it the lifting lemma, because it says that we can "lift" a vector out of the quotient without changing the lifetime.

Proof. Since W is N-invariant, we know that N induces a well-defined endomorphism $\overline{N}\colon \overline{V} \to \overline{V}$. We also know that $\overline{N}^k = \overline{N^k}$ and so $\overline{N}$ is also nilpotent. This means that the lifetime with respect to $\overline{N}$ of a vector in V/W is well-defined. Moreover, the lifetime $l_{\overline{N}}(\overline{u})$ of any vector $\overline{u}$ in $\overline{V}$ is *a priori* less than the lifetime $l_N(v)$ of any vector v in $\pi^{-1}(\overline{u})$. Indeed, if $0 \neq \overline{N}^k(\overline{u}) = \overline{N^k(v)}$ for any $k \geq 1$, then $N^k(v)$ is not in U and in particular not zero.

To shorten notation in the rest of the proof, we set $\bar{l} = l_{\overline{N}}(\overline{u})$, $l = l_N(w)$ and let v be a fixed vector in V so that $\overline{v} = \overline{u}$.

Now set $z = N^{\bar{l}+1}(v)$. Since $\bar{l}$ is the lifetime of $\overline{u} = \overline{v}$ with respect to $\overline{N}$, we see that z is in $W = \ker(\pi = \pi_W)$. Indeed,

$$0 = \overline{N}^{\bar{l}+1}(\overline{v}) = \overline{N^{\bar{l}+1}}(\overline{v}) = \overline{N^{\bar{l}+1}(v)} = \pi(N^{\bar{l}+1}(v)) = \pi(z).$$

Moreover, $N^{l-\bar{l}}(z) = N^{l-\bar{l}}(N^{\bar{l}+1}(v)) = N^{l+1}(v) = 0$ because $N^{l+1}(v) \neq 0$ would imply that v has a longer lifetime than w. So z is also in $\ker(N^{l-\bar{l}})$.

We can now put these two facts together. Since z lies in $W = \mathrm{Span}(N^l(w), N^{l-1}(w), \dots, N(w), w)$ there are coefficients α_i so that $z = \alpha_l N^l(w) + \alpha_{l-1}N^{l-1}(w) + \dots + \alpha_1 N(w) + \alpha_0 w$. However, z also lies in $\ker(N^{l-\bar{l}})$, so

$$0 = N^{l-\bar{l}}(z) = \sum_{i=0}^{l} \alpha_l N^{i+l-\bar{l}}(w).$$

Since $i + l - \bar{l} \geq l + 1$ if and only if $i \geq \bar{l} + 1$, all terms $N^{i+l-\bar{l}}(w)$ vanish for $i \geq \bar{l} + 1$. Therefore the sum collapses to

$$0 = N^{l-\bar{l}}(z) = \sum_{i=0}^{\bar{l}} \alpha_l N^{i+l-\bar{l}}(w) = \alpha_0 N^{l-\bar{l}}(w) + \cdots + \alpha_{\bar{l}} N^l(w).$$

and since $N^l(w), N^{l-1}(w), \ldots, N(w), w$ are linearly independent, we see that all $\alpha_0, \ldots,$ $\alpha_{\bar{l}}$ must vanish. Hence

$$z = \alpha_l N^l(w) + \cdots + \alpha_{l-\bar{l}} N^{l-\bar{l}}(w) = N^{\bar{l}+1}\left(\alpha_l N^{l-\bar{l}-1}(w) + \cdots + \alpha_{l-\bar{l}} N^{l-2\bar{l}-1}(w)\right).$$

Define $r = \alpha_l N^{l-\bar{l}-1}(w) + \cdots + \alpha_{l-\bar{l}} N^{l-2\bar{l}-1}(w)$ so $z = N^{\bar{l}+1}(r)$ and note that r lies in $W = \ker(\pi)$.

We now define our final candidate $v' = v - r$. First we must verify that $\overline{v'} = \overline{u}$. Indeed, $\overline{v'} = \pi(v') = \pi(v - r) = \pi(v) - \pi(r) = \pi(v) = \overline{v} = \overline{u}$. Finally, we must verify that the length of $l_N(v')$ is $\bar{l} = \bar{l}_{\overline{N}}(\overline{u})$. On the one hand we calculate

$$N^{\bar{l}+1}(v') = N^{\bar{l}+1}(v) - N^{\bar{l}+1}(r) = z - z = 0$$

so $l_N(v') \leq \bar{l}$. On the other hand we know that for any vector v in $\pi^{-1}(\overline{u})$, $l_N(v) \geq l_{\overline{N}}(\overline{u})$. □

Lemma 15.7 (Linear independence of lifts)**.** *Suppose $\overline{S}$ is linearly independent in V/W, then there exists a linearly independent set S in V so that $\pi(S) = \overline{S}$.*

Proof. For every element $\overline{s}$ in $\overline{S}$ we choose an element s in V so that $\pi(s) = \overline{s}$. Now suppose $\alpha_1 s_1 + \cdots + \alpha_k s_k = 0$. Then $\alpha_1 \overline{s_1} + \cdots + \alpha_k \overline{s_k} = 0$ and all the α_i must be zero. □

15.4 Jordan Normal form of Nilpotent endomorphisms

Before we prove the existence of the Jordan normal form of arbitrary endomorphisms (whose characteristic polynomials split), we will prove that every *nilpotent* endomorphism N (of finite dimensional) vector spaces has a Jordan normal form. The good news here is, that the hard work has already been done in the lifting lemma (Lemma 15.6).

We know that we must find an N-invariant linear subspace decomposition $V = V_1 \oplus \cdots \oplus V_k$ of V with bases $\mathcal{B}_i$ of V_i so that $[N|_{V_i}]_{\mathcal{B}_i} = J_{0,n_i}$ where $\dim(V_i) = n_i$.

Suppose we have a basis $\mathcal{B}_i = (b_1, \ldots, b_{n_i})$ of V_i so that $[N|_{V_i}]_{\mathcal{B}_i} = J_{0,n_i}$. Then, by definition

$$[N(b_j)]_{\mathcal{B}_i} = \begin{cases} 0 & \text{if} \quad j = 1 \\ e_{j-1} & \text{if} \quad j > 1 \end{cases} \quad \text{equivalently} \quad N(b_j) = \begin{cases} 0 & \text{if} \quad j = 1 \\ b_{j-1} & \text{if} \quad j > 1. \end{cases}$$

Therefore $b_{n-1} = N(b_n)$ and consequently $b_{n-j} = N^j(b_n)$. Notably, $0 \neq b_1 = N^{n_i-1}(b_n)$ and $N(b_1) = B^{n_i}(b_n) = 0$. Therefore b_n is a vector in V_i with lifetime n_i with respect to N.

Conversely, given any non-zero vector v_i we obtain a linearly independent set $\mathcal{B}_{v_i} \overset{\text{def.}}{=} (N^{l(v_i)}(v_i), \dots, N(v_i), v_i)$. If we set $V_i = \text{Span}(\mathcal{B}_{v_i})$ we obtain that $[N|_{V_i}]_{\mathcal{B}_{v_i}} = J_{0,n_i}$ where $n_i = \dim(V_i) = l(v_i)$. This gives us the following

Lemma 15.8. *Let N be a nilpotent endomorphism of the finite dimensional k-vector space V. Finding an N-invariant linear subspace decomposition $V = V_1 \oplus \dots \oplus V_k$ of V with bases $\mathcal{B}_i$ of V_i so that $[N|_{V_i}]_{\mathcal{B}_i} = J_{0,n_i}$ is equivalent to finding vectors $v_1, \dots, v_k$ so that*

$$\mathcal{B} = (\mathcal{B}_{v_1}, \dots, \mathcal{B}_{v_k}) = (N^{l(v_1)}(v_1), \dots, N(v_1), v_1, N^{l(v_2)}(v_2), \dots, N(v_2), v_2, \dots, N^{l(v_k)}, \dots, N(v_k), v_k)$$

is a basis of V.

Remark. Please note that the ordering of the basis vectors $N^j(v_i)$ within the bases $\mathcal{B}_{v_i}$ *is* important, but the ordering of the bases $\mathcal{B}_{v_i}$ within the entire basis $\mathcal{B}$ is *not* important.

Permuting the bases of the linear subspaces $\mathcal{B} = (\mathcal{B}_{v_2}, \mathcal{B}_{v_1}, \mathcal{B}_{v_3}, \dots, \mathcal{B}_{n_k}$ amounts to permuting the Jordan blocks. However permuting the vectors within will break the Jordan Form up.

Consider the following matrix with block structure $J_{0,3}, J_{0,2}$

$$[N]_{\mathcal{B}} = \begin{pmatrix} 0 & 1 & 0 & & \\ 0 & 0 & 1 & & \\ 0 & 0 & 0 & & \\ & & & 0 & 1 \\ & & & 0 & 0 \end{pmatrix}$$

If $\mathcal{B} = (\mathcal{B}_v, \mathcal{B}_w) = (N^2(v), N(v), v, N(w), w)$ is the basis, then switching $\mathcal{B}_v, \mathcal{B}_w$ will result in switching the blocks

$$[N]_{(\mathcal{B}_w, \mathcal{B}_v)} = \begin{pmatrix} 0 & 1 & & & \\ 0 & 0 & & & \\ & & 0 & 1 & 0 \\ & & 0 & 0 & 1 \\ & & 0 & 0 & 0 \end{pmatrix}$$

however changing the order of the vectors within $\mathcal{B}_v$ would lead to

$$[N]_{(v,N^2(v),N(v),N(w),w)} = \begin{pmatrix} 0 & 0 & 1 & & \\ 1 & 0 & 0 & & \\ 0 & 0 & 0 & & \\ & & & 0 & 1 \\ & & & 0 & 0 \end{pmatrix}$$

which is no longer in Jordan Form.

We can now prove the existence of the Jordan normal form for nilpotent endomorphisms. We will do so via induction and the lifting lemma.

Proof of the existence of a Jordan normal form for nilpotent endormophisms. Suppose N is a nilpotent endomorphism of the n-dimensional k-vector space whose characteristic polynomial splits over k and we know that for every endomorphism of an $(n' \leq n-1)$-dimensional k-vector space whose characteristic polynomial splits we can find vectors $v_1, \dots, v_k$ so that $\mathcal{B} = (\mathcal{B}_{v_1}, \dots, \mathcal{B}_{v_k})$ is a basis.

Let w be a vector in V with maximal lifetime with respect to N, write down $\mathcal{B}_w = (N^{l(w)}(w), \dots, N(w), w)$ and set $W = \operatorname{Span}(\mathcal{B}_w), \overline{V} = V/W$. Then $\dim\left(\overline{V}\right) < n = \dim(V)$. Moreover, since W is N-invariant, N induces a well-defined endomorphism $\overline{N}$ on $\overline{V}$. Then by the induction hypothesis, we there exist vectors $\overline{u_1} \dots \overline{u_l}$ in $\overline{V}$ so that

$$\overline{\mathcal{B}} = (\mathcal{B}_{\overline{u_1}}, \dots, \mathcal{B}_{\overline{u_k}}) = (\overline{N}^{l_{\overline{N}}(\overline{u_1})}(\overline{u_1}), \overline{N}^{l_{\overline{N}}(\overline{u_1})-1}(\overline{u_1}), \dots)$$

is a basis of $\overline{V}$.

We can now apply the lifting lemma (Lemma 15.6) to each of the vectors $\overline{u_i}$ to obtain vectors u_i with $l_N(u_i) = l_{\overline{N}}(\overline{u_i})$. Therefore $\left|(\mathcal{B}_{u_1}, \dots, \mathcal{B}_{u_k})\right| = \left|(\mathcal{B}_{\overline{u_1}}, \dots, \mathcal{B}_{\overline{u_k}})\right| = \dim(\overline{V})$ and $(\mathcal{B}_{u_1}, \dots, \mathcal{B}_{u_k})$ is a basis of $W^{\perp}$.

Finally, since $\mathcal{B}_w$ is a basis of W, and $(\mathcal{B}_{u_1}, \dots, \mathcal{B}_{u_k})$ a basis of $W^{\perp}$ we obtain a full basis $(\mathcal{B}_w, \mathcal{B}_{u_1}, \dots, \mathcal{B}_{u_k})$. This completes the proof. □