Linear Algebra

Dave Penneys

August 7, 2008

Contents

Introduction

These notes are designed to be a self contained treatment of linear algebra. The author assumes the reader has taken an elementary course in matrix theory and is well versed with matrices and Gaussian elimination. It is my intention that each time a definition is given, there will be at least two examples given. Most times, the examples will be presented without proof that they are, in fact, examples, and it will be left to the reader to verify that the examples set forth are examples.

Chapter 1

Background Material

1.1 Sets

Throughout these notes, sets will be denoted with curly brackets or letters. A vertical bar in the middle of the curly brackets will mean "such that." For example, $\mathbb{N} = \{1, 2, 3, \dots\}$ is the set of natural numbers, $\mathbb{Z} = \{0, 1, -1, 2, -2, \dots\}$ is the set of integers, $\mathbb{Q} = \big\{p/q \big| p, q \in \mathbb{Z} \text{ and } q \neq 0\big\}$ is the set of rational numbers, $\mathbb{R}$ is the set of real numbers (which we will not define), and $\mathbb{C} = \mathbb{R} + i\mathbb{R} = \big\{x + iy \big| x, y \in \mathbb{R}\big\}$ where $i^2 + 1 = 0$ is the set of complex numbers. If $z = x + iy$ is in $\mathbb{C}$ where x, y are in $\mathbb{R}$, then its complex conjugate is the number $\overline{z} = x - iy$. The modulus, or absolute value, of z is

$$|z| = \sqrt{x^2 + y^2} = \sqrt{z\overline{z}}.$$

Definition 1.1.1. To say x is an element of the set X, we write $x \in X$. We say X is a subset of Y, denoted $X \subset Y$, if $x \in X \Rightarrow x \in Y$ (the double arrow $\Rightarrow$ means "implies"). Sets X and Y are equal if both $X \subset Y$ and $Y \subset X$. If $X \subset Y$ and we want to emphasize that X and Y may be equal, we will write $X \subseteq Y$. If X and Y are sets, we write

(1) $X \cup Y = \big\{x \big| x \in X \text{ or } x \in Y\big\}$,

(2) $X \cap Y = \big\{x \big| x \in X \text{ and } x \in Y\big\}$,

(3) $X \setminus Y = \big\{x \in X \big| x \notin Y\big\}$, and

(4) $X \times Y = \big\{(x, y) \big| x \in X \text{ and } y \in Y\big\}$. $X \times Y$ is called the (Cartesian) product of X and Y.

There is a set with no elements in it, and it is denoted $\emptyset$ or $\{\}$. A subset $X \subset Y$ is called proper if $Y \setminus X \neq \emptyset$, i.e., there is a $y \in Y \setminus X$.

Remark 1.1.2. Note that sets can contain other sets. For example, $\{\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}\}$ is a set. In particular, there is a difference between $\{\emptyset\}$ and $\emptyset$. The first is the set containing the empty set. The second is the empty set. There is something in the first set, namely the empty set, so it is not empty.

Examples 1.1.3.

(1) $\mathbb{N} \cup \mathbb{Z} = \mathbb{Z}$, $\mathbb{N} \cap \mathbb{Z} = \mathbb{N}$, and $\mathbb{Z} \setminus \mathbb{N} = \{0, -1, -2, \dots\}$.

(2) $\emptyset \cup X = X$, $\emptyset \cap X = \emptyset$, $X \setminus X = \emptyset$, and $X \setminus \emptyset = X$ for all sets X.

(3) $\mathbb{R} \setminus \mathbb{Q}$ is the set of irrational numbers.

(4) $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2 = \big\{(x, y)\big|x, y \in \mathbb{R}\big\}$.

Definition 1.1.4. If X is a set, then the power set of X, denoted $\mathcal{P}(X)$, is $\big\{S\big|S \subset X\big\}$.

Examples 1.1.5.

(1) If $X = \emptyset$, then $\mathcal{P}(X) = \{\emptyset\}$.

(2) If $X = \{x\}$, then $\mathcal{P}(X) = \{\emptyset, \{x\}\}$.

(3) If $X = \{x, y\}$, then $\mathcal{P}(X) = \{\emptyset, \{x\}, \{y\}, \{x, y\}\}$.

Exercises

Exercise 1.1.6. A relation on a set X is a subset $\mathcal{R}$ of $X \times X$. We usually write $x\mathcal{R}y$ if $(x, y) \in \mathcal{R}$. The relation $\mathcal{R}$ is called

(i) reflexive if $x\mathcal{R}x$ for all $x \in X$,

(ii) symmetric if $x\mathcal{R}y$ implies $y\mathcal{R}x$ for all $x, y \in X$,

(iii) antisymmetric if $x\mathcal{R}y$ and $y\mathcal{R}x$ implies $x = y$, and

(iv) transitive if $x\mathcal{R}y$ and $y\mathcal{R}z$ implies $x\mathcal{R}z$.

(v) skew-symmetric if $x\mathcal{R}y$ and $x\mathcal{R}z$ implies $y\mathcal{R}z$ for all $x, y, z \in X$.

Find the following examples of a relation $\mathcal{R}$ on a set X:

1. X and $\mathcal{R}$ such that $\mathcal{R}$ is transitive, but not symmetric, antisymmetric, or reflexive.

2. X and $\mathcal{R}$ such that $\mathcal{R}$ is reflexive, symmetric, and transitive, but not antisymmetric,

3. X and $\mathcal{R}$ such that $\mathcal{R}$ is reflexive, antisymmetric, and transitive, but not symmetric, and

4. X and $\mathcal{R}$ such that $\mathcal{R}$ is reflexive, symmetric, antisymmetric, and transitive.

Exercise 1.1.7. Show that a reflexive relation $\mathcal{R}$ on X is symmetric and transitive if and only if $\mathcal{R}$ is skew-transitive.

1.2 Functions

Definition 1.2.1. A function (or map) f from the set X to the set Y, denoted $f\colon X \to Y$, is a rule that associates to each $x \in X$ a unique $y \in Y$, denoted $f(x)$. The sets X and Y are called the domain and codomain of f respectively. The function f is called

(1) injective (or one-to-one) if $f(x) = f(y)$ implies $x = y$,

(2) surjective (or onto) if for all $y \in Y$ there is an $x \in X$ such that $f(x) = y$, and

(3) bijective if f is both injective and surjective.

To say the element $x \in X$ maps to the element $y \in Y$ via f, i.e. $f(x) = y$, we sometimes write $f\colon x \mapsto y$ or $x \mapsto y$ when f is understood. The image, or range, of f, denoted $\operatorname{im}(f)$ or $f(X)$, is $\big\{y \in Y \big|$ there is an $x \in X$ with $f(x) = y\big\}$. Another way of saying that f is surjective is that $\operatorname{im}(f) = Y$. The graph of f is the set $\big\{(x, y) \in X \times Y \big| y = f(x)\big\}$.

Examples 1.2.2.

(1) The natural inclusion $\mathbb{N} \to \mathbb{Z}$ is injective.

(2) The absolute value function $\mathbb{Z} \to \mathbb{N} \cup \{0\}$ is surjective.

(3) Neither of the first two is bijective. The map $\mathbb{N} \to \mathbb{Z}$ given by $1 \mapsto 0$, $2 \mapsto 1$, $3 \mapsto -1$, $4 \mapsto 2$, $5 \mapsto -2$ and so forth is bijective.

Definition 1.2.3. Suppose $f\colon X \to Y$ and $g\colon Y \to Z$. The composite of g with f, denoted $g \circ f$, is the function $X \to Z$ given by $(g \circ f)(x) = g(f(x))$.

Examples 1.2.4.

(1)

(2)

Proposition 1.2.5. *Suppose $f\colon X \to Y$ and $g\colon Y \to Z$.*

(1) If f, g are injective, then so is $g \circ f$.

(2) If f, g are surjective, then so is $g \circ f$.

Proof. Exercise. □

Proposition 1.2.6. *Function composition is associative.*

Proof. Exercise. □

Definition 1.2.7. Let $f\colon X \to Y$, and let $W \subset X$. Then the restriction of f to W, denoted $f|_W\colon W \to Y$, is the function given by $f|_W(w) = f(w)$ for all $w \in W$.

Examples 1.2.8.

(1)

(2)

Proposition 1.2.9. *Let $f\colon X \to Y$.*

(1) f is injective if and only if it has a left inverse, i.e. a function $g\colon Y \to X$ such that $g \circ f = \mathrm{id}_X$, the identity on X.

(2) f is surjective if and only if it has a right inverse, i.e. a function $g\colon Y \to X$ such that $f \circ g = \mathrm{id}_Y$.

Proof.

(1) Suppose f is injective. Then for each $y \in \mathrm{im}(f)$, there is a unique x with $f(x) = y$. Pick $x_0 \in X$, and define a function $g\colon Y \to X$ as follows:

$$g(y) = \begin{cases} x & \text{if } f(x) = y \\ x_0 & \text{else.} \end{cases}$$

It is immediate that $g \circ f = \mathrm{id}_f$. Suppose now that there is a g such that $g \circ f = \mathrm{id}_X$. Then if $f(x_1) = f(x_2)$, applying g to the equation yields $x_1 = g \circ f(x_1) = g \circ f(x_2) = x_2$, so f is injective.

(2) Suppose f is surjective. Then the sets $I_y = \left\{x \in X \middle| f(x) = y\right\}$ are nonempty for each $y \in Y$. By the axiom of choice, we may choose a representative $x_y \in I_y$ for each $y \in Y$. Construct a function $g\colon Y \to X$ by setting $g(y) = x_y$. It is immediate that $f \circ g = \mathrm{id}_Y$. Suppose now that f has a right inverse g. Then if $y \in Y$, we have that $f \circ g(y) = y$, so $y \in \mathrm{im}(f)$ and f is surjective. □

Remark 1.2.10. The axiom of choice is formulated as follows:

Let X be a collection of nonempty sets. A choice function f is a function such that for every $S \in X$, $f(S) \in S$.

Axiom of Choice: There exists a choice function f on X if X is a set of nonempty sets.

Note that in the proof of 1.2.9 (2), our collection of nonempty sets is $\left\{I_y \middle| y \in Y\right\}$, and our choice function is $I_y \mapsto x_y$.

We give a corollary whose proof uses a uniqueness technique found frequently in mathematics.

Corollary 1.2.11. *$f\colon X \to Y$ is bijective if and only if it admits an inverse $g\colon Y \to X$ such that $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$.*

Proof. It is clear by 1.2.9 that if an inverse exists, f is bijective as it admits both a left and right inverse. Suppose now that f is bijective. Then by the preceding proposition, there is a left inverse g and a right inverse h. We then see that

$$g = g \circ \mathrm{id}_Y = g \circ (f \circ h) = (g \circ f) \circ h = \mathrm{id}_X \circ h = h,$$

so $g = h$ is an inverse of f. Note that this direction uses the axiom of choice as it was used in 1.2.9. To prove this fact without the axiom of choice, we define $g\colon Y \to X$ by $g(y) = x$ if $f(x) = y$. As the sets I_y as in the proof of 1.2.9 each only have one element in them, the axiom of choice is unnecessary. □

Remarks 1.2.12.

(1) The proof of 1.2.11 also shows that if f admits an inverse, then it is unique.

(2) Note that we used the associativity of composition of functions in the proof of 1.2.11.

Definition 1.2.13. For $n \in \mathbb{N}$, let $[n] = \{1, 2, \ldots, n\}$. The set X

(1) has n elements if there exists a bijection $[n] \to A$,

(2) is finite if there is a surjection $[n] \to X$ for some $n \in \mathbb{N}$,

(3) is infinite if there is an injection $\mathbb{N} \to X$ (equivalently if X is not finite),

(4) is countable if there is a surjection $\mathbb{N} \to X$,

(5) is denumerable if there is a bijection $\mathbb{N} \to X$,

(6) is uncountable if X is not countable.

Examples 1.2.14.

(1) $[n]$ has n elements and $\mathbb{N}$ is denumerable.

(2) $\mathbb{Q}$ is countable, and $\mathbb{R}$ is uncountable.

(3) $[n]$ is countable, but not denumerable.

(4) If X has n elements, then $\mathcal{P}(X)$ has 2^n elements.

Definition 1.2.15. If A is a finite set, then $|A| \in \mathbb{N}$ is the number n such that A has n elements.

Exercises

Exercise 1.2.16. Prove 1.2.5.

Exercise 1.2.17. Show that $\mathbb{Q}$ is countable.
Hint: Find a surjection $\mathbb{Z}^2 \setminus \mathbb{Z} \times \{0\} \to \mathbb{Q}$, and construct a surjection $\mathbb{N} \to \mathbb{Z}^2 \setminus \mathbb{Z} \times \{0\}$. Then use 1.2.5.

1.3 Fields

Definition 1.3.1. A binary operation on a set X is a function $\#\colon X \times X \to X$ given by $(x, y) \mapsto x\#y$. A binary operation $\#\colon X \times X \to X$ is called

(1) associative if $x\#(y\#z) = (x\#y)\#z$ for all $x, y, z \in X$, and

(2) commutative if $x\#y = y\#x$ for all $x, y \in X$.

Examples 1.3.2.

(1) Addition and multiplication on $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are associative and commutative binary operations.

(2) The cross product $\times\colon \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ is not commutative or associative.

Definition 1.3.3. A field $(\mathbb{F}, +, \cdot)$ is a triple consisting of a set $\mathbb{F}$ and two binary operations $+$ and $\cdot$ on $\mathbb{F}$ called addition and multiplication respectively such that the following axioms are satisfied:

(F1) *additive associativity*: $(a+b)+c = a+(b+c)$ for all $a, b, c \in \mathbb{F}$;

(F2) *additive identity*: there exists $0 \in \mathbb{F}$ such that $a + 0 = a = 0 + a$ for all $a \in \mathbb{F}$;

(F3) *additive inverse*: for each $a \in \mathbb{F}$, there exists an element $b \in \mathbb{F}$ such that $a + b = 0 = b + a$;

(F4) *additive commutativity*: $a + b = b + a$ for all $a, b \in \mathbb{F}$;

(F5) *distributivity*: $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(a + b) \cdot c = a \cdot c + b \cdot c$ for all $a, b, c \in \mathbb{F}$;

(F6) *multiplicative associativity*: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in \mathbb{F}$;

(F7) *multiplicative identity*: there exists $1 \in \mathbb{F} \setminus \{0\}$ such that $a \cdot 1 = a = 1 \cdot a$ for all $a \in \mathbb{F}$;

(F8) *multiplicative inverse*: for each $a \in \mathbb{F} \setminus \{0\}$, there exists $b \in \mathbb{F} \setminus \{0\}$ such that $a \cdot b = 1 = b \cdot a$;

(F9) *multiplicative commutativity*: $a \cdot b = b \cdot a$ for all $a, b \in \mathbb{F}$.

Examples 1.3.4.

(1) $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields.

(2) $\mathbb{Q}(\sqrt{n}) = \big\{x + y\sqrt{n} \big| x, y \in \mathbb{Q}\big\}$ is a field.

Remark 1.3.5. These axioms are also used in defining the following algebraic structures:

AXIOMS	NAME	AXIOMS	NAME
(F1)	*semigroup*	(F1)-(F5)	*nonassociative ring*
(F1)-(F2)	*monoid*	(F1)-(F6)	*ring*
(F1)-(F3)	*group*	(F1)-(F7)	*ring with unity*
(F1)-(F4)	*abelian group*	(F1)-(F8)	*division ring*
		(F1)-(F9)	*field*

Remarks 1.3.6.

(1) The additive inverse of $a \in \mathbb{F}$ in (F3) is usually denoted $-a$.

(2) The multiplicative inverse of $a \in \mathbb{F} \setminus \{0\}$ in (F8) is usually denoted a^{-1} or $1/a$.

(3) For simplicity, we will often denote the field by $\mathbb{F}$ instead of $(\mathbb{F}, +, \cdot)$.

Definition 1.3.7. Let $\mathbb{F}$ be a field. A subfield $\mathbb{K} \subset \mathbb{F}$ is a subset such that $+|_{\mathbb{K}\times\mathbb{K}}$ and $\cdot|_{\mathbb{K}\times\mathbb{K}}$ are well defined binary operations on $\mathbb{K}$, $(\mathbb{K}, +|_{\mathbb{K}\times\mathbb{K}}, \cdot|_{\mathbb{K}\times\mathbb{K}})$ is a field, and the multiplicative identity of $\mathbb{K}$ is the the multiplicative identity of $\mathbb{F}$.

Examples 1.3.8.

(1) $\mathbb{Q}$ is a subfield of $\mathbb{R}$ and $\mathbb{R}$ is a subfield of $\mathbb{C}$.

(2) $\mathbb{Q}(\sqrt{2})$ is a subfield of $\mathbb{R}$.

(3) The irrational numbers $\mathbb{R} \setminus \mathbb{Q}$ do not form a subfield of $\mathbb{R}$.

Finite Fields

Theorem 1.3.9 (Euclidean Algorithm)**.** *Suppose $x \in \mathbb{Z}_{\geq 0}$ and $n \in \mathbb{N}$. Then there are unique $k, r \in \mathbb{Z}_{\geq 0}$ with $r < n$ such that $x = kn + r$.*

Proof. There is a smallest $k \in \mathbb{Z}_{\geq 0}$ such that $kn \leq x$. Set $r = x - kn$. Then $r < n$ and $x = kn + r$. It is obvious that k, r are unique. □

Definition 1.3.10. For $x \in \mathbb{Z}_{\geq 0}$ and $n \in \mathbb{N}$, we define

$$x \mod n = r$$

if $x = kn + r$ with $r < n$.

Examples 1.3.11.

(1)

(2)

Definition 1.3.12.

(1) For $x, y \in \mathbb{Z}$, we say x divides y, denoted $x|y$, if there is a $k \in \mathbb{Z}$ such that $x = ky$.

(2) For $x, y \in \mathbb{N}$, the greatest common divisor, denoted $\gcd(x, y)$, is the largest number $k \in \mathbb{N}$ such that $k|x$ and $k|y$.

(3) We call $x, y \in \mathbb{N}$ relatively prime if $\gcd(x, y) = 1$.

(4) A number $p \in \mathbb{N} \setminus \{1\}$ is called prime if $\gcd(p, n) = 1$ for all $n \in [p-1]$.

Proposition 1.3.13. *$x, y \in \mathbb{N}$ are relatively prime if and only if there are $r, s \in \mathbb{Z}$ such that $rx + sy = 1$.*

Proof. Now if $\gcd(x, y) = k$, then $k|(rx + sy)$ for all $r, s \in \mathbb{Z}$, so existence of $r, s \in \mathbb{Z}$ such that $rx + sy = 1$ implies $\gcd(x, y) = 1$.

Now suppose $\gcd(x, y) = 1$. Let $S = \left\{sx + ty \in \mathbb{N} \middle| s, t \in \mathbb{Z}\right\}$. Then S has a smallest element $n = sx + ty$ for some $s, t \in \mathbb{Z}$. By the Euclidean Algorithm 1.3.9, there are unique $k, r \in \mathbb{Z}_{\geq 0}$ with $r < n$ such that $x = kn + r$. But then

$$r = x - kn = x - k(sx + ty) = (1 - ks)x + (-kt)y \in S.$$

But $r < n$, so $r = 0$, and $x = kn$. Similarly, we have an $l \in \mathbb{Z}_{\geq 0}$ such that $y = ln$. Hence $n|x$ and $n|y$, so $n = 1$. □

Definition 1.3.14. For $n \in \mathbb{N} \setminus \{1\}$, we define $\mathbb{Z}/n = \{0, 1, \ldots, n-1\}$, and we define binary operations # and $*$ on $\mathbb{Z}/n$ by

$$x \# y = (x+y) \mod n \text{ and } x * y = (xy) \mod n.$$

Proposition 1.3.15. *$(\mathbb{Z}/n, \#, *)$ is a field if and only if n is prime.*

Proof. First note that $\mathbb{Z}/n$ has additive inverses as the additive inverse of $x \in \mathbb{Z}/n$ is $n - x$. It then follows that $\mathbb{Z}/n$ is a commutative ring with unit (one must check that addition and multiplication are compatible with the operation $x \mapsto x \mod n$). Thus $\mathbb{Z}/n$ is a field if and only if it has multiplicative inverses. We claim that $x \in \mathbb{Z}/n \setminus \{0\}$ has a multiplicative inverse if and only if $\gcd(x, n) = 1$, which will immediately imply the result.

Suppose $\gcd(x, n) = 1$. By 1.3.13, we have $r, s \in \mathbb{N}$ with $rx + sn = 1$. Hence

$$(r \mod n) * x = (rx) \mod n = (rx + sn) \mod n = 1 \mod n = 1,$$

so x has a multiplicative inverse.

Now suppose x has a multiplicative inverse y. Then

$$(xy) \mod n = 1,$$

so there is an $k \in \mathbb{Z}$ such that $xy = kn + 1$. Thus $xy + (-k)n = 1$, and $\gcd(x, n) = 1$ by 1.3.13. □

Exercises

Exercise 1.3.16 (Fields)**.** Show that $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} | a, b \in \mathbb{Q}\}$ is a field where addition and multiplication are the restriction of addition and multiplication in $\mathbb{R}$.

Exercise 1.3.17. Construct an addition and multiplication table for $\mathbb{Z}/2$, $\mathbb{Z}/3$, and $\mathbb{Z}/5$ and deduce they are fields from the tables.

1.4 Matrices

This section is assumed to be prerequisite knowledge for the student, and definitions in this section will be given without examples, and all results will be stated without proof. For this section, $\mathbb{F}$ is a field.

Definition 1.4.1. An $m \times n$ matrix A over $\mathbb{F}$ is a function $A \colon [m] \times [n] \to \mathbb{F}$. Usually, A is denoted by an $m \times n$ array of elements of $\mathbb{F}$, and the i, j^{th} entry, denoted $A_{i,j}$, is $A(i, j)$ where $(i, j) \in [m] \times [n]$. The set of all $m \times n$ matrices over $\mathbb{F}$ is denoted $M_{m \times n}(\mathbb{F})$, and we will write $M_n(\mathbb{F}) = M_{n \times n}(\mathbb{F})$. The i^{th} row of the matrix A is the $1 \times n$ matrix $A|_{\{i\} \times [n]}$, and the j^{th} row is the $m \times 1$ matrix $A|_{[m] \times \{j\}}$. The identity matrix $I \in M_n(\mathbb{F})$ is the matrix given by

$$I_{i,j} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

Definition 1.4.2.

(1) If $A, B \in M_{m\times n}(\mathbb{F})$, then $A+B$ is the matrix in $M_{m\times n}(\mathbb{F})$ given by $(A+B)_{i,j} = A_{i,j} + B_{i,j}$.

(2) If $A \in M_{m\times n}(\mathbb{F})$ and $\lambda \in \mathbb{F}$, then $\lambda A \in M_{m\times n}(\mathbb{F})$ is the matrix given by $(\lambda A)_{i,j} = \lambda A_{i,j}$.

(3) If $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times p}(\mathbb{F})$, then $AB \in M_{m\times p}(\mathbb{F})$ is the matrix given by

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

(4) If $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{m\times p}(\mathbb{F})$, then $[A|B] \in M_{m\times(n+p)}(\mathbb{F})$ is the matrix given by

$$[A|B]_{i,j} = \begin{cases} A_{i,j} & \text{if } j \leq n \\ B_{i,j-n} & \text{else.} \end{cases}$$

Remark 1.4.3. Note that matrix addition and multiplication are associative.

Proposition 1.4.4. *The identity matrix $I \in M_n(\mathbb{F})$ is the unique matrix in $M_n(\mathbb{F})$ such that $AI = A$ for all $A \in M_{m\times n}(\mathbb{F})$ for all $m \in \mathbb{N}$ and $IB = B$ for all $B \in M_{n\times p}(\mathbb{F})$ for all $p \in \mathbb{N}$.*

Proof. It is clear that $AI = A$ for all $A \in M_{m\times n}(\mathbb{F})$ for all $m \in \mathbb{N}$ and $IB = B$ for all $B \in M_{n\times p}(\mathbb{F})$ for all $p \in \mathbb{N}$. If $J \in M_n(\mathbb{F})$ is another such matrix, then $J = IJ = I$. □

Definition 1.4.5. A matrix $A \in M_n(\mathbb{F})$ is invertible if there is a matrix $B \in M_n(\mathbb{F})$ such that $AB = BA = I$.

Proposition 1.4.6. *If $A \in M_n(\mathbb{F})$ is invertible, then the inverse is unique.*

Remark 1.4.7. In this case, the unique inverse of A is usually denoted A^{-1}.

Definition 1.4.8.

(1) The transpose of the matrix $A \in M_{m\times n}(\mathbb{F})$ is the matrix $A^T \in M_{n\times m}(\mathbb{F})$ given by $(A^T)_{i,j} = A_{j,i}$.

(2) The adjoint of the matrix $A \in M_{m\times n}(\mathbb{C})$ is the matrix $A^* \in M_{n\times m}(\mathbb{C})$ given by $(A^*)_{i,j} = \overline{A_{j,i}}$.

Proposition 1.4.9.

(1) If $A, B \in M_{m\times n}(\mathbb{F})$ and $\lambda \in \mathbb{F}$, then $(A+\lambda B)^T = A^T + \lambda B^T$. If $\mathbb{F} = \mathbb{C}$, then $(A+\lambda B)^* = A^* + \overline{\lambda} B^*$.

(2) If $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times p}(\mathbb{F})$, then $(AB)^T = B^T A^T$. If $\mathbb{F} = \mathbb{C}$, then $(AB)^* = B^* A^*$.

(3) If $A \in M_n(\mathbb{F})$ is invertible, then A^T is invertible, and $(A^T)^{-1} = (A^{-1})^T$. If $\mathbb{F} = \mathbb{C}$, then A^* is invertible, and $(A^*)^{-1} = (A^{-1})^*$.

Definition 1.4.10. Let $A \in M_n(\mathbb{F})$. We call A

(1) upper triangular if $i > j$ implies $A_{i,j} = 0$,

(2) lower triangular if A^T is upper triangular, i.e. $j > i$ implies $A_{i,j} = 0$, and

(3) diagonal if A is both upper and lower triangular, i.e. $i \neq j$ implies $A_{i,j} = 0$.

Definition 1.4.11. Let $A \in M_n(\mathbb{F})$. We call A

(1) block upper triangular if there are square matrices $A_1, \ldots, A_m$ with $m \geq 2$ such that

$$A = \begin{pmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{pmatrix}$$

where the $*$ denotes entries in $\mathbb{F}$.

(2) block lower triangular if A^T is block upper triangular, and

(3) block diagonal if A is both block upper triangular and block lower triangular.

Definition 1.4.12. Matrices $A, B \in M_n(\mathbb{F})$ are similar, denoted $A \sim B$, if there is an invertible $S \in M_n(\mathbb{F})$ such that $S^{-1}AS = B$.

Exercises

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

Exercise 1.4.13. Show that similarity gives a relation on $M_n(\mathbb{F})$ that is reflexive, symmetric, and transitive (see 1.1.6).

Exercise 1.4.14. Prove 1.4.9.

1.5 Systems of Linear Equations

This section is assumed to be prerequisite knowledge for the student, and definitions in this section will be given without examples, and all results will be stated without proof. For this section, $\mathbb{F}$ is a field.

Definition 1.5.1.

(1) An $\mathbb{F}$-linear equation, or a linear equation over $\mathbb{F}$, is an equation of the form

$$\sum_{i=1}^{n} \lambda_i x_i = \lambda_1 x_1 + \cdots + \lambda_n x_n = \mu \text{ where } \lambda_1, \ldots, \lambda_n, \mu \in \mathbb{F}.$$

The x_i's are called the variables, and λ_i is called the coefficient of the variable x_i. An $\mathbb{F}$-linear equation is completely determined by a matrix

$$A = \begin{pmatrix} \lambda_1 & \cdots & \lambda_n \end{pmatrix} \in M_{1\times n}(\mathbb{F})$$

and a number $\mu \in \mathbb{F}$. In this sense, we can give an equivalent definition of an $\mathbb{F}$-linear equation as a pair (A, μ) with $A \in M_{1\times n}(\mathbb{F})$ and a $\mu \in \mathbb{F}$.

(2) A solution of the $\mathbb{F}$-linear equation

$$\sum_{i=1}^{n} \lambda_i x_i = \lambda_1 x_1 + \cdots + \lambda_n x_n = \mu$$

is an element

$$(r_1, \ldots, r_n) \in \mathbb{F}^n = \underbrace{\mathbb{F} \times \cdots \times \mathbb{F}}_{n \text{ copies}}$$

such that

$$\sum_{i=1}^{n} \lambda_i r_i = \lambda_1 r_1 + \cdots + \lambda_n r_n = \mu.$$

Equivalently, a solution to the $\mathbb{F}$-linear equation (A, μ) is an $x \in M_{n\times 1}(\mathbb{F})$ such that $Ax = \mu$.

(3) A system of $\mathbb{F}$-linear equations is a finite number m of $\mathbb{F}$-linear equations:

$$\sum_{i=1}^{n} \lambda_{1,i} x_i = \lambda_{1,1} x_1 + \cdots + \lambda_{1,n} x_n = \mu_1$$
$$\vdots$$
$$\sum_{i=1}^{n} \lambda_{m,i} x_i = \lambda_{m,1} x_1 + \cdots + \lambda_{m,n} x_n = \mu_n$$

where $\lambda_{i,j} \in \mathbb{F}$ for all i, j. Equivalently, a system of $\mathbb{F}$-linear equations is a pair (A, y) where $A \in M_{m\times n}(\mathbb{F})$ and $y \in M_{m\times 1}(\mathbb{F})$.

(4) A solution to the system of $\mathbb{F}$-linear equations

$$\sum_{i=1}^{n} \lambda_{1,i} x_i = \lambda_{1,1} x_1 + \cdots + \lambda_{1,n} x_n = \mu_1$$
$$\vdots$$
$$\sum_{i=1}^{n} \lambda_{m,i} x_i = \lambda_{m,1} x_1 + \cdots + \lambda_{m,n} x_n = \mu_n$$

is an element $(r_1, \ldots, r_n) \in \mathbb{F}^n$ such that

$$\sum_{i=1}^{n} \lambda_{1,i} r_i = \lambda_{1,1} r_1 + \cdots + \lambda_{1,n} r_n = \mu_1$$
$$\vdots$$
$$\sum_{i=1}^{n} \lambda_{m,i} r_i = \lambda_{m,1} r_1 + \cdots + \lambda_{m,n} r_n = \mu_n.$$

Equivalently, a solution to the system of $\mathbb{F}$-linear equations (A, y) is an $x \in M_{n\times 1}(\mathbb{F})$ such that $Ax = y$.

Definition 1.5.2.

(1) An elementary row operation on a matrix $A \in M_{m\times n}(\mathbb{F})$ is performed by

- (i) doing nothing,
- (ii) switching two rows,
- (iii) multiplying one row by a constant $\lambda \in \mathbb{F}$, or
- (iv) replacing the i^{th} row with the sum of the i^{th} row and a scalar (constant) multiple of another row.

(2) An $n \times n$ elementary matrix is a matrix obtained from the identity matrix $I \in M_n(\mathbb{F})$ by performing not more than one elementary row operation.

(3) Matrices $A, B \in M_{m\times n}(\mathbb{F})$ are row equivalent if there are elementary matrices $E_1, \ldots, E_n$ such that $A = E_n \cdots E_1 B$.

Remark 1.5.3. Elementary row operations are invertible, i.e. every elementary row operation can be undone by another elementary row operation.

Proposition 1.5.4. *Performing an elementary row operation on a matrix $A \in M_{m\times n}(\mathbb{F})$ is equivalent to multiplying A by the elementary matrix obtained by doing the same elementary row operation to the identity.*

Corollary 1.5.5. *Elementary matrices are invertible.*

Proposition 1.5.6. *Suppose $A \in M_n(\mathbb{F})$. Then A is invertible if and only if A is row equivalent to I.*

Definition 1.5.7. A pivot of the i^{th} row of the matrix $A \in M_{m\times n}(\mathbb{F})$ is the first nonzero entry in the i^{th} row. If there is no such entry, then the row has no pivots.

Definition 1.5.8. Let $A \in M_{m\times n}(\mathbb{F})$.

(1) A is said to be in row echelon form if the pivot in the $(i+1)^{\text{th}}$ row (if it exists) is in a column strictly to the right of the pivot in the i^{th} row (if it exists) for $i = 1, \ldots, n-1$, i.e. if the pivot of the i^{th} row is $A_{i,j}$ and the pivot of the $(i+1)^{\text{th}}$ row is $A_{(i+1),k}$, then $k > j$.

(2) A is said to be in reduced row echelon form, or is said to be row reduced, if A is in row echelon form, all pivots of A are equal to 1, and all entries that occur above a pivot are zeroes.

Theorem 1.5.9 (Gaussian Elimination Algorithm)**.** *Every matrix over $\mathbb{F}$ is row equivalent to a matrix in row echelon form, and every matrix over $\mathbb{F}$ is row equivalent to a unique matrix in reduced row echelon form.*

Theorem 1.5.10. *Suppose $A \in M_{m\times n}(\mathbb{F})$. The system of $\mathbb{F}$-linear equations (A, y) has a solution if and only if the augmented matrix $[A|y]$ can be row reduced so that no pivot occurs in the $(n+1)^{th}$ column. The solution is unique if and only if A can be row reduced to the identity matrix.*

Exercises

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

Exercise 1.5.11. Prove 1.5.5.

Exercise 1.5.12. Prove 1.5.9.

Exercise 1.5.13. Prove 1.5.10.

Exercise 1.5.14. Prove 1.5.6.

Chapter 2

Vector Spaces

The objects that we will study this semester are vector spaces. The main theorem in this chapter is that vector spaces over a field $\mathbb{F}$ are classified by their dimension. In this chapter $\mathbb{F}$ will denote a field.

2.1 Definition

Definition 2.1.1. A vector space consists of

(1) a scalar field $\mathbb{F}$;

(2) a set V, whose elements are called vectors;

(3) a binary operation $+$ called vector addition on V such that

(V1) addition is associative, i.e. $u + (v + w) = (u + v) + w$ for all $u, v, w \in V$,

(V2) addition is commutative, i.e. $u + v = v + u$ for all $u, v \in V$

(V3) *vector additive identity*: there is a vector $0 \in V$ such that $0 + v = v$ for all $v \in V$, and

(V4) *vector additive inverse*: for each $v \in V$, there is a $-v \in V$ such that $v + (-v) = 0$;

(4) and a map $\cdot \colon \mathbb{F} \times V \to V$ given by $(\lambda, v) \mapsto \lambda \cdot v$ called a scalar multiplication such that

(V5) *scalar unit*: $1 \cdot x = x$ for all $v \in V$ and

(V6) *scalar associativity*: $\lambda \cdot (\mu \cdot v) = (\lambda\mu) \cdot v$ for all $v \in V$ and $\lambda, \mu \in \mathbb{F}$;

such that the distributive properties hold:

(V7) $\lambda \cdot (u + v) = (\lambda \cdot u) + (\lambda \cdot v)$ for all $\lambda \in \mathbb{F}$ and $u, v \in V$ and

(V8) $(\lambda + \mu) \cdot v = (\lambda \cdot v) + (\mu \cdot v)$ for all $\lambda, \mu \in \mathbb{F}$ and $v \in V$.

Examples 2.1.2.

(1) $(\mathbb{F}, +, \cdot)$ is a vector space over $\mathbb{F}$ where 0 is the additive identity and $-x$ is the additive inverse of $x \in \mathbb{F}$.

(2) Let $M_{m\times n}(\mathbb{F})$ denote the $m \times n$ matrices over $\mathbb{F}$. $(M_{n\times k}(\mathbb{F}), +, \cdot)$ is a vector space where $+$ is addition of matrices and $\cdot$ is the usual scalar multiplication; if $A = (A_{ij}), B = (B_{ij}) \in M_{m\times n}(\mathbb{F})$ and $\lambda \in \mathbb{F}$, then $(A+B)_{ij} = A_{ij} + B_{ij}$ and $(\lambda A)_{ij} = \lambda A_{ij}$.

(3) Note that $\mathbb{F}^n = M_{n\times 1}(\mathbb{F})$ is a vector space over $\mathbb{F}$.

(4) Let $\mathbb{K} \subset \mathbb{F}$ be a subfield. Then $\mathbb{F}$ is a vector space over $\mathbb{K}$.

(5) $F(X, \mathbb{F}) = \{f \colon X \to \mathbb{F}\}$ is a vector space over $\mathbb{F}$ with pointwise addition and scalar multiplication:

$$(\lambda f + g)(x) = \lambda f(x) + g(x).$$

(6) $C(X, \mathbb{F})$, the set of continuous functions from $X \subset \mathbb{F}$ to $\mathbb{F}$ is a vector space over $\mathbb{F}$. We will write $C(a,b) = C((a,b), \mathbb{R})$, and similarly for closed or half-open intervals.

(7) $C^1(a,b)$, the set of $\mathbb{R}$-valued, continuously differentiable functions on the interval $(a,b) \subset \mathbb{R}$ is a vector space over $\mathbb{R}$. This example can be generalized to $C^n(a,b)$, the n-times continuously differentiable functions on (a,b), and $C^\infty(a,b)$, the infinitely differentiable functions.

Now that we know what a vector space is, we begin to develop the necessary tools to prove that vector spaces over $\mathbb{F}$ are classified by their dimension. From this point on, let $(V, +, \cdot)$ be a vector space over the field $\mathbb{F}$.

Definition 2.1.3. Let $(V, +, \cdot)$ be a vector space over $\mathbb{F}$. Then $W \subset V$ is a vector subspace (or subspace) if $+|_{W\times W}$ is an addition on W (W is closed under addition), $\cdot|_{\mathbb{F}\times W}$ is a scalar multiplication on W (W is closed under scalar multiplication), and $(W, +|_{W\times W}, \cdot|_{\mathbb{F}\times W})$ is a vector space over $\mathbb{F}$. A subspace $W \subset V$ is called proper if $W \neq V$.

Examples 2.1.4.

(1) Every vector space has a zero subspace (0) consisting of the vector 0. Also, if V is a vector space, V is a subspace of V.

(2) $\mathbb{R}^2$ is not a subspace of $\mathbb{R}^3$. $\mathbb{R}^2$ is not even a subset of $\mathbb{R}^3$. There are subspaces of $\mathbb{R}^3$ which look exactly like $\mathbb{R}^2$. We will explain what this means when we discuss the concept of isomorphism of vector spaces.

(3) If $A \in M_{m\times n}(\mathbb{F})$ is an $m \times n$ matrix, then $NS(A) \subset \mathbb{R}^n$ is the subspace consisting of all $x \in \mathbb{R}^n$ such that $Ax = 0$.

(4) $C^n(a,b)$ is a subspace of $C(a,b)$ for all $n \in \mathbb{N} \cup \{\infty\}$.

(5) $C(a,b)$ is a subspace of $F((a,b), \mathbb{R})$.

Proposition 2.1.5. *A subset $W \subset V$ is a subspace if it is closed under addition and scalar multiplication, i.e. $\lambda u + v \in W$ for all $u, v \in W$ and $\lambda \in \mathbb{F}$.*

Proof. If W is closed under addition and scalar multiplication, then clearly $a|_{W \times W}$ and $s|_{\mathbb{F} \times W}$ will satisfy the distributive property. It remains to check that W has additive inverses and W has 0. Since $-1 \in \mathbb{F}$, we have that if $w \in W$, $-1 \cdot w \in W$. It is an easy exercise to check that $-1 \cdot w = -w$, the additive inverse of w. Now since $w, -w \in W$, we have $w + (-w) = 0 \in W$, and we are finished. □

Exercises

V will denote a vector space over $\mathbb{F}$. When confusion can arise, 0_V will denote the additive identity of V, and $0_{\mathbb{F}}$ will denote the additive identity of $\mathbb{F}$.

Exercise 2.1.6. Show that $(\mathbb{R}_{>0}, \#, \star)$ is a vector space over $\mathbb{R}$ where $x \# y = xy$ is the addition and $r \star x = x^r$ is the scalar multiplication.

Exercise 2.1.7.

Show that if V is a vector space over $\mathbb{F}$, then the additive identity is unique. Show if $v \in V$, then the additive inverse of v is unique.

Exercise 2.1.8. Let $\lambda \in \mathbb{F}$. Show that $\lambda \cdot 0_V = 0_V$.

Exercise 2.1.9. Let $v \in V$ and $\lambda \in \mathbb{F} \setminus \{0\}$. Show that $\lambda \cdot v = 0$ implies $v = 0$.

Exercise 2.1.10. Let $v \in V$. Show that $0_{\mathbb{F}} \cdot v = 0_V$.

Exercise 2.1.11. Let $v \in V$. Show that $(-1) \cdot v = -v$, the additive inverse of v.

Exercise 2.1.12 (Complexification)**.** Let $(V, +, \cdot)$ be a vector space over $\mathbb{R}$. Let $V_{\mathbb{C}} = V \times V$, and define functions $+ \colon V_{\mathbb{C}} \times V_{\mathbb{C}} \to V_{\mathbb{C}}$ and $\cdot \colon \mathbb{C} \times V_{\mathbb{C}} \to V_{\mathbb{C}}$ by

$$(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2) \text{ and } (x + iy) \cdot (u_1, v_1) = (xu_1 - yv_1, yu_1 + xv_1).$$

(1) Show $(V_{\mathbb{C}}, +, \cdot)$ is a vector space over $\mathbb{C}$ called the complexification of the real vector space V.

(2) Show that $(0, v) = i(v, 0)$ for all $v \in V$.

Note: *This implies that we can think of $V \subset V_{\mathbb{C}}$ as all vectors of the form $(v, 0)$, and we can write $(u, v) = u + iv$. Addition and scalar multiplication can then be rewritten in the naive way as*

$$(u_1+iv_1)+(u_2+iv_2) = (u_1+u_2)+i(v_1+v_2) \textit{ and } (x+iy)\cdot(u_1+iv_1) = xu_1-yv_1+i(yu_1+xv_1).$$

For $w = u + iv \in V_{\mathbb{C}}$, the real part of $u + iv \in V_{\mathbb{C}}$ is $\operatorname{Re}(w) = u \in V$, the imaginary part is $\operatorname{Im}(v) = v \in V$, and the conjugate is $\overline{w} = u - iv \in V_{\mathbb{C}}$.

(3) Show that $u_1 + iv_1 = u_2 + iv_2$ if and only if $u_1 = u_2$ and $v_1 = v_2$. Hence, two vectors are equal if and only if their real and imaginary parts are equal.

(4) Find $(\mathbb{R}^n)_{\mathbb{C}}$ and $(M_{m \times n}(\mathbb{R}))_{\mathbb{C}}$.

Hint: Use the identification $V_{\mathbb{C}} = V + iV$ described in the note.

Exercise 2.1.13 (Quotient Spaces)**.** Let $W \subset V$ be a subspace. For $v \in V$, define

$$v + W = \left\{ v + w \middle| w \in W \right\}.$$

$v + W$ is called a coset of W. Let $V/W = \left\{ v + W \middle| v \in V \right\}$, the set of cosets of W.

(1) Show that $u + W = v + W$ if and only if $u - v \in W$.

(2) Define an addition # and a scalar multiplication $*$ on V/W so that $(V/W, \#, *)$ is a vector space.

2.2 Linear Combinations, Span, and (Internal) Direct Sum

Definition 2.2.1. Let $v_1, \ldots, v_n \in V$. A linear combination of $v_1, \ldots, v_n$ is a vector of the form

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

where $\lambda_i \in \mathbb{F}$ for all $i = 1, \ldots, n$. It is convention that the empty linear combination is equal to zero (i.e. the sum of no vectors is zero).

Definition 2.2.2. Let $S \subset V$ be a subset. Then

$$\operatorname{span}(S) = \bigcap \left\{ W \middle| W \text{ is a subspace of } V \text{ with } S \subset W \right\}.$$

Examples 2.2.3.

(1) If $A \in M_{m \times n}(\mathbb{F})$ is a matrix, then the row space of A is the subspace of $\mathbb{R}^n$ spanned by the rows of A and the column space of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A. These subspaces are denoted $RS(A)$ and $CS(A)$ respectively.

(2) $\operatorname{span}(\emptyset) = (0)$, the zero subspace.

Remarks 2.2.4. Suppose $S \subset V$.

(1) $\operatorname{span}(S)$ is the smallest subspace of V containing S, i.e. if W is a subspace containing S, then $\operatorname{span}(S) \subset W$.

(2) Since $\operatorname{span}(S)$ is closed under addition and scalar multiplication, we have that all (finite) linear combinations of elements of S are contained in $\operatorname{span}(S)$. As the set of all linear combinations of elements of S forms a subspace, we have that

$$\operatorname{span}(S) = \left\{ \sum_{i=1}^{n} \lambda_i v_i \middle| n \in \mathbb{Z}_{\geq 0},\ \ \lambda_i \in \mathbb{F},\ \ v_i \in S \text{ for all } i = 1, \ldots, n \right\}.$$

Note that this still makes sense for $S = \emptyset$ as $\operatorname{span}(S)$ contains the empty linear combination, zero.

Definition 2.2.5. If V is a vector space and $W_1, \dots, W_n \subset V$ are subspaces, then we define the subspace

$$W_1 + \cdots + W_n = \sum_{i=1}^{n} W_i = \left\{ \sum_{i=1}^{n} w_i \middle| w_i \in W_i \text{ for all } i \in [n] \right\} \subset V.$$

Examples 2.2.6.

(1) $\operatorname{span}\{(1,0)\} + \operatorname{span}\{(0,1)\} = \mathbb{R}^2$.

(2)

Proposition 2.2.7. *Suppose $W_1, \dots, W_n \subset V$ are subspaces. Then*

$$\sum_{i=1}^{n} W_i = \operatorname{span}\left(\bigcup_{i=1}^{n} W_i \right).$$

Proof. Suppose $v \in \displaystyle\sum_{i=1}^{n} W_i$. Then v is a linear combination of elements of the W_i's. so $v \in \operatorname{span}\left(\displaystyle\bigcup_{i=1}^{n} W_i \right)$.

Now suppose $v \in \operatorname{span}\left(\displaystyle\bigcup_{i=1}^{n} W_i \right)$. Then v is a linear combination of elements of the W_i's, so there are $w_1^1, \dots, w_{m_1}^1, \dots, w_1^n, \dots, w_{m_n}^n$ and scalars $\lambda_1^1, \dots, \lambda_{m_1}^1, \dots, \lambda_1^n, \dots, \lambda_{m_n}^n$ such that $w_j^i \in W_i$ for all $j \in [m_j]$ and

$$v = \sum_{i=1}^{n} \sum_{j=1}^{m_j} \lambda_j^i w_j^i.$$

For $i \in [n]$, set

$$u_i = \sum_{j=1}^{m_j} \lambda_j^i w_j^i \in W_i$$

to see that $v = \displaystyle\sum_{i=1}^{n} u_i \in \sum_{i=1}^{n} W_i$. □

Definition-Proposition 2.2.8. *Suppose $W_1, \dots, W_n \subset V$ are subspaces, and let*

$$W = \sum_{i=1}^{n} W_i.$$

The following conditions are equivalent:

(1) $W_i \cap \displaystyle\sum_{j \neq i} W_j = (0)$ for all $i \in [n]$,

(2) for each $v \in W$, there are unique $w_i \in W_i$ for $i \in [n]$ such that $v = \sum_{i=1}^{n} w_i$.

(3) if $w_i \in W_i$ for $i \in [n]$ such that $\sum_{i=1}^{n} w_i = 0$, then $w_i = 0$ for all $i \in [n]$.

If any of the above three conditions are satisfied, we call W the direct sum of the W_i's, denoted

$$W = \bigoplus_{i=1}^{n} W_i.$$

Proof.

<u>(1) ⇒ (2):</u> Suppose $W_i \cap \sum_{j \neq i} W_j = (0)$ for all $i \in [n]$, and let $v \in W$. Since $W = \sum_{i=1}^{n} W_i$, there are $w_i \in W_i$ for $i \in [n]$ such that $v = \sum_{i=1}^{n} w_i$. Suppose $v = \sum_{i=1}^{n} w_i'$ with $w_i' \in W_i$ for all $i \in [n]$. Then

$$0 = v - v = \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i' = \sum_{i=1}^{n} w_i - w_i'.$$

Since $w_i' - w_i \in W_i$ for all $i \in [n]$, we have

$$w_j' - w_j = \sum_{i \neq j} w_i - w_i' \in W_j \cap \sum_{i \neq j} W_i = (0),$$

so $w_j - w_j' = 0$. Similarly, $w_i = w_i'$ for all $i \in [n]$, and the expression is unique.

<u>(2) ⇒ (3):</u> Trivial.

<u>(3) ⇒ (1):</u> Now suppose (3) holds. Suppose

$$w \in W_i \cap \sum_{j \neq i} W_j$$

for $i \in [n]$. Then we have $w_j \in W_j$ for all $j \neq i$ such that

$$w = \sum_{j \neq i} w_j.$$

Setting $w_i = -w$, we have that

$$0 = w - w = \sum_{j=1}^{n} w_j,$$

so $w_j = 0$ for all $j \in [n]$, and $w_i = 0 = -w$. Thus $w = 0$. □

Remarks 2.2.9.

(1) If $V = \bigoplus_{i=1}^{n} W_i$ and $v = \sum_{i=1}^{n} w_i$ where $w_i \in W_i$ for all $i \in [n]$, then w_i is called the W_i-component of w for $i \in [n]$.

(2) The proof (1) $\Rightarrow$ (2) highlights another important proof technique called the "in two places at once" technique.

Examples 2.2.10.

(1) $V = V \oplus (0)$ for all vector spaces V.

(2) $\mathbb{R}^2 = \operatorname{span}\{(1,0)\} \oplus \operatorname{span}\{(0,1)\}$.

(2) $C(\mathbb{R},\mathbb{R}) = \{\text{even functions}\} \oplus \{\text{odd functions}\}$.

(2) Suppose $Y \subset X$, a set. Then

$$F(X,\mathbb{F}) = \left\{f \in F(X,\mathbb{F})\middle| f(y) = 0 \text{ for all } y \in Y\right\} \oplus \left\{f \in F(X,\mathbb{F})\middle| f(x) = 0 \text{ for all } x \notin Y\right\}.$$

Exercises

2.3 Linear Independence and Bases

Definition 2.3.1. A subset $S \subset V$ is linearly independent if for every finite subset $\{v_1, \dots, v_n\} \subset S$,

$$\sum_{i=1}^{n} \lambda_i v_i = 0 \text{ implies that } \lambda_i = 0 \text{ for all } i \in [n].$$

We say the vectors $v_1, \dots, v_n$ are linearly independent if $\{v_1, \dots, v_n\}$ is linearly independent. If a set S is not linearly independent, it is linearly dependent.

Examples 2.3.2.

(1) Letting $e_i = (0, \dots, 0, 1, 0, \dots, 0)^T \in \mathbb{F}^n$ where the 1 is in the i^{th} slot for $i \in [n]$, we get that $\left\{e_i\middle| i \in [n]\right\}$ is linearly independent.

(2) Define $E^{i,j} \in M_{m\times n}(\mathbb{F})$ by

$$(E^{i,j})_{k,l} = \begin{cases} 1 & \text{if } i = k, j = l \\ 0 & \text{else.} \end{cases}$$

Then $\left\{E^{i,j}\middle| i \in [m], j \in [n]\right\}$ is linearly independent.

(3) The functions $\delta_x \colon X \to \mathbb{F}$ given by

$$\delta_x(y) = \begin{cases} 0 & \text{if } x \neq y \\ 1 & \text{if } x = y \end{cases}$$

are linearly independent in $F(X,\mathbb{F})$, i.e. $\left\{\delta_x\middle| x \in X\right\}$ is linearly independent.

Remark 2.3.3. Note that the zero element of a vector space is never in a linearly independent set.

Proposition 2.3.4. *Let $S_1 \subset S_2 \subset V$.*

(1) If S_1 is linearly dependent, then so is S_2.

(2) If S_2 is linearly independent, then so is S_1.

Proof.

(1) If S_1 is linearly dependent, then there is a finite subset $\{v_1, \dots, v_n\} \subset S_1$ and scalars $\lambda_1, \dots, \lambda_n$ not all zero such that

$$\sum_{i=1}^{n} \lambda_i v_i = 0.$$

Then as $\{v_1, \dots, v_n\}$ is a finite subset of S_2, S_2 is not linearly independent.

(2) Let $\{v_1, \dots, v_n\} \subset S_1$, and suppose

$$\sum_{i=1}^{n} \lambda_i v_i = 0.$$

If S_2 is infinite, then $\lambda_i = 0$ for all i as $\{v_1, \dots, v_n\}$ is a finite subset of S_2. If S_2 is finite, then we have $S_2 = S_1 \cup \{w_1, \dots, w_m\}$, and we have that

$$\sum_{i=1}^{n} \lambda_i v_i + \sum_{j=1}^{m} \mu_j w_j = 0 \text{ with } \mu_j = 0 \text{ for all } j = 1, \dots, m,$$

so $\lambda_i = 0$ for all $i = 1, \dots, n$. □

Remark 2.3.5. Note that in the proof of 2.3.4 (1), we have scalars $\lambda_1, \dots, \lambda_n$ which are not all zero. This means that there is a $\lambda_i \neq 0$ for some $i \in \{1, \dots, n\}$. This is different than if $\lambda_1, \dots, \lambda_n$ were all not zero. The order of "not" and "all" is extremely important, and it is often confused by many beginning students.

Proposition 2.3.6. *Suppose S is a linearly independent subset of V, and suppose $v \notin$ span(S). Then $S \cup \{v\}$ is linearly independent.*

Proof. Let $\{v_1, \dots, v_n\}$ be a finite subset of S. Then we know

$$\sum_{i=1}^{n} \lambda_i v_i = 0 \Longrightarrow \lambda_i = 0 \text{ for all } i.$$

Now suppose

$$\mu v + \sum_{i=1}^{n} \mu_i v_i = 0.$$

If $\mu \neq 0$, then we have

$$v = \frac{-1}{\mu}\sum_{i=1}^{n}\mu_i v_i = \sum_{i=1}^{n}\frac{-\mu_i}{\mu}v_i.$$

Hence v is a linear combination of the v_i's, and $v \in \operatorname{span}(S)$, a contradiction. Hence $\mu = 0$, and all $\mu_i = 0$. □

Proposition 2.3.7. *Let $S \subset V$ be a linearly independent, and let $v \in \operatorname{span}(S) \setminus \{0\}$. Then there are unique $v_1, \dots, v_n \in S$ and $\lambda_1, \dots, \lambda_n \in \mathbb{F} \setminus \{0\}$ such that*

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

Proof. As $v \in \operatorname{span}(S)$, v can be written as a linear combination of elements of S. As $v \neq 0$, there are $v_1, \dots, v_n \in S$ and $\lambda_1, \dots, \lambda_n \in \mathbb{F} \setminus \{0\}$ such that

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

Suppose there are $w_1, \dots, w_m \in S$ and $\mu_1, \dots, \mu_m \in \mathbb{F} \setminus \{0\}$ such that

$$v = \sum_{j=1}^{m}\mu_j w_j.$$

Let $T = \{v_1, \dots, v_n\} \cap \{w_1, \dots, w_m\}$. If $T = \emptyset$, then

$$0 = v - v = \sum_{i=1}^{n}\lambda_i v_i - \sum_{j=1}^{m}\mu_j w_j,$$

so $\lambda_i = 0$ for all $i \in [n]$ and $\mu_j = 0$ for all $j \in [m]$ as S is linearly independent, a contradiction. Let $\{u_1, \dots, u_k\} = \{v_1, \dots, v_n\} \cap \{w_1, \dots, w_m\}$. After reindexing, we may assume $u_i = v_i = w_i$ for all $i \in [k]$. Thus

$$0 = v - v = \sum_{i=1}^{k}\lambda_i v_i + \sum_{i=k+1}^{n}\lambda_i v_i - \sum_{j=1}^{k}\mu_j w_j - \sum_{j=k+1}^{m}\mu_j w_j = \sum_{i=1}^{k}(\lambda_i - \mu_i)u_i + \sum_{i=k+1}^{n}\lambda_i v_i - \sum_{j=k+1}^{m}\mu_j w_j,$$

so $\lambda_i = \mu_i$ for all $i \in [k]$, $\lambda_i = 0$ for all $i = k+1, \dots, n$, and $\mu_j = 0$ for all $j = k+1, \dots, m$, so we have that $k = n = m$, and the expression is unique. □

Definition 2.3.8. A subset $B \subset V$ is called a basis for V if B is linearly independent and $V = \operatorname{span}(B)$ (B spans V).

Examples 2.3.9.

(1) The set $\{e_i \,|\, i \in [n]\}$ is the standard basis of $\mathbb{F}^n$.

(2) The matrices $E^{i,j}$ which have as entries all zeroes except a 1 in the $(i,j)^{\text{th}}$ entry form the standard basis of $M_{m\times n}(\mathbb{F})$.

(3) The functions $\delta_x\colon X\to\mathbb{F}$ given by

$$\delta_x(y)=\begin{cases}0 & \text{if } x\neq y\\ 1 & \text{if } x=y\end{cases}$$

form a basis of $F(X,\mathbb{F})$ if and only if X is finite. Note that the function $f(x)=1$ for all $x\in X$ is not a finite linear combination of these functions if X is infinite.

Definition 2.3.10. A vector space is finitely generated if there is a finite subset $S\subset V$ such that $\operatorname{span}(S)=V$. Such a set S is called a finite generating set for V.

Examples 2.3.11.

(1) Any vector space with a finite basis is finitely generated. For example, $\mathbb{F}^n$ and $M_{m\times n}(\mathbb{F})$ are finitely generated.

(2) The matrices E_{ij} which have as entries all zeroes except a 1 in the ij^{th} entry form the standard basis of $M_{m\times n}(\mathbb{F})$.

(3) We will see shortly that $C(a,b)$ for $a<b$ is not finitely generated.

Exercises

Exercise 2.3.12 (Symmetric and Exterior Algebras)**.**

(1) Find a basis for the following real vector spaces:

  (a) $S^2(\mathbb{R}^n)=\left\{A\in M_n(\mathbb{R})\middle|A=A^T\right\}$, i.e. the real symmetric matrices and

  (b) $\bigwedge^2(\mathbb{R}^n)=\left\{A\in M_n(\mathbb{R})\middle|A=-A^T\right\}$, ie. the real antisymmetric (or skew-symmetric) matrices.

(2) Show that $M_n(\mathbb{R})=S^2(\mathbb{R}^n)\oplus\bigwedge^2(\mathbb{R}^n)$.

Exercise 2.3.13 (Even and Odd Functions)**.** A function $f\in\mathbb{C}(\mathbb{R},\mathbb{R})$ is called even if $f(x)=f(-x)$ for all $x\in\mathbb{R}$ and odd if $f(x)=-f(-x)$ for all $x\in\mathbb{R}$. Show that

(1) the subset of even, respectively odd, functions is a subspace of $C(\mathbb{R},\mathbb{R})$ and

(2) $C(\mathbb{R},\mathbb{R})=\{\text{even functions}\}\oplus\{\text{odd functions}\}$.

Exercise 2.3.14. Let $S_1,S_2\subset V$ be subsets of V.

(1) What is $\operatorname{span}(\operatorname{span}(S_1))$?

(2) Show that if $S_1\subset S_2$, then $\operatorname{span}(S_1)\subset\operatorname{span}(S_2)$.

(3) Suppose $S_1\amalg S_2$ is linearly independent. Show

$$\operatorname{span}(S_1\amalg S_2)=\operatorname{span}(S_1)\oplus\operatorname{span}(S_2).$$

Exercise 2.3.15.

(1) Let $B = \{v_1, \ldots, v_n\}$ be a basis for V . Suppose W is a k-dimensional subspace of V. Show that for any subset $\{v_{i_1}, \ldots, v_{i_m}\}$ of B with $m > n - k$, there is a nonzero vector $w \in W$ that is a linear combination of $\{v_{i_1}, \ldots, v_{i_m}\}$.

(2) Let W be a subspace of V having the property that there exists a unique subspace W' such that $V = W \oplus W'$. Show that $W = V$ or $W = (0)$.

2.4 Finitely Generated Vector Spaces and Dimension

For this section, V will denote a finitely generated vector space over $\mathbb{F}$.

Lemma 2.4.1. *Let $A \in M_{m\times n}(F)$ be a matrix.*

(1) $y \in CS(A) \subset \mathbb{F}^m$ if and only if there is an $x \in \mathbb{F}^n$ such that $Ax = y$.

(2) If $v_1, \ldots, v_n$ are n vectors in $\mathbb{F}^m$ with $m < n$, then $\{v_1, \ldots, v_n\}$ is linearly dependent.

Proof.

(1) Let $A_1, \ldots, A_n$ be the columns of A. We have

$$\begin{aligned} y \in CS(A) &\iff \text{ there are } \lambda_1, \ldots, \lambda_n \text{ such that } y = \sum_{i=1}^{n} \lambda_i A_i \\ &\iff \text{ there are } \lambda_1, \ldots, \lambda_n \text{ such that if } x = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}, \text{ then } Ax = y \\ &\iff \text{ there is an } x \in \mathbb{F}^n \text{ such that } Ax = y. \end{aligned}$$

(2) We must show there are scalars $\lambda_1, \ldots, \lambda_n$, not all zero, such that

$$\sum_{i=1}^{n} \lambda_i v_i = 0.$$

Form a matrix $A \in M_{m\times n}(\mathbb{F})$ by letting the i^{th} column be v_i: $A = [v_1|v_2|\cdots|v_m]$. By (1), the above condition is now equivalent to finding a nonzero $x \in \mathbb{F}^n$ such that $Ax = 0$. To solve this system of linear equations, we augment the matrix so it has $n+1$ columns by letting the last column be all zeroes. Performing Gaussian elimination, we get a row reduced matrix U which is row equivalent to the matrix $B = [A|0]$. Now, the number of pivots (the first entry of a row that is nonzero) of U must be less than or equal to m as there are only m rows. Hence, at least one of the first n columns does not have a pivot in it. These columns correspond to free variables when we are looking for solutions of the original system of linear equations, i.e. there is a nonzero x such that $Ax = 0$. □

Theorem 2.4.2. *Suppose S_1 is a finite set that spans V and $S_2 \subset V$ is linearly independent. Then S_2 is finite and $|S_2| \leq |S_1|$.*

Proof. Let $S_1 = \{v_1, \ldots, v_m\}$ and $\{w_1, \ldots, w_n\} \subset S_2$. We show $n \leq m$. Since S_1 spans V, there are scalars $\lambda_{i,j} \in \mathbb{F}$ such that

$$w_j = \sum_{i=1}^{m} \lambda_{i,j} v_i \text{ for all } j = 1, \ldots, n.$$

Let $A \in M_{m \times n}(\mathbb{F})$ by $A_{i,j} = \lambda_{i,j}$. If $m > n$, then the rows of A are linearly dependent by 2.4.1, so there is an $x = (\mu_1, \ldots, \mu_n)^T \in \mathbb{R}^n \setminus \{0\}$ such that $Ax = 0$. This means that

$$\sum_{j=1}^{n} A_{i,j}\mu_j = \sum_{j=1}^{n} \lambda_{i,j}\mu_j = 0 \text{ for all } i \in [m].$$

This implies that

$$\sum_{j=1}^{n} \mu_j w_j = \sum_{j=1}^{n} \mu_j \sum_{i=1}^{m} \lambda_{i,j} v_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \lambda_{i,j}\mu_j \right) v_i = 0,$$

a contradiction as $\{w_1, \ldots, w_n\}$ is linearly independent. Hence $n \leq m$. □

Corollary 2.4.3. *Suppose V has a finite basis B with n elements. Then every basis of V has n elements.*

Proof. Let B' be another basis of V. Then by 2.4.2, B' is finite and $|B'| \leq |B|$. Applying 2.4.2 after switching B, B' yields $|B| \leq |B'|$, and we are finished. □

Corollary 2.4.4. *Every infinite subset of a finitely generated vector space is linearly dependent.*

Definition 2.4.5. The vector space V is finite dimensional if there is a finite subset $B \subset V$ such that B is a basis for V. The number of elements of B is the dimension of V, denoted $\dim(V)$. A vector space that is not finite dimensional is infinite dimensional. In this case, we will write $\dim(V) = \infty$.

Examples 2.4.6.

(1) $M_{m \times n}(\mathbb{F})$ has dimension mn.

(2) If $a \neq b$, then $C(a, b)$ and $C^n(a, b)$ are infinite dimensional for all $n \in \mathbb{N} \cup \{\infty\}$.

(3) $F(X, \mathbb{F})$ is finite dimensional if and only if X is finite.

Theorem 2.4.7 (Contraction)**.** *Let $S \subset V$ be a finite subset such that S spans V.*

(1) There is a subset $B \subset S$ such that B is a basis of V.

(2) If $\dim(V) = n$ *and* S *has* n *elements, then* S *is a basis for* V*.*

Proof.

(1) IfS is not a basis, then S is not linearly independent. Thus, there is some vector $v \in S$ such that v is a linear combination of the other elements of S. Set $S_1 = S \setminus \{v\}$. It is clear that S_1 spans V. If S_1 is not a basis, repeat the process to get S_2. Repeating this process as many times as necessary, since S is finite, we will eventually get that S_n is linearly independent for some $n \in \mathbb{N}$, and thus a basis for V.

(2) Suppose S is not a basis for V. Then by (1), there is a proper subset T of S such that T is a basis for V. But T has fewer elements than S, a contradiction to 2.4.3. □

Corollary 2.4.8 (Existence of Bases)**.** *Let* V *be a finitely generated vector space. Then* V *has a basis.*

Proof. This follows immediately from 2.4.7 (1). □

Corollary 2.4.9. *The vector space* V *is finitely generated if and only if* V *is finite dimensional.*

Proof. By 2.4.8, we see that a finitely generated vector space has a finite basis. The other direction is trivial. □

Lemma 2.4.10. *A subspace* W *of a finite dimensional vector space* V *is finitely generated.*

Proof. We construct a maximal linearly independent subset S of W as as follows: if $W = (0)$, we are finished. Otherwise, choose $w_1 \in W \setminus \{0\}$, and set $S_1 = \{w_1\}$, and note that S_1 is linearly independent by 2.3.6. If $W_1 = \operatorname{span}(S_1) = W$, we are finished. Otherwise, choose $w_2 \in W \setminus W_1$, and set $S_2 = \{w_1, w_2\}$, which is again linearly independent by 2.3.6. If $W_2 = \operatorname{span}(S_2) = W$, we are finished. Otherwise we may repeat this process. This algorithm will terminate as all linearly independent subsets of W must have less than or equal to $\dim(V)$ elements by 2.4.2. Now it is clear by 2.3.6 that our maximal linearly independent set $S \subset W$ must be a basis for W by 2.3.6. Hence W is finitely generated. □

Theorem 2.4.11 (Extension)**.** *Let* W *be a subspace of* V*. Let* B *be a basis of* W *(we know one exists by 2.4.8 and 2.4.10).*

(1) If $|B| = \dim(V)$*, then* $W = V$*.*

(2) There is a basis C *of* V *such that* $B \subset C$*.*

Proof.

(1) Suppose $B = \{w_1, \dots, w_n\}$ and $\dim(V) = n$. Suppose $W \neq V$. Then there is a $w_{n+1} \in V \setminus W$, and $B_1 = B \cup \{w_{n+1}\}$ is linearly independent by 2.3.6. This is a contradiction to 2.4.2 as every linearly independent subset of V must have less than or equal to n elements. Hence $W = V$.

(2) If $\operatorname{span}(B) \neq V$, then pick $v_1 \in V \setminus \operatorname{span}(B)$. It is immediate from 2.3.6 that $B_1 = B \cup \{v_1\}$ is linearly independent. If $\operatorname{span}(B_1) \neq V$, then we may pick $v_2 \in V \setminus \operatorname{span}(B_1)$, and once again by 2.3.6, $B_2 = B_1 \cup \{v_2\}$ is linearly independent. Repeating this process as many times as necessary, we will have that eventually B_n will have $\dim(V)$ elements and be linearly independent, so its span must be V by (1). □

Corollary 2.4.12. *Let W be a subspace of V, and let B be a basis of W. If $\dim(V) = n < \infty$, then $\dim(W) \leq \dim(V)$. If in addition W is a proper subspace, then $\dim(W) < \dim(V)$.*

Proof. This is immediate from 2.4.11. □

Proposition 2.4.13. *Suppose V is finite dimensional and $W_1, \dots, W_n$ are subspaces of V such that*

$$V = \bigoplus_{i=1}^{n} W_i.$$

(1) For $i = 1, \dots, n$, let $n_i = \dim(W_i)$ and let $B_i = \{v_1^i, \dots, v_{n_i}^u\}$ be a basis for W_i. Then $B = \coprod_{i=1}^{n} B_i$ is a basis for V.

(2) $\dim(V) = \sum_{i=1}^{n} \dim(W_i)$.

Proof.

(1) It is clear that B spans V. We must show it is linearly independent. Suppose

$$\sum_{i=1}^{n} \sum_{j=1}^{n_i} \lambda_j^i v_j^i = 0.$$

Then setting $w_j = \sum_{j=1}^{n_i} \lambda_j^i w_j^i$, we have $\sum_{i=1}^{n} w_i = 0$, so $w_i = 0$ for all $i \in [n]$ by 2.2.8. Now since B_i is a basis for all $i \in [n]$, $\lambda_j^i = 0$ for all i, j.

(2) This is immediate from (1). □

Exercises

Exercise 2.4.14 (Dimension Formulas)**.** Let W_1, W_2 be subspaces of a vector space V.

(1) Show that $\dim(W_1 + W_2) + \dim(W_1 \cap W_2) = \dim(W_1) + \dim(W_2)$.

(2) Suppose $W_1 \cap W_2 = (0)$. Find an expression for $\dim(W_1 \oplus W_2)$ similar to the formula found in (1).

(3) Let W be a subspace of V, and suppose $V = W_1 \oplus W_2$. Show that if $W_1 \subset W$ or $W_2 \subset W$, then

$$W = (W \cap W_1) \oplus (W \cap W_2).$$

Is this still true if we omit the condition "$W_1 \subset W$ or $W_2 \subset W$"?

Exercise 2.4.15 (Complexification 2)**.** Let V be a vector space over $\mathbb{R}$.

(1) Show that $\dim_{\mathbb{R}}(V) = \dim_{\mathbb{C}}(V_{\mathbb{C}})$. Deduce that $\dim_{\mathbb{R}}(V_{\mathbb{C}}) = 2\dim_{\mathbb{R}}(V)$.

(2) If $T \in L(V)$, we define the complexification of the operator T to be the operator $T_{\mathbb{C}} \in L(V_{\mathbb{C}})$ given by

$$T_{\mathbb{C}}(u + iv) = Tu + iTv.$$

(a) Show that the complexicifation of multiplication by $A \in M_n(\mathbb{R})$ on $\mathbb{R}^n$ is multiplication by $A \in M_n(\mathbb{C})$ on $\mathbb{C}^n$.

(b) Show that the complexification of multiplication by $f \in F(X, \mathbb{R})$ on $F(X, \mathbb{R})$ is multiplication by $f \in F(X, \mathbb{C})$ on $F(X, \mathbb{C})$.

2.5 Existence of Bases

A question now arises: do all vector spaces have bases? The answer to this question is yes as we will see shortly. We saw in the previous section that this result is very easy for finitely generated vector spaces. However, to prove this in full generality, we will need Zorn's Lemma which is logically equivalent to the Axiom of Choice, i.e. assuming one, we can prove the other. We will not prove the equivalence of these two statements as it is beyond the scope of this course. For this section, V will denote a vector space over $\mathbb{F}$ (which is not necessarily finitely generated).

Definition 2.5.1.

(1) A partial order on the set P is a subset $\mathcal{R}$ of $P \times P$ (a relation $\mathcal{R}$ on P), such that

(i) (reflexive) $(p, p) \in \mathcal{R}$ for all $p \in P$,

(ii) (antisymmetric) $(p, q) \in \mathcal{R}$ and $(q, p) \in \mathcal{R}$, then $p = q$, and

(iii) (transitive) $(p, q), (q, r) \in \mathcal{R}$ implies $(p, r) \in \mathcal{R}$.

Usually, we denote $(p, q) \in \mathcal{R}$ as $p \leq q$. Hence, we may restate these conditions as:

(i) $p \leq p$ for all $p \in P$,

(ii) $p \leq q$ and $q \leq p$ implies $p = q$, and

(iii) $p \leq q$ and $q \leq r$ implies $p \leq r$.

Note that we may not always compare $p, q \in P$. In this sense the order is partial. The pair $(P, \leq)$, or sometimes P if $\leq$ is understood, is called a partially ordered set.

(2) Suppose P is partially ordered set. The subset T is totally ordered if for every $s, t \in T$, we have either $s \leq t$ or $t \leq s$. Note that we can compare all elements of a totally ordered set.

(3) An upper bound for a totally ordered subset $T \subset P$, a partially ordered set, is an element $x \in P$ such that $t \leq x$ for all $t \in T$.

(4) An element $m \in P$ is called maximal if $p \in P$ with $m \leq p$ implies $p = m$.

Examples 2.5.2.

(1) Let X be a set, and let $\mathcal{P}(X)$ be the power set of X, i.e. the set of all subsets of X. Then we may partially order $\mathcal{P}(X)$ by inclusion, i.e. we say $S_1 \leq S_2$ for $S_1, S_2 \in \mathcal{P}(X)$ if $S_1 \subseteq S_2$. Note that this is not a total order unless X has only one element. Furthermore, note that $\mathcal{P}(X)$ has a unique maximal element: X.

(2) We may partially order $\mathcal{P}(X)$ by reverse inclusion, i.e. we say $S_2 \leq S_1$ for $S_1, S_2 \in \mathcal{P}(X)$ if $S_1 \subseteq S_2$. Note once again that this is not a total order unless X has one element. Also, $\mathcal{P}(X)$ has a unique maximal element for this order as well: $\emptyset$.

Lemma 2.5.3 (Zorn)**.** *Suppose every nonempty totally ordered subset T of a nonempty partially ordered set P has an upper bound. Then P has a maximal element.*

Remark 2.5.4. Note that the maximal element may not be (and probably is not) unique.

Theorem 2.5.5 (Existence of Bases)**.** *Let V be a vector space. Then V has a basis.*

Proof. Let P be the set of all linearly independent subsets of V. We partially order P by inclusion. Let T be a totally ordered subset of P. We show that T has an upper bound. Our candidate for an upper bound will be

$$X = \bigcup_{t \in T} t,$$

the union of all sets contained in T.

We must show $X \in P$, i.e. X is linearly independent. Let $Y = \{y_1, \ldots, y_n\}$ be a finite subset of X. Then there are $t_1, \ldots, t_n \in T$ such that $y_i \in t_i$ for all $i = 1, \ldots, n$. Since T is totally ordered, one of the t_i's contains the others. Call this set s. Then $Y \subset s$, but s is linearly independent. Hence Y is linearly independent, and thus so is X.

We must show that X is indeed an upper bound for T. This is obvious as $t \subset X$ for all $t \in T$. Hence $t \leq X$ for all $t \in T$.

Now, we invoke Zorn's lemma to get a maximal element B of P. We claim that B is a basis for V, i.e. it spans V (B is linearly independent as it is in P). Suppose not. Then there is some $v \in V \setminus \operatorname{span}(B)$. Then $B \cup \{v\}$ is linearly independent, and $B \subset B \cup \{v\}$. Hence, $B \leq B \cup \{v\}$, but $B \neq B \cup \{v\}$. This is a contradiction to the maximality of B. Hence B is a basis for V. □

Theorem 2.5.6 (Extension)**.** *Let W be a subspace of V. Let B be a basis of W. Then there is a basis C of V such that $B \subset C$.*

Proof. Let P be the set whose elements are linearly independent subsets of V containing B partially ordered by inclusion. As in the proof of 2.5.5, we see that each totally ordered subset has an upper bound, so P has a maximal element $C \in P$ (which means $B \subset C$). Once again, as in the proof of 2.5.5, we must have that C is a basis for V. □

Theorem 2.5.7 (Contraction)**.** *Let $S \subset V$ be a subset such that S spans V. Then there is a subset $B \subset S$ such that B is a basis of V.*

Proof. Let P be the set whose elements are linearly independent subsets of S partially ordered by inclusion. As in the proof of 2.5.5, we see that each totally ordered subset has an upper bound, so P has a maximal element $B \in P$ (which means $B \subset S$). Once again, as in the proof of 2.5.5, we must have that B is a basis for V. □

Proposition 2.5.8. *Let $W \subset V$ be a subspace. Then there is a (non-unique) subspace U of V such that $V = U \oplus W$.*

Proof. This is merely a restatement of 2.5.6. Let B be a basis of W, and extend B to a basis C of V. Set $U = \operatorname{span}(C \setminus B)$. Then clearly $U \cap W = (0)$ and $U + W = V$, so $V = U \oplus W$ by 2.2.8. Note that if V is finitely generated, we may use 2.4.11 instead of 2.5.6. □

Exercises

Exercise 2.5.9.

Chapter 3

Linear Transformations

For this chapter, unless stated otherwise, V, W will be vector spaces over $\mathbb{F}$.

3.1 Linear Transformations

Definition 3.1.1. A linear transformation from the vector space V to the vector space W, denoted $T\colon V \to W$, is a function from V to W such that

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

for all $\lambda \in \mathbb{F}$ and $u, v \in V$. This condition is called $\mathbb{F}$-linearity. Often $T(v)$ is denoted Tv. Sometimes we will refer to a linear transformation as a linear operator, an operator, or a map. The set of all linear transformations from V to W is denoted $L(V, W)$. If $V = W$, we write $L(V) = L(V, V)$.

Examples 3.1.2.

(1) There is a zero linear transformation $0\colon V \to W$ given by $0(v) = 0 \in W$ for all $v \in V$. There is an identity linear transformation $I \in L(V)$ given by $Iv = v$ for all $v \in V$.

(2) Let $A \in M_{m\times n}(\mathbb{F})$. Then left multiplication by A defines a linear transformation $L_A\colon \mathbb{F}^n \to \mathbb{F}^m$ given by $x \mapsto Ax$.

(3) The projection maps $\mathbb{F}^n \to \mathbb{F}$ given by $e_i^*(\lambda_1, \dots, \lambda_n)^T = \lambda_i$ for $i \in [n]$ are $\mathbb{F}$-linear.

(4) Integration from a to b where $a, b \in \mathbb{R}$ is a map $C[a, b] \to \mathbb{R}$.

(5) The derivative map $D\colon C^n(a, b) \to C^{n-1}(a, b)$ given by $f \mapsto f'$ is an operator.

(6) Let $x \in X$, a set. Then evaluation at x is a linear transformation $\mathrm{ev}_x\colon F(X, \mathbb{F}) \to \mathbb{F}$ given by $f \mapsto f(x)$.

(7) Suppose $B = \{v_1, \dots, v_n\}$ is a basis for V. Then the projection maps $v_j^*\colon V \to \mathbb{F}$ given by

$$v_j^*\left(\sum_{i=1}^{n} \lambda_i v_i\right) = \lambda_j$$

are linear transformations.

Definition 3.1.3. Let V be a vector space over $\mathbb{R}$, and let $V_{\mathbb{C}}$ be as in 2.1.12. If $T \in L(V)$, we define the complexification of the operator T to be the operator $T_{\mathbb{C}} \in L(V_{\mathbb{C}})$ given by

$$T_{\mathbb{C}}(u + iv) = Tu + iTv.$$

Examples 3.1.4.

(1) The complexification of multiplication by the matrix A on $\mathbb{R}^n$ is multiplication by the matrix A on $\mathbb{C}^n$.

(2) The complexification of multiplication by x on $\mathbb{R}[x]$ is multiplication by x on $\mathbb{C}[x]$.

Remarks 3.1.5.

(1) Note that a linear transformation is completely determined by what it does to a basis. If $\{v_i\}$ is a basis for V and we know what Tv_i is for all i, then by linearity, we know Tv for any vector $v \in V$. Hence, to define a linear transformation, one only needs to specify where a basis goes, and then we may "extend it by linearity," i.e. if $\{v_i\}$ is a basis for V, we specify Tv_i for all i, and then we decree that

$$T\left(\sum_{j=1}^{n} \lambda_j v_j\right) = \sum_{j=1}^{n} \lambda_j T v_j$$

for all finite subsets $\{v_1, \dots, v_n\} \subset \{v_i\}$. This rule defines T on all of V as $\{v_i\}$ spans V, and T is well defined since $\{v_i\}$ is linearly independent.

For example, in example 7 above, v_j^* is the unique map that sends v_j to one, and v_i to zero for $i \neq j$.

(2) If T_1 and T_2 are linear transformations $V \to W$ and $\lambda \in \mathbb{F}$, then we may define $T_1 + T_2 \colon V \to W$ by $(T_1 + T_2)v = T_1 v + T_2 v$, and we may define $\lambda T_1 \colon V \to W$ by $(\lambda T_1)v = \lambda(T_1 v)$. Hence, if V and W are vector spaces over $\mathbb{F}$, then $L(V, W)$, the set of all $\mathbb{F}$-linear transformations $V \to W$, is a vector space (it is easy to see that $+$ is an addition and $\cdot$ is a scalar multiplication which satisfy the distributive property, each linear transformation has an additive inverse, and there is a zero linear transformation).

Example 3.1.6. The map $L \colon M_{m\times n}(\mathbb{F}) \to L(\mathbb{F}^n, \mathbb{F}^m)$ given by $A \mapsto L_A$ is a linear transformation.

Exercises

Exercise 3.1.7. Let $v \in V$. Show that $\operatorname{ev}_v \colon L(V, W) \to W$ given by $T \mapsto Tv$ is a linear operator.

3.2 Kernel and Image

Note that we may talk about injectivity and surjectivity of operators as they are functions. The immediate question to ask is, “Why should we care about linear transformations?” The answer is that these maps are the “natural” maps to consider when we are thinking about vector spaces. As we shall see, linear transformations preserve vector space structure in the sense of the following definition and proposition:

Definition 3.2.1. Let $T \in L(V, W)$. Let the kernel of T, denoted $\ker(T)$, be $\big\{v \in V \big| Tv = 0\big\}$ and let the image, or range, of T, denoted $\operatorname{im}(T)$ or TV, be $\big\{w \in W \big| w = Tv \text{ for some } v \in V\big\}$.

Examples 3.2.2.

(1) Let $A \in M_{m\times n}(\mathbb{F})$ and consider L_A. Then $\ker(L_A) = NS(A)$ and $\operatorname{im}(L_A) = CS(A)$.

(2) Consider $\operatorname{ev}_x \colon F(X, \mathbb{F}) \to \mathbb{F}$. Then $\ker(\operatorname{ev}_x) = \big\{f \colon X \to \mathbb{F} \big| f(x) = 0\big\}$ and $\operatorname{im}(\operatorname{ev}_x) = \mathbb{F}$ as $\lambda\delta_x \mapsto \lambda$.

(3) If $B = \{v_1, \dots, v_n\}$ is a basis for V, then $\ker(v_1^*) = \operatorname{span}\{v_2, \dots, v_n\}$ and $\operatorname{im}(v_1^*) = \mathbb{F}$.

(4) Let $D \colon C^1[0, 1] \to C[0, 1]$ be the derivative map $f \mapsto f'$. Then $\ker(D) = \{\text{constant functions}\}$ and $\operatorname{im}(D) = C[0, 1]$ as

$$D \int\limits_0^x f(t)\ dt = f(x) \text{ for all } x \in [0, 1]$$

by the fundamental theorem of calculus, i.e. D has a right inverse.

Proposition 3.2.3. *Let $T \in L(V, W)$. $\ker(T)$ and $\operatorname{im}(T)$ are vector spaces.*

Proof. It suffices to show they are closed under addition and scalar multiplication. Clearly $Tv = 0 = Tu$ implies $T(\lambda u + v) = 0$ for all $u, v \in \ker(T)$ and $\lambda \in \mathbb{F}$, so $\ker(T)$ is a subspace of V, and thus a vector space. Let $y, z \in \operatorname{im}(T)$ and $\mu \in \mathbb{F}$, and let $w, x \in V$ such that $Tw = y$ and $Tx = z$. Then $T(\lambda w + x) = \lambda y + z \in \operatorname{im}(T)$, so $\operatorname{im}(T)$ is a subspace of W, and thus a vector space. □

Proposition 3.2.4. *$T \in L(V, W)$ is injective if and only if $\ker(T) = (0)$.*

Proof. Suppose T is injective. Then since $T0 = 0$, if $Tv = 0$, then $v = 0$. Suppose now that $\ker(T) = (0)$. Let $x, y \in V$ such that $Tx = Ty$. Then $T(x - y) = 0$, so $x - y = 0$ and $x = y$. Hence T is injective. □

Proposition 3.2.5. *Suppose $T \in L(V, W)$ is bijective, and let $\{v_i\}$ be a basis for V. Then $\{Tv_i\}$ is a basis for W.*

Proof. We show $\{Tv_i\}$ is linearly independent. Suppose $\{Tv_1, \dots, Tv_n\}$ is a finite subset of $\{Tv_i\}$, and suppose

$$\sum_{j=1}^n \lambda_j T v_j = T\left(\sum_{j=1}^n \lambda_j v_j\right) = 0.$$

Then since T is injective, by 3.2.4

$$\sum_{j=1}^{n} \lambda_j v_j = 0.$$

Now since $\{v_i\}$ is linearly independent, we have $\lambda_j = 0$ for all $j = 1, \ldots, n$.

We show $\{Tv_i\}$ spans W. Let $w \in W$. Then there is a $v \in V$ such that $Tv = w$ as T is surjective. Then $v \in \operatorname{span}\{v_i\}$, so there are $v_1, \ldots, v_n \in \{v_i\}$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ such that

$$v = \sum_{j=1}^{n} \lambda_j v_j.$$

Then

$$w = Tv = T \sum_{j=1}^{n} \lambda_j v_j = \sum_{j=1}^{n} \lambda_j T v_j,$$

so $w \in \operatorname{span}\{Tv_i\}$. □

Definition 3.2.6. An operator $T \in L(V, W)$ is called invertible, or an isomorphism (of vector spaces), if there is a linear operator $S \in L(W, V)$ such that $T \circ S = \mathrm{id}_W$ and $S \circ T = \mathrm{id}_V$, where id means the identity linear transformation. Often, when composing linear transformations, we will omit the $\circ$ and write ST and TS. We call vector spaces V, W isomorphic, denoted $V \cong W$, if there is an isomorphism $T \in L(V, W)$.

Examples 3.2.7.

(1) $\mathbb{F}^n \cong V$ if $\dim(V) = n$.

(2) $F(X, \mathbb{F}) \cong \mathbb{F}^n$ if $|X| = n$.

Proposition 3.2.8. *$T \in L(V, W)$ is invertible if and only if it is bijective.*

Proof. It is clear that an invertible operator is bijective.

Suppose T is bijective. Then there is an inverse function $S \colon W \to V$. It remains to show S is a linear transformation. Suppose $y, z \in W$ and $\lambda \in \mathbb{F}$. Then there are $w, x \in W$ such that $Tw = y$ and $Tx = z$. Then

$$S(\lambda y + z) = S(\lambda Tw + Tz) = ST(\lambda w + x) = \lambda w + x = \lambda STw + STx = \lambda Sy + Sz.$$

Hence S is a linear transformation. □

Remark 3.2.9. If T is an isomorphism, then T maps bases to bases. In this way, we can identify the domain and codomain of T as the same vector space. In particular, if $V \cong W$, then $\dim(V) = \dim(W)$ (we allow the possibility that both are ∞).

Proposition 3.2.10. *Suppose* $\dim(V) = \dim(W) = n < \infty$. *Then there is an isomorphism* $T \colon V \to W$.

Proof. Let $\{v_1, \dots, v_n\}$ be a basis for V and $\{w_1, \dots, w_n\}$ be a basis for W. Define a linear transformation $T \colon V \to W$ by $Tv_i = w_i$ for $i = 1, \dots, n$, and extend by linearity. It is easy to see that the inverse of T is the operator $S \colon W \to V$ defined by $Sw_i = v_i$ for all $i = 1, \dots, n$ and extending by linearity. □

Remark 3.2.11. We see that if $U \cong V$ via isomorphism T and $V \cong W$ via isomorphism S, then $U \cong W$ via isomorphism ST.

Our task is now to prove the Rank-Nullity Theorem. One particular application of this theorem is that injectivity, surjectivity, and bijectivity are all equivalent for $T \in L(V, W)$ when V, W are of the same finite dimension.

Lemma 3.2.12. *Let $T \in L(V, W)$, and let M be a subspace such that $V = \ker(T) \oplus M$ (one exists by 2.5.8).*

(1) $T|_M$ is injective.

(2) $\operatorname{im}(T|_M) = \operatorname{im}(T)$.

(3) $T|_M$ is an isomorphism $M \to \operatorname{im}(T)$.

Proof.

(1) Suppose $u, v \in M$ such that $Tu = Tv$. Then $T(u - v) = 0$, so $u - v \in M \cap \ker(T) = (0)$. Hence $u - v = 0$, and $u = v$.

(2) Let $B = \{u_1, \dots, u_n\}$ be a basis for $\ker(T)$, and let $C = \{w_1, \dots, w_m\}$ be a basis for M. Then by 2.4.13, $B \cup C$ is a basis for V. Hence $\{Tu_1, \dots, Tu_n, Tw_1, \dots, Tw_m\}$ spans $\operatorname{im}(T)$. But since $Tu_i = 0$ for all i, we have that $\{Tw_1, \dots, Tw_m\}$ spans $\operatorname{im}(T)$. Thus $\operatorname{im}(T) = \operatorname{im}(T|_M)$.

(3) By (1), $T|_M$ is injective, and by (2), $T|_M$ is surjective onto $\operatorname{im}(T)$. Hence it is bijective and thus an isomorphism. □

Theorem 3.2.13 (Rank-Nullity)**.** *Suppose V is finite dimensional, and let $T \in L(V, W)$. Then*

$$\dim(V) = \dim(\operatorname{im}(T)) + \dim(\ker(T)).$$

Proof. By 2.5.8 there is a subspace M of V such that $V = \ker(T) \oplus M$. By 2.4.13, $\dim(V) = \dim(\ker(T)) + \dim(M)$. We must show $\dim(M) = \dim(\operatorname{im}(T))$. This follows immediately from 3.2.12 and 3.2.9. □

Remark 3.2.14. Suppose V is finite dimensional, and let $T \in L(V, W)$. Then $\dim(\operatorname{im}(T))$ is sometimes called the rank of T, denoted $\operatorname{rank}(T)$, and $\dim(\ker(T))$ is sometimes called the nullity of T, denoted $\operatorname{nullity}(T)$. Using this terminology, 3.2.13 says that

$$\dim(V) = \operatorname{rank}(T) + \operatorname{nullity}(T),$$

which is how 3.2.13 gets the name "Rank-Nullity Theorem."

Corollary 3.2.15. *Let V, W be finite dimensional vector spaces with* $\dim(V) = \dim(W)$*, and let $T \in L(V, W)$. The following are equivalent:*

(1) T is injective,

(2) T is surjective, and

(3) T is bijective.

Proof.

(3) ⇒ (1): Obvious.

(1) ⇒ (2): Suppose T is injective. Then $\dim(\ker(T)) = 0$, so $\dim(V) = \dim(\mathrm{im}(T)) = \dim(W)$, and T is surjective.

(2) ⇒ (3): Suppose T is surjective. Then by similar reasoning, $\dim(\ker(T)) = 0$, so T is injective. Hence T is bijective. □

Exercises

Exercise 3.2.16. Let $T \in L(V, W)$, let $S = \{v_1, \dots, v_n\} \subset V$, and recall that $TS = \{Tv_1, \dots, Tv_n\} \subset W$.

(1) Prove or disprove the following statements:

(a) If S is linearly independent, then TS is linearly independent.

(b) If TS is linearly independent, then S is linearly independent.

(c) If S spans V, then TS spans W.

(d) If TS spans W, then S spans V.

(e) If S is a basis for V, then TS is a basis for W.

(f) If TS is a basis for W, then S is a basis for V.

(2) For each of the false statements above (if there are any), find a condition on T which makes the statement true.

Exercise 3.2.17 (Rank of a Matrix)**.**

(1) Let $A \in M_n(\mathbb{F})$. Prove that $\dim(CS(A)) = \dim(RS(A))$. This number is called the rank of A, denoted $\mathrm{rank}(A)$.

(2) Show that $\mathrm{rank}(A) = \mathrm{rank}(L_A)$.

Exercise 3.2.18. [Square Matrix Theorem] Prove the following (celebrated) theorem from matrix theory. For $A \in M_n(\mathbb{F})$, the following are equivalent:

(1) A is invertible,

(2) There is a $B \in M_n(\mathbb{F})$ such that $AB = I$,

(3) There is a $B \in M_n(\mathbb{F})$ such that $BA = I$,

(4) A is nonsingular, i.e. if $Ax = 0$ for $x \in \mathbb{F}^n$, then $x = 0$,

(5) For all $y \in \mathbb{F}^n$, there is a unique $x \in \mathbb{F}^n$ such that $Ax = y$,

(6) For all $y \in \mathbb{F}^n$, there is an $x \in \mathbb{F}^n$ such that $Ax = y$,

(7) $CS(A) = \mathbb{F}^n$,

(8) $RS(A) = \mathbb{F}^n$,

(9) $\operatorname{rank}(A) = n$, and

(10) A is row equivalent to I.

Exercise 3.2.19 (Quotient Spaces 2)**.** Let $W \subset V$ be a subspace.

(1) Show that the map $q\colon V \to V/W$ given by $v \mapsto v+W$ is a surjective linear transformation such that $\ker(q) = W$.

(2) Suppose V is finite dimensional. Using q, find $\dim(V/W)$ in terms of $\dim(V)$ and $\dim(W)$.

(3) Show that if $T \in L(V,U)$ and $W \subset \ker(T)$, then T factors uniquely through q, i.e. there is a unique linear transformation $\widetilde{T}\colon V/W \to U$ such that the diagram

$$\begin{tikzcd}
V \arrow[rr, "T"] \arrow[dr, "q"'] & & U \\
& V/W \arrow[ur, "\widetilde{T}"'] &
\end{tikzcd}$$

commutes, i.e. $T = \widetilde{T} \circ q$. Show that if $W = \ker(T)$, then $\widetilde{T}$ is injective.

3.3 Dual Spaces

We now study $L(V, \mathbb{F})$ in the case of a finite dimensional vector space V over $\mathbb{F}$.

Definition 3.3.1. Let V be a vector space over $\mathbb{F}$. The dual of V, denoted V^*, is the vector space of all linear transformations $V \to \mathbb{F}$, i.e. $V^* = L(V, \mathbb{F})$. Elements of V^* are called linear functionals.

Proposition 3.3.2. *Let V be a vector space over $\mathbb{F}$, and let $\varphi \in V^*$. Then either $\varphi = 0$ or φ is surjective.*

Proof. We know that $\dim(\operatorname{im}(\varphi)) \leq 1$. If it is 0, then $\varphi = 0$. If it is 1, then φ is surjective. □

Definition-Proposition 3.3.3. *Let $B = \{v_1, \dots, v_n\}$ be a basis for the finite dimensional vector space V. Then $B^* = \{v_1^*, \dots, v_n^*\}$ is a basis for V^* called the dual basis.*

Proof. We must show that B^* is a basis for V^*. First, we show B^* is linearly independent. Suppose

$$\sum_{i=1}^{n} \lambda_i v_i^* = 0,$$

the zero linear transformation. Then we have that

$$\left(\sum_{i=1}^{n} \lambda_i v_i^*\right) v_j = \lambda_j = 0$$

for all $j = 1, \ldots, n$. Thus, B^* is linearly independent. We show B^* spans V^*. Suppose $\varphi \in V^*$. Let $\lambda_j = \varphi(v_j)$ for $j = 1, \ldots, n$. Then we have that

$$\left(\varphi - \sum_{i=1}^{n} \lambda_i v_i^*\right) v_j = 0$$

for all $j = 1, \ldots, n$. Since a linear transformation is completely determined by its values on a basis, we have that

$$\varphi - \sum_{i=1}^{n} \lambda_i v_i^* = 0 \Longrightarrow \varphi = \sum_{i=1}^{n} \lambda_i v_i^*.$$

□

Remark 3.3.4. One of the most useful techniques to show linear independence is the "kill-off" method used in the previous proof:

$$\left(\sum_{i=1}^{n} \lambda_i v_i^*\right) v_j = \lambda_j = 0.$$

Applying the v_j allows us to see that each λ_j is zero. We will see this technique again when we discuss eigenvalues and eigenvectors.

Proposition 3.3.5. *Let V be a finite dimensional vector space over $\mathbb{F}$. There is a canonical isomorphism* $\mathrm{ev}\colon V \to V^{**} = (V^*)^*$ *given by* $v \mapsto \mathrm{ev}_v$*, the linear transformation* $V^* \to \mathbb{F}$ *given by the evaluation map:* $\mathrm{ev}_v(\varphi) = \varphi(v)$.

Proof. First, it is clear that ev is a linear transformation as $\mathrm{ev}_{\lambda u+v} = \lambda\,\mathrm{ev}_u + \mathrm{ev}_v$. Hence, we must show the map ev is bijective. Suppose $\mathrm{ev}_v = 0$. Then ev_v is the zero linear functional on V^*, i.e., for all $\varphi \in V^*$, $\varphi(v) = 0$. Since φ is completely determined by where it sends a basis of V, we know that $\varphi = 0$. Hence, the map ev is injective. Suppose now that $x \in V^{**}$. Let $B = \{v_1, \ldots, v_n\}$ be a basis for V and let $B^* = \{v_1^*, \ldots, v_n^*\}$ be the dual basis. Setting $\lambda_j = x(v_j^*)$ for $j = 1, \ldots, n$, we see that if

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

then $(x - \mathrm{ev}_u)v_j^* = 0$ for all $j = 1, \dots, n$. Once more since a linear transformation is completely determined by where it sends a basis, we have $x - \mathrm{ev}_u = 0$, so $x = \mathrm{ev}_u$ and ev is surjective. □

Remark 3.3.6. The word “canonical” here means “completely determined,” or “the best one.” It is independent of any choices of basis or vector in the space.

Proposition 3.3.7. *Let V be a finite dimensional vector space over $\mathbb{F}$. Then V is isomorphic to V^*, but not canonically.*

Proof. Let $B = \{v_1, \dots, v_n\}$ be a basis for V. Then B^* is a basis for V^*. Define a linear transformation $T\colon V \to V^*$ by $v_i \mapsto v_i^*$ for all $v_i \in B$, and extend this map by linearity. Then T is an isomorphism as there is the obvious inverse map. □

Remark 3.3.8. It is important to ask why there is no canonical isomorphism between V and V^*. The naive, but incorrect, thing to ask is whether $v \mapsto v^*$ would give a canonical isomorphism $V \to V^*$. Note that the definition of v^* requires a basis of V as in 3.3.3. We cannot define the coordinate projection v^* without a distinguished basis. For example, if $V = \mathbb{F}^2$ and $v = (1, 0)$, we can complete $\{v\}$ to a basis in many different ways. Let $B_1 = \{v, (0, 1)\}$ and $B_2 = \{v, (1, 1)\}$. Then we see that $v^*(1, 1) = 1$ relative to B_1, but $v^*(1, 1) = 0$ relative to B_2.

Exercises

Exercise 3.3.9 (Dual Spaces of Infinite Dimensional Spaces)**.** Suppose $B = \big\{v_n \big| n \in \mathbb{N}\big\}$ is a basis of V. Is it true that $V^* = \operatorname{span}\big\{v_n^* \big| n \in \mathbb{N}\big\}$?

3.4 Coordinates

For this section, V, W will denote finite dimensional vector spaced over $\mathbb{F}$. We now discuss the way in which all finite dimensional vector spaces over $\mathbb{F}$ look like $\mathbb{F}^n$ (non-uniquely). We then study $L(V, W)$, with the main result being that if $\dim(V) = n$ and $\dim(W) = m$, then $L(V, W) \cong M_{m\times n}(\mathbb{F})$ (non-uniquely) which is left to the reader as an exercise.

Definition 3.4.1. Let V be a finite dimensional vector space over $\mathbb{F}$, and let $B = (v_1, \dots, v_n)$ be an ordered basis for V. The map $[\cdot]_B\colon V \to \mathbb{R}^n$ given by

$$v = \sum_{i=1}^{n} \lambda_i v_i \longmapsto \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} v_1^*(v) \\ v_2^*(v) \\ \vdots \\ v_n^*(v) \end{pmatrix}$$

is called the coordinate map with respect to B. We say that $[v]_B$ are coordinates for v with respect to the basis B. Note that we denote an ordered basis as a sequence of vectors in parentheses rather than a set of vectors contained in curly brackets.

Proposition 3.4.2. *The coordinate map is an isomorphism.*

Proof. It is clear that the coordinate map is a linear transformation as $[\lambda u+v]_B = \lambda[u]_B+[v]_B$ for all $u, v \in V$ and $\lambda \in \mathbb{F}$. We must show $[\cdot]_B$ is bijective. We show it is injective. Suppose that $[v]_B = (0, \ldots, 0)^T$. Then since

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

for unique $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$, we must have that $\lambda_i = 0$ for all $i = 1, \ldots, n$, and $v = 0$. Hence the map is injective. We show the map is surjective. Suppose $(\lambda_1, \ldots, \lambda_n)^T \in \mathbb{F}^n$. Then define the vector u by

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

Then $[u]_B = (\lambda_1, \ldots, \lambda_n)^T$, so the map is surjective. □

Definition 3.4.3. Let V and W be finite dimensional vector spaces, let $B = (v_1, \ldots, v_n)$ be an ordered basis for V, let $C = (w_1, \ldots, w_m)$ be an ordered basis for W, and let $T \in L(V, W)$. The matrix $[T]_B^C \in M_{m\times n}(\mathbb{F})$ is given by

$$([T]_B^C)_{ij} = w_i^*(Tv_j) = e_i^*([Tv_j]_C)$$

where w_i^* is projection onto the i^{th} coordinate in W and e_i^* is the projection onto the i^{th} coordinate in $\mathbb{F}^m$.

We can also define $[T]_B^C$ as the matrix whose $(i,j)^{\text{th}}$ entry is $\lambda_{i,j}$ where the $\lambda_{i,j}$ are the unique elements of $\mathbb{F}$ such that

$$Tv_j = \sum_{i=1}^{m} \lambda_{i,j} w_i.$$

We can see this in terms of matrix augmentation as follows:

$$[T]_B^C = \left[[Tv_1]_C \middle| \cdots \middle| [Tv_n]_C \right].$$

The matrix $[T]_B^C$ is called coordinate matrix for T with respect to B and C. If $V = W$ and $B = C$, then we denote $[T]_B^B$ as $[T]_B$.

Remark 3.4.4. Note that the j^{th} column of $[T]_B^C$ is $[Tv_j]_C$.

Examples 3.4.5.

(1) Let $B = (v_1, \ldots, v_4)$ be an ordered basis for the finite dimensional vector space V. Let $S \in L(V, V)$ be given by $Sv_i = v_{i+1}$ for $i \neq 4$ and $Sv_4 = v_1$. Then

$$[S]_B = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

(2) Let S be as above, but consider the ordered basis $B' = (v_4, \dots, v_1)$. Then

$$[S]_{B'} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

(3) Let S be as above, but consider the ordered basis $B' = (v_3, v_2, v_4, v_1)$. Then

$$[S]_{B'} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Proposition 3.4.6. *Let V and W be finite dimensional vector spaces, let $B = (v_1, \dots, v_n)$ be an ordered basis for V, let $C = (w_1, \dots, w_m)$ be an ordered basis for W, and let $T \in L(V, W)$. Then the diagram*

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "{[\cdot]_B}"] & W \arrow[d, "{[\cdot]_C}"] \\ V \arrow[r, "{[T]_B^C}"] & W \end{tikzcd}$$

commutes, i.e. $[T]_B^C[v]_B = [Tv]_C$ for all $v \in V$.

Proof. To show both matrices are equal, we must show that their entries are equal. We know $([T]_B^C)_{i,j} = e_i^*[Tv_j]_C$. Hence

$$\begin{aligned} ([T]_B^C[v]_B)_{i,1} &= \sum_{j=1}^n ([T]_B^C)_{i,j}([v]_B)_{j,1} = \sum_{j=1}^n e_i^*[Tv_j]_C([v]_B)_{j,1} = e_i^* \sum_{j=1}^n ([v]_B)_{j,1}[Tv_j]_C \\ &= e_i^* \left[\sum_{j=1}^n ([v]_B)_{j,1} Tv_j \right]_C = e_i^* \left[T \sum_{j=1}^n ([v]_B)_{j,1} v_j \right]_C = e_i^*[Tv]_C = ([Tv]_C)_{i,1} \end{aligned}$$

as $[\cdot]_C$, e_i^*, and T are linear transformations, and

$$v = \sum_{j=1}^n ([v]_B)_{j1} v_j.$$

□

Remarks 3.4.7. When we proved 2.4.3 and 2.5.6 (1), we used the coordinate map implicitly. The algorithm of Gaussian elimination proves the hard result of 2.4.2, which in turn, proves these technical theorems after we identify the vector space with $\mathbb{F}^n$ for some n.

Proposition 3.4.8. *Suppose $S, T \in L(V, W)$ and $\lambda \in \mathbb{F}$, and suppose that $B = \{v_1, \dots, v_n\}$ is a basis for V and $C = \{w_1, \dots, w_m\}$ is a basis for W. Then $[S + \lambda T]_B^C = [S]_B^C + \lambda[T]_B^C$. Thus $[\cdot]_B^C \colon L(V, W) \to M_{m\times n}(\mathbb{F})$ is a linear transformation.*

Proof. First, we note that this follows immediately from 3.4.2 and 3.4.6 as

$$[S+\lambda T]_B^C x = [(S+\lambda T)[x]_B^{-1}]_C = [S[x]_B^{-1}+\lambda T[x]_B^{-1}]_C = [S[x]_B^{-1}]_C+\lambda[T[x]_B^{-1}]_C = [S]_B^C x+\lambda[T]_B^C x$$

for all $x \in \mathbb{F}^n$.

We give a direct proof as well. We have that

$$Sv_j = \sum_{i=1}^{m}([S]_B^C)_{i,j}w_i \text{ and } Tv_j = \sum_{i=1}^{m}([T]_B^C)_{i,j}w_i$$

for all $j \in [n]$. Thus

$$(S+\lambda T)v_j = Sv_j + \lambda Tv_j = \sum_{i=1}^{m}([S]_B^C)_{i,j}w_i + \lambda\sum_{i=1}^{m}([T]_B^C)_{i,j}w_i = \sum_{i=1}^{m}(([S]_B^C)_{i,j} + \lambda([T]_B^C)_{i,j})w_i.$$

for all $j \in [n]$, and $([S+\lambda T]_B^C)_{i,j} = ([S]_B^C)_{i,j} + \lambda([T]_B^C)_{i,j}$, so $[S+\lambda T]_B^C = [S]_B^C + \lambda[T]_B^C$. □

Proposition 3.4.9. *Suppose $T \in L(U,V)$ and $S \in L(V,W)$, and suppose that $A = \{u_1,\dots,u_m\}$ is a basis for U, $B = \{v_1,\dots,v_n\}$ is a basis for V, and $C = \{w_1,\dots,w_p\}$ is a basis for W. Then $[ST]_A^C = [S]_B^C[T]_A^B$.*

Proof. We have that

$$\begin{aligned} STu_j &= S\sum_{i=1}^{n}([T]_A^B)_{i,j}v_i = \sum_{i=1}^{n}([T]_A^B)_{i,j}(Sv_i) = \sum_{i=1}^{n}([T]_A^B)_{i,j}\left(\sum_{k=1}^{p}([S]_B^C)_{k,i}w_k\right) \\ &= \sum_{k=1}^{p}\left(\sum_{i=1}^{n}([S]_B^C)_{k,i}([T]_A^B)_{i,j}\right)w_k = \sum_{k=1}^{p}\left([S]_B^C[T]_A^B\right)_{k,j}w_k \end{aligned}$$

for all $j \in [m]$, so $[ST]_A^C = [S]_B^C[T]_A^B$. □

Exercises

V, W will denote vector spaces. Let $T \in L(V,W)$.

Exercise 3.4.10.

Exercise 3.4.11.

Exercise 3.4.12. Suppose $\dim(V) = n < \infty$ and $\dim(W) = m < \infty$, and let B, C be bases for V, W respectively.

(1) Show that $L(V,W) \cong M_{m\times n}(\mathbb{F})$.

(2) Show T is invertible if and only if $[T]_B^C$ is invertible.

Exercise 3.4.13. Suppose V is finite dimensional and B, C are two bases of V. Show $[T]_B \sim [T]_C$.

Chapter 4

Polynomials

In this section, we discuss the background material on polynomials needed for linear algebra. The two main results of this section are the Euclidean Algorithm and the Fundamental Theorem of Algebra. The first is a result on factoring polynomials, and the second says that every polynomial in $\mathbb{C}[z]$ has a root. The latter result relies on a result from complex analysis which is stated but not proved. For this section, $\mathbb{F}$ is a field.

4.1 The Algebra of Polynomials

Definition 4.1.1. A polynomial p over $\mathbb{F}$ is a sequence $p = (a_i)_{i\in\mathbb{Z}_{\geq 0}}$ where $a_i \in \mathbb{F}$ for all $i \in \mathbb{Z}_{\geq 0}$ such that there is an $n \in \mathbb{N}$ such that $a_i = 0$ for all $i > n$. The minimal $n \in \mathbb{Z}_{\geq 0}$ such that $a_i = 0$ for all $i > n$ (if it exists) is called the degree of p, denoted $\deg(p)$, and we define the degree of the zero polynomial, the sequence of all zeroes, denoted 0, to be $-\infty$. The leading coefficient of of p, denoted $\mathrm{LC}(p)$ is $a_{\deg(p)}$, and the leading coefficient of 0 is 0. Note that $p = 0$ if and only if $\mathrm{LC}(p) = 0$. A polynomial $p \in \mathbb{F}[z]$ is called monic if $\mathrm{LC}(p) = 1$.

Often, we identify a polynomial with the function it induces. The zero polynomial 0 induces the zero function $0\colon \mathbb{F} \to \mathbb{F}$ by $z \mapsto 0$. A nonzero polynomial $p = (a_i)$ induces a function still denoted $p\colon \mathbb{F} \to \mathbb{F}$ given by

$$p(z) = \sum_{i=0}^{\deg(p)} a_i z^i.$$

Sometimes when we identify the polynomial with the function, we call p a polynomial over $\mathbb{F}$ with indeterminate z, and a_i is called the coefficient of the z^i term for $i = 1, \dots, n$. If we say that p is the polynomial given by

$$p(z) = \sum_{i=0}^{n} a_i z^i,$$

then it is implied that $p = (a_i)$ where $a_i = 0$ for all $i > n$. Note that it is still possible in this case that $\deg(p) < n$.

The set of all polynomials over $\mathbb{F}$ is denoted $\mathbb{F}[z]$. The set of all polynomials of degree less than or equal to $n \in \mathbb{Z}_{\geq 0}$ is denoted $P_n(\mathbb{F})$. If $p, q \in \mathbb{F}[z]$ with $p = (a_i)$ and $q = (b_i)$, then we define the polynomial $p + q = (c_i) \in \mathbb{F}[z]$ by $c_i = a_i + b_i$ for all $i \in \mathbb{Z}_{\geq 0}$. Note that if we identify p, q with the functions they induce:

$$p(z) = \sum_{i=0}^{\deg(p)} a_i z^i \text{ and } q(z) = \sum_{i=0}^{\deg(q)} b_i z^i,$$

the polynomial $p + q \in \mathbb{F}[z]$ is given by

$$(p+q)(z) = p(z) + q(z) = \sum_{i=0}^{\deg(p)} a_i z^i + \sum_{i=0}^{\deg(q)} b_i z^i = \sum_{i=1}^{\max\{\deg(p),\deg(q)\}} (a_i + b_i) z^i.$$

We define the polynomial $pq = (c_i) \in \mathbb{F}[z]$ by

$$c_k = \sum_{i+j=k} a_i b_j \text{ for all } k \in \mathbb{Z}_{\geq 0}.$$

Alternatively, if we are using the function notation,

$$(pq)(z) = p(z)q(z) = \sum_{i=0}^{\deg(p)} a_i z^i \sum_{i=0}^{\deg(q)} b_i z^i = \sum_{k=0}^{\deg(p)+\deg(q)} \left(\sum_{i+j=k} a_i b_j \right) z^k.$$

If $\lambda \in \mathbb{F}$ and $p = (a_i) \in \mathbb{F}[z]$, then we define the polynomial $\lambda p = (c_i)$ by $c_i = \lambda a_i$ for all $i \in \mathbb{Z}_{\geq 0}$. Alternatively, we have

$$(\lambda p)(z) = \lambda p(z) = \lambda \sum_{i=0}^{\deg(p)} a_i z^i = \sum_{i=0}^{\deg(p)} (\lambda a_i) z^i.$$

Examples 4.1.2.

(1) $p(z) = z^2 + 1$ is a polynomial in $\mathbb{F}[z]$.

(2) A linear polynomial is a polynomial of degree 1, i.e. it is of the form $\lambda z - \mu$ for $\lambda, \mu \in \mathbb{F}$ with $\lambda \neq 0$.

Remarks 4.1.3.

(1) It is clear that $\mathrm{LC}(pq) = \mathrm{LC}(p)\,\mathrm{LC}(q)$, so pq is monic if and only if p, q are monic.

(2) A consequence of (1) is that $pq = 0$ implies $p = 0$ or $q = 0$ for $p, q \in \mathbb{F}[z]$.

(3) Another consequence of (1) is that $\deg(pq) = \deg(p) + \deg(q)$ for all $p, q \in \mathbb{F}[z]$.

(4) $\deg(p+q) \leq \max\{\deg(p), \deg(q)\}$ for all $p, q \in \mathbb{F}[z]$, and if $p, q \neq 0$, we have $\deg(p+q) \leq \deg(p) + \deg(q)$.

Definition 4.1.4. An $\mathbb{F}$-algebra, or an algebra over $\mathbb{F}$, is a vector space $(A, +, \cdot)$ and a binary operation $*$ on A called multiplication such that the following axioms hold:

(A1) *multiplicative associativity*: $(a * b) * c = a * (b * c)$ for all $a, b, c \in A$,

(A2) *distributivity*: $a * (b + c) = a * b + b * c$ and $(a + b) * c = a * c + b * c$ for all $a, b, c \in A$, and

(A3) *compatibility with scalar multiplication*: $\lambda \cdot (a * b) = (\lambda \cdot a) * b$ for all $\lambda \in \mathbb{F}$ and $a, b \in A$.

The algebra is called unital if there is a multiplicative identity $1 \in A$ such that $a*1 = 1*a = a$ for all $a \in A$, and the algebra is called commutative if $a * b = b * a$ for all $a, b \in A$.

Examples 4.1.5.

(1) The zero vector space (0) is a unital, commutative algebra over $\mathbb{F}$ with multiplication given as $0 * 0 = 0$. The unit in the algebra in this case is 0.

(2) $M_n(\mathbb{F})$ is an algebra over $\mathbb{F}$.

(3) $L(V)$ is an algebra over $\mathbb{F}$ if V is a vector space over $\mathbb{F}$ where multiplication is given by $S * T = S \circ T = ST$.

(4) $C(a, b)$ is an algebra over $\mathbb{R}$. This is also true for closed and half open intervals.

(5) $C^n(a, b)$ is an algebra over $\mathbb{R}$ for all $n \in \mathbb{N} \cup \{\infty\}$.

Proposition 4.1.6. $\mathbb{F}[z]$ *is an* $\mathbb{F}$*-algebra with addition, multiplication, and scalar multiplication defined as in 4.1.1.*

Proof. Exercise. □

Exercises

Exercise 4.1.7 (Ideals)**.** An ideal I in an $\mathbb{F}$-algebra A is a subspace $I \subset A$ such that $a*x \in I$ and $x * a \in I$ for all $x \in I$ and $a \in A$.

(1) Show that $I = \left\{p \in \mathbb{F}[z] \middle| p(0) = 0\right\}$ is an ideal in $\mathbb{F}[z]$.

(2) Show that $I_x = \left\{f \in C[a, b] \middle| f(x) = 0\right\}$ is an ideal in $C[a, b]$ for all $x \in [a, b]$.

(3) Show that $I = \left\{f \in C(\mathbb{R}, \mathbb{R}) \middle| \text{there are } a, b \in \mathbb{R} \text{ such that } f(x) = 0 \text{ for all } x < a, x > b\right\}$ is an ideal in $C(\mathbb{R}, \mathbb{R})$ (note that the $a, b \in \mathbb{R}$ depend on the f).

Exercise 4.1.8 (Ideals of Matrix Algebras)**.** Find all ideals of $M_n(\mathbb{F})$.

Exercise 4.1.9. Let $P_n(\mathbb{F})$ be the subset of $\mathbb{F}[z]$ consisting of all polynomials of degree less than or equal to n, i.e.

$$P_n(\mathbb{F}) = \left\{p \in \mathbb{F}[z] \middle| \deg(p) \leq n\right\}.$$

Show that $P_n(\mathbb{F})$ is a subspace of $\mathbb{F}[z]$ and $\dim(P_n(\mathbb{F}) = n + 1$.

Exercise 4.1.10. Let $B = (1, x, x^2, x^3)$, respectively $C = (1, x, x^2)$, be an ordered basis of $P_3(\mathbb{F})$, respectively $P_2(\mathbb{F})$; let $B' = (x^3, x^2, x, 1)$, respectively $C' = (x^2, x, 1)$ be another ordered basis of $P_3(\mathbb{F})$, respectively $P_2(\mathbb{F})$; and let $D \in L(P_3(\mathbb{F}), P_2(\mathbb{F}))$ be the derivative map

$$az^3 + bz^2 + cz + d \longmapsto 3az^2 + 2bz + c \text{ for all } a, b, c \in \mathbb{F}.$$

Find $[D]_{B}^{C}$. and $[D]_{B'}^{C'}$.

4.2 The Euclidean Algorithm

Theorem 4.2.1 (Euclidean Algorithm)**.** *Let $p, q \in \mathbb{F}[z] \setminus \{0\}$ with $\deg(q) \leq \deg(p)$. Then there are unique $k, r \in \mathbb{F}[z]$ such that $p = qk + r$ and $\deg(r) < \deg(q)$.*

Proof. Suppose

$$p(z) = \sum_{i=0}^{m} a_i z^i \text{ and } q(z) = \sum_{i=0}^{n} b_i z^i$$

where $m \geq n$, $a_m \neq 0$, and $b_n \neq 0$. Then set

$$k_1(z) = \frac{a_m}{b_n} z^{m-n} \text{ and } p_2(z) = p(z) - k_1(z)q(z),$$

and note that $\deg(p_2) < \deg(p)$. If $\deg(p_2) < \deg(q)$, we are finished by setting $k = k_1$ and $r = p_2$. Otherwise, set

$$k_2(z) = \frac{\operatorname{LC}(p_2)}{b_n} z^{\deg(p_2)-n} \text{ and } p_3(z) = p_2(z) - k_2(z)q(z),$$

ad note that $\deg(p_3) < \deg(p_2)$. If $\deg(p_3) < \deg(q)$, we are finished by setting $k = k_1 + k_2$ and $r = p_3$. Otherwise, set

$$k_3(z) = \frac{\operatorname{LC}(p_3)}{b_n} z^{\deg(p_3)-n} \text{ and } p_4(z) = p_3(z) - k_3(z)q(z),$$

ad note that $\deg(p_4) < \deg(p_3)$. If $\deg(p_4) < \deg(q)$, we are finished by setting $k = k_1 + k_2 + k_3$ and $r = p_4$. Otherwise, we may repeat this process as many times as necessary, and note that this algorithm will terminate as $\deg(p_{j+1}) < \deg(p_j)$ (eventually $\deg(p_j) < \deg(q)$ for some q).

It remains to show k, r are unique. Suppose $p = k_1 q + r_1 = k_2 q + r_2$ with $\deg(r_1), \deg(r_2) < \deg(q)$. Then $0 = (k_1 - k_2)q + (r_1 - r_2)$. Now since $\deg(r_1 - r_2) < \deg(q)$, we have that

$$0 = \operatorname{LC}((k_1 - k_2)q + (r_1 - r_2)) = \operatorname{LC}((k_1 - k_2)q) = \operatorname{LC}(k_1 - k_2)\operatorname{LC}(q),$$

so $k_1 = k_2$ as $q \neq 0$. It immediately follows that $r_1 = r_2$. □

Exercises

Exercise 4.2.2 (Principal Ideals)**.** Suppose A is a unital, commutative algebra over $\mathbb{F}$. An ideal $I \subset A$ is called principal if there is an $x \in I$ such that $I = \big\{ax\big|a \in A\big\}$. Show that every ideal of $\mathbb{F}[z]$ is principal (see 4.1.7).

4.3 Prime Factorization

Definition 4.3.1. For $p, q \in \mathbb{F}[z]$, we say p divides q, denoted $p|q$, if there is a polynomial $k \in \mathbb{F}[z]$ such that $kp = q$.

Examples 4.3.2.

(1) $(z \pm 1) \mid (z^2 - 1)$.

(2) $(z \pm i) \mid (z^2 + 1)$ in $\mathbb{C}[z]$, but there are no nonconstant polynomials in $\mathbb{R}[z]$ that divide $z^2 + 1$.

Remark 4.3.3. Note that the above defines a relation on $\mathbb{F}[z]$ that is reflexive and transitive (see 1.1.6). It is also antisymmetric on the set of monic polynomials.

Definition 4.3.4. A polynomial $p \in \mathbb{F}[z]$ with $\deg(p) \geq 1$ is called irreducible (or prime) if $q \in \mathbb{F}[z]$ with $q \mid p$ and $\deg(q) < \deg(p)$ implies q is constant.

Examples 4.3.5.

(1) Every linear (degree 1) polynomial is irreducible.

(2) $z^2 + 1$ is irreducible over $\mathbb{R}$ (in $\mathbb{R}[z]$), but not over $\mathbb{C}$ (in $\mathbb{C}[z]$).

Definition 4.3.6. Polynomials $p_1, \ldots, p_n \in \mathbb{F}[z]$ where $n \geq 2$ and $\deg(p_i) \geq 1$ for all $i \in [n]$ are called relatively prime if $q \mid p_i$ for all $i \in [n]$ implies q is constant.

Examples 4.3.7.

(1) Any $n \geq 2$ distinct monic linear polynomials in $\mathbb{F}[z]$ is relatively prime.

(2) Any $n \geq 2$ distinct quadratics $z^2 + az + b \in \mathbb{R}[z]$ such that $a^2 - 4b < 0$ are relatively prime.

(3) Any quadratic $z^2 + az + b \in \mathbb{R}[z]$ with $a^2 - 4b < 0$ and any linear polynomial in $\mathbb{R}[z]$ are relatively prime.

Proposition 4.3.8. *$p_1, \ldots, p_n \in \mathbb{F}[z]$ where $n \geq 2$ are relatively prime if and only if there are $q_1, \ldots, q_n \in \mathbb{F}[z]$ such that*

$$\sum_{i=1}^{n} q_i p_i = 1.$$

Proof. Suppose there are $q_1, \dots, q_n \in \mathbb{F}[z]$ such that

$$\sum_{i=1}^{n} q_i p_i = 1,$$

and suppose $q|p_i$ for all $i \in [n]$. Then $q|1$, so q must be constant, and the p_i's are relatively prime.

Suppose now that the p_i's are relatively prime. Choose polynomials $q_1, \dots, q_n \in \mathbb{F}[z]$ such that

$$f = \sum_{i=1}^{n} q_i p_i$$

has minimal degree which is not $-\infty$ (in particular, $f \neq 0$). Now we know $\deg(f) \leq \deg(p_j)$ as we could have $q_j = 1$ and $q_i = 0$ for all $i \neq j$. By the Euclidean algorithm, for $j \in [n]$, there are unique k_j, r_j with $\deg(r_j) < \deg(f)$ such that $p_j = k_j f + r_j$. Since

$$r_j = p_j - kf = p_j - k\sum_{i=1}^{n} q_i p_i = (1 - kq_j)p_j + \sum_{i \neq j} kq_i p_i,$$

we must have that $r_j = 0$ for all $j \in [n]$ since $\deg(f)$ was chosen to be minimal. Thus $f|p_i$ for all $i \in [n]$, so f must be constant. As $f \neq 0$, we may divide by f, and replacing q_i with q_i/f for all $i \in [n]$, we get the desired result. □

Corollary 4.3.9. *Suppose $p, q_1, \dots, q_n \in \mathbb{F}[z]$ where $n \geq 2$ and p is irreducible. Then if $p \mid (q_1 \cdots q_n)$, $p \mid q_i$ for some $i \in [n]$.*

Proof. We proceed by induction on $n \geq 2$.

$\underline{n = 2:}$ Suppose $q_1, q_2 \in \mathbb{F}[z]$ and $p \mid q_1 q_2$. If $p \mid q_1$, we are finished. Otherwise, p, q_1 are relatively prime (if k is nonconstant and $k \mid p$, then $k = p$, but $k \nmid q_1$). By 4.3.8, there are $g_1, g_2 \in \mathbb{F}[z]$ such that $g_1 p + g_2 q_1 = 1$. We then get

$$q_2 = g_1 q_2 p + g_2 q_1 q_2.$$

As $p|g_1 q_2 p$ and $p|g_2 q_1 q_2$, we have $p|q_2$.

$\underline{n - 1 \Rightarrow n:}$ We have $p|(q_1 \cdots q_{n-1})q_n$, so by the case $n = 2$, either $p|q_n$, in which case we are finished, or $p|q_1 \cdots q_{n-1}$, in which case we apply the induction hypothesis to get $p|q_i$ for some $i \in [n-1]$. □

Lemma 4.3.10. *Suppose $p, q, r \in \mathbb{F}[z]$ with $p \neq 0$ such that $pq = pr$. Then $q = r$.*

Proof. We have $p(q - r) = 0$. As $\mathrm{LC}(p(q - r)) = \mathrm{LC}(p)\,\mathrm{LC}(q - r)$, we must have that $\mathrm{LC}(q - r) = 0$, and $q = r$. □

Theorem 4.3.11 (Unique Factorization)**.** *Let $p \in \mathbb{F}[z]$ with $\deg(p) \geq 1$. Then there are unique monic irreducible polynomials $p_1, \ldots, p_n$ and a unique constant $\lambda \in \mathbb{F} \setminus \{0\}$ such that*

$$p = \lambda p_1 \cdots p_n = \lambda \prod_{i=1}^{n} p_i.$$

Proof.

Existence: We proceed by induction on $\deg(p)$.

$\deg(p) = 1$: This case is trivial as $p/\operatorname{LC}(p)$ is a monic irreducible polynomial and $p = \operatorname{LC}(p)(p/\operatorname{LC}(p))$.

$\deg(p) > 1$: We assume the result holds for all polynomials of degree less than $\deg(p)$. If p is irreducible, then so is $p/\operatorname{LC}(p)$, and we have $p = \operatorname{LC}(p)(p/\operatorname{LC}(p))$. If p is not irreducible, there is are nonconstant polynomials $q, r \in \mathbb{F}[z]$ with $\deg(q), \deg(r) < \deg(p)$ such that $p = qr$. As the result holds for q, r, we have that the result holds for p.

Uniqueness: Suppose

$$p = \lambda \prod_{i=1}^{n} p_i = \mu \prod_{j=1}^{m} q_i$$

where all p_i, q_j's are monic, irreducible polynomials. Then $\lambda = \mu = \operatorname{LC}(p)$. Now p_n divides $q_1 \cdots q_m$, so $p_n | q_i$ for some $i \in [m]$. After relabeling, we may assume $i = m$. But then $p_n = q_m$, so

$$p_1 \cdots p_{n-1} p_n = q_1 \cdots q_{m-1} p_n.$$

By 4.3.10, we have that $p_1 \cdots p_{n-1} = q_1 \cdots q_{m-1}$, and we may repeat this process. Eventually, we see that $n = m$ and that the q_j's are at most a rearrangement of the p_i's. □

Exercises

Exercise 4.3.12. Suppose $p, q \in \mathbb{F}[z]$ are relatively prime. Show that p^n and q^m are relatively prime for $n, m \in \mathbb{N}$.

Exercise 4.3.13. Let $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ be distinct and let $\mu_1, \ldots, \mu_n \in \mathbb{F}$. Show that there is a unique polynomial $p \in \mathbb{F}[z]$ of degree $n-1$ such that $p(\lambda_i) = \mu_i$ for all $i \in [n]$. Deduce that a polynomial of degree d is uniquely determined by where it sends $d+1$ (distinct) points of $\mathbb{F}$.

Exercise 4.3.14 (Greatest Common Divisors and Least Common Multiples)**.** Let $p_1, p_2 \in \mathbb{F}[z]$ with degree ≥ 1.

(1) Show that there is a unique monic polynomial $q \in \mathbb{F}[z]$ of minimal degree such that $p_1 \mid q$ and $p_2 \mid q$. This polynomial, usually denoted $\operatorname{lcm}(p_1, p_2)$, is called the least common multiple of p_1 and p_2.

(2) Show that there is a unique polynomial $q \in \mathbb{F}[z]$ of maximal degree with $\operatorname{LT}(q) = 1$ such that $q \mid p_1$ and $q \mid p_2$. This polynomial, usually denoted $\gcd(p_1, p_2)$, is called the greatest common divisor of p_1 and p_2.

4.4 Irreducible Polynomials

For this section, $\mathbb{F}$ will denote a field and $\mathbb{K} \subset \mathbb{F}$ will be a subfield such that $\mathbb{F}$ is a finite dimensional $\mathbb{K}$-vector space.

Definition 4.4.1. $\lambda \in \mathbb{F}$ is called a root of $p \in \mathbb{K}[z]$ where $\deg(p) \geq 1$ if $p(\lambda) = 0$.

Definition-Proposition 4.4.2. *Let $\lambda \in \mathbb{F}$. The irreducible polynomial of λ over the field $\mathbb{K}$ is the unique monic irreducible polynomial $\operatorname{Irr}_{\mathbb{K},\lambda} \in \mathbb{K}[z]$ of minimal degree such that $\operatorname{Irr}_{\mathbb{K},\lambda}(\lambda) = 0$.*

Proof.

Existence: Since $\lambda \in \mathbb{F}$, $\lambda^k \in \mathbb{F}$ for all $k \in \mathbb{Z}_{\geq 0}$. The set $\left\{\lambda^k \middle| k \in \mathbb{Z}_{\geq 0}\right\}$ cannot be linearly independent over $\mathbb{K}$ as $\mathbb{F}$ is a finite dimensional $\mathbb{K}$-vector space, so there are scalars $\mu_0, \dots, \mu_n$ with $\mu_n \neq 0$ such that

$$\sum_{i=0}^{n} \mu_i \lambda^i = 0 \Longrightarrow \sum_{i=0}^{n} \frac{\mu_i}{\mu_n} \lambda^i = 0.$$

Define $p \in \mathbb{K}[z]$ by

$$p(z) = \sum_{i=0}^{n} \frac{\mu_i}{\mu_n} z^i.$$

Then $p(\lambda) = 0$, so there is a monic polynomial in $\mathbb{K}[z]$ with λ as a root.

Pick a monic polynomial $q \in \mathbb{K}[z]$ of minimal degree such that $q(\lambda) = 0$. Then q is irreducible (otherwise, there are $k, r \in \mathbb{K}[z]$ of degree ≥ 1 such that $q = kr$, so $q(\lambda) = k(\lambda)r(\lambda) = 0$, so either $k(\lambda) = 0$ or $r(\lambda) = 0$, a contradiction as q was chosen of minimal degree).

Uniqueness: Suppose $p, q \in \mathbb{K}[z]$ are both monic polynomials of minimal degree such that $p(\lambda) = q(\lambda) = 0$. Then $\deg(p) = \deg(q)$, so by the Euclidean algorithm, there are $k, r \in \mathbb{K}[z]$ with $\deg(r) < \deg(q)$ such that $p = kq + r$. Now $p(\lambda) = k(\lambda)q(\lambda) + r(\lambda)$, so $r(\lambda) = 0$. This is only possible if $r = 0$ as $\deg(r) < \deg(q)$. Hence $p = kq$ with $\deg(p) = \deg(q)$, so k is a constant. As p, q are both monic, $k = 1$, and $p = q$. □

Corollary 4.4.3. *Suppose $\lambda \in \mathbb{F}$. Then $\lambda \in \mathbb{K}$ if and only if $\deg(\operatorname{Irr}_{\mathbb{K},\lambda}) = 1$.*

Corollary 4.4.4. *Suppose $\lambda \in \mathbb{F}$ and $\dim_{\mathbb{K}}(\mathbb{F}) = n$. Then $\deg(\operatorname{Irr}_{\mathbb{K},\lambda}) \leq n$.*

Lemma 4.4.5. *Suppose $\lambda \in \mathbb{F}$ is a root of $p \in \mathbb{K}[x]$ where $\deg(p) \geq 1$. Then $\operatorname{Irr}_{\mathbb{K},\lambda} | p$.*

Proof. The proof is similar to 4.4.2. Without loss of generality, we may assume p is monic (if not, we replace p with $p/\operatorname{LC}(p)$. Now since $\deg(\operatorname{Irr}_{\mathbb{K},\lambda}) \leq \deg(p)$, by the Euclidean algorithm, there are $k, r \in \mathbb{K}[z]$ with $\deg(r) < \deg(\operatorname{Irr}_{\mathbb{K},\lambda})$ such that $p = k \operatorname{Irr}_{\mathbb{K},\lambda} + r$. We immediately see $r(\lambda) = 0$, so $r = 0$. □

Example 4.4.6. We know that $\mathbb{C}$ is a two dimensional real vector space. If $\lambda \in \mathbb{C}$ is a root of $p \in \mathbb{R}[z]$, then $\overline{\lambda}$ is also a root of p. In fact, since $p(\lambda) = 0$, we have that $\overline{p(\lambda)} = p(\overline{\lambda}) = 0$ since the coefficients of p are all real numbers. Hence, the irreducible polynomial in $\mathbb{R}[z]$ for $\lambda \in \mathbb{C}$ is given by

$$\operatorname{Irr}_{\mathbb{R},\lambda}(z) = \begin{cases} z - \lambda & \text{if } \lambda \in \mathbb{R} \\ (z-\lambda)(z-\overline{\lambda}) = z^2 - 2\operatorname{Re}(\lambda)z + |\lambda|^2 & \text{if } \lambda \notin \mathbb{R}. \end{cases}$$

Definition 4.4.7. The polynomial $p \in \mathbb{F}[z]$ splits (into linear factors) if there are $\lambda, \lambda_1, \ldots, \lambda_n \in \mathbb{F}$ such that

$$p(z) = \lambda \prod_{i=1}^{n} (z - \lambda_i).$$

Examples 4.4.8.

(1) Every constant polynomial trivially splits in $\mathbb{F}[z]$.

(2) The polynomial $z^2 + 1$ splits in $\mathbb{C}[z]$ but not in $\mathbb{R}[z]$.

Definition 4.4.9. The field $\mathbb{F}$ is called algebraically closed if every $p \in \mathbb{F}[z]$ splits.

Exercises

4.5 The Fundamental Theorem of Algebra

Definition 4.5.1. A function $f\colon \mathbb{C} \to \mathbb{C}$ is entire if there is a power series representation

$$f(z) = \sum_{n=0}^{\infty} a_n z^n$$

where $a_n \in \mathbb{C}$ for all $n \in \mathbb{Z}_{\geq 0}$ that converges for all $z \in \mathbb{C}$.

Examples 4.5.2.

(1) Every polynomial in $\mathbb{C}[z]$ is entire as its power series representation is itself.

(2) $f(z) = e^z$ is entire as

$$e^z = \sum_{j=0}^{\infty} \frac{z^n}{n!}$$

(3) $f(z) = \cos(z)$ is entire as

$$\cos(z) = \sum_{j=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}$$

(4) $f(z) = \sin(z)$ is entire as

$$\sin(z) = \sum_{j=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}$$

In order to prove the Fundamental Theorem of Algebra, we will need a theorem from complex analysis. We state it without proof.

Theorem 4.5.3 (Liouville)**.** *Every bounded, entire function $f\colon \mathbb{C} \to \mathbb{C}$ is constant.*

Theorem 4.5.4 (Fundamental Theorem of Algebra)**.** *Every nonconstant polynomial in $\mathbb{C}[x]$ has a root.*

Proof. Let p be a polynomial in $\mathbb{C}[z]$. If p has no roots, then $1/p$ is entire and bounded. By Liouville's Theorem, $1/p$ is constant, so p is constant. □

Remark 4.5.5. Let $p \in \mathbb{C}[z]$ be a nonconstant polynomial, and let $n = \deg(p)$. Then by 4.5.4, p has a root, say μ_1, so there is a $p_1 \in \mathbb{C}[z]$ such that $p(z) = (z - \mu_1)p_1(z)$. If p_1 is nonconstant, then apply 4.5.4 again to see p_1 has a root μ_2, so there is a $p_2 \in \mathbb{C}[z]$ with $\deg(p_2) = n - 2$ such that $p(z) = (z - \mu_1)(z - \mu_2)p_2(z)$. Repeating this process, we see that p_n must be constant as a polynomial has at most n roots. Hence there are $\mu_1, \dots, \mu_n \in \mathbb{C}$ such that

$$p(z) = k \prod_{j=1}^{n} (z - \mu_j) = k(z - \mu_1) \cdots (z - \mu_n) \text{ for some } k \in \mathbb{C}.$$

Hence every polynomial in $\mathbb{C}[z]$ splits.

Proposition 4.5.6. *Every nonconstant polynomial $p \in \mathbb{R}[z]$ splits into linear and quadratic factors.*

Proof. We know p has a root $\lambda \in \mathbb{C}$ by 4.5.4, so by 4.4.6, $\operatorname{Irr}_\lambda | p$, i.e. there is a $p_2 \in \mathbb{R}[z]$ with $\deg(p_2) < \deg(p)$ such that $p = \operatorname{Irr}_\lambda p_2$. If p_2 is constant we are finished. If not, we may repeat the process for p_2 to get p_3, and so forth. This process will terminate as $\deg(p_n) < \deg(p_{n+1})$. □

Exercises

4.6 The Polynomial Functional Calculus

Definition 4.6.1 (Polynomial Functional Calculus)**.** Given a polynomial

$$p(z) = \sum_{j=0}^{n} \lambda_j z^j \in \mathbb{F}[z],$$

and $T \in L(V)$, we can define an operator by

$$p(T) = \sum_{j=0}^{n} \lambda_j T^j$$

with the convention that $T^0 = I$.

Proposition 4.6.2. *The polynomial functional calculus satisfies:*

(1) $(p+q)(T) = p(T) + q(T)$ *for all* $p, q \in \mathbb{F}[z]$ *and* $T \in L(V)$*,*

(2) $(pq)(T) = p(T)q(T) = q(T)p(T)$ *for all* $p, q \in \mathbb{F}[z]$ *and* $T \in L(V)$*,*

(3) $(\lambda p)(T) = \lambda p(T)$ *for all* $\lambda \in \mathbb{F}$*,* $p \in \mathbb{F}[z]$ *and* $T \in L(V)$*, and*

(4) $(p \circ q)(T) = p(q(T))$ *for all* $p, q \in \mathbb{F}[z]$ *and* $T \in L(V)$*.*

Proof. Exercise. □

Exercises

Exercise 4.6.3. Prove 4.6.2.

Chapter 5

Eigenvalues, Eigenvectors, and the Spectrum

For this chapter, V will denote a vector space over $\mathbb{F}$. We begin our study of operators in $L(V)$ by studying the spectrum of an element $T \in L(V)$ which gives a lot of information about the operator. We then discuss methods for calculating the spectrum $\operatorname{sp}(T)$ when V is finite dimensional, namely the characteristic and minimal polynomials. For $A \in M_n(\mathbb{F})$, we will identify $L_A \in L(\mathbb{F}^n)$ with the matrix A.

5.1 Eigenvalues and Eigenvectors

Definition 5.1.1. Let $T \in L(V)$.

(1) The spectrum of T, denoted $\operatorname{sp}(T)$, is $\big\{\lambda \in \mathbb{F} \big| T - \lambda I \text{ is not invertible}\big\}$.

(2) If $v \in V \setminus \{0\}$ such that $Tv = \lambda v$ for some $\lambda \in \mathbb{F}$, then v is called an eigenvector of T with corresponding eigenvalue λ.

(3) If λ is an eigenvalue of T, i.e. there is an eigenvector v of T with corresponding eigenvalue λ, then $E_\lambda = \big\{w \in V \big| Tw = \lambda w\big\}$ is the eigenspace associated to the eigenvalue λ. It is clear that E_λ is a subspace of V.

Proposition 5.1.2. *Suppose V is finite dimensional. Then $\lambda \in \mathbb{F}$ is an eigenvalue of $T \in L(V)$ if and only if $\lambda \in \operatorname{sp}(V)$.*

Proof. By 3.2.15, $T - \lambda I$ is not invertible if and only if $T - \lambda I$ is not injective. It is clear that $T - \lambda I$ is not injective if and only if there is an eigenvector v of T with corresponding eigenvalue λ. □

Remark 5.1.3. Note that 5.1.2 depends on the finite dimensionality of V as 3.2.15 depends on the finite dimensionality of V.

Examples 5.1.4.

(1) Suppose $A \in M_n(\mathbb{F})$ is an upper triangular matrix. Then $\operatorname{sp}(L_A)$ is the set of distinct values on the diagonal of A. This can be seen as $A - \lambda I$ is not invertible if and only if λ appears on the diagonal of A.

(2) The eigenvalues of the operator

$$A = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \in M_n(\mathbb{F})$$

are $0, 1$ corresponding to respective eigenvectors

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

If we had an eigenvector associated to the eigenvalue λ, then

$$\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{2}\begin{pmatrix} a+b \\ a+b \end{pmatrix} = \lambda \begin{pmatrix} a \\ b \end{pmatrix},$$

so $a + b = 2\lambda a = 2\lambda b$. Thus $\lambda(a - b) = 0$, so either $\lambda = 0$, or $a = b$. In the latter case, we have $\lambda = 1$ as $a = b \neq 0$. Hence $\operatorname{sp}(L_A) = \{0, 1\}$.

(3) The operator

$$B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{R})$$

has no eigenvalues. In fact, if

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -b \\ a \end{pmatrix} = \lambda \begin{pmatrix} a \\ b \end{pmatrix},$$

then we must have that $\lambda a = -b$ and $\lambda b = a$, so $-b = \lambda a = \lambda^2 b$, and $(\lambda^2 + 1)b = 0$. Now the first is nonzero as $\lambda \in \mathbb{R}$, so $b = 0$. Thus $a = 0$, and B has no eigenvectors.

If instead we consider $B \in M_2(\mathbb{C})$, then B has eigenvalues $\pm i$ corresponding to eigenvectors

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \pm i \end{pmatrix}.$$

(4) Consider $L_z \in L(\mathbb{F}[z])$ given by $L_z(p(z)) = zp(z)$. Then the set of eigenvalues of L_z is empty as $zp(z) = \lambda p(z)$ if and only if $(z - \lambda)p(z) = 0$ if and only if $p = 0$ for all $\lambda \in \mathbb{F}$.

(5) Recall that a polynomial is really a sequence of elements of $\mathbb{F}$ which is eventually zero. Now the operator L_z discussed above is really the right shift operator R on $\mathbb{F}[z]$:

$$R(a_0, a_1, a_2, a_3, \dots) = L_z(a_0, a_1, a_2, a_3, \dots) = (0, a_0, a_1, a_2, \dots).$$

One can also define the left shift operator $L \in L(\mathbb{F}[z])$ by

$$L(a_0, a_1, a_2, a_3, \dots) = (a_1, a_2, a_3, a_4, \dots).$$

Note that $LR = I$, the identity operator in $L(\mathbb{F}[z])$, but R is not surjective as the constant polynomials are not hit, and L is not injective as the constant polynomials are killed. Thus L, R are not invertible, so $0 \in \operatorname{sp}(R)$ and $0 \in \operatorname{sp}(L)$. Thus R is an example of an operator with no eigenvalues, but $0 \in \operatorname{sp}(R)$. In fact, the only eigenvalue of L is 0 as $Lp = \lambda p$ implies

$$\deg(\lambda p) = \deg(Lp) = \begin{cases} \deg(p) - 1 & \text{if } \deg(p) \geq 2 \\ -\infty & \text{if } \deg(p) < 2 \end{cases}$$

which is only possible if $\lambda p = 0$, and $Lp = 0$ only for the constant polynomials.

Definition 5.1.5. Suppose $T \in L(V)$. A subspace $W \subset V$ is called T-invariant if $TW \subset W$.

Examples 5.1.6. Suppose $T \in L(V)$.

(1) (0) and V are the trivial T-invariant subspaces.

(2) $\ker(T)$ and $\operatorname{im}(T)$ are T-invariant subspaces.

(3) An eigenspace of T is a T-invariant subspace.

Notation 5.1.7. Given $T \in L(V)$ and a T-invariant subspace $W \subset V$, we define the restriction of T to W in $L(W)$, denoted $T|_W$, by $T|_W(w) = Tw$ for all $w \in W$. Note that $T|_W$ is just the restriction of the function $T\colon V \to V$ to W with the codomain restricted as well.

Lemma 5.1.8 (Polynomial Eigenvalue Mapping)**.** *Suppose $p \in \mathbb{F}[z]$, $T \in L(V)$, and v is an eigenvector for T corresponding to $\lambda \in \operatorname{sp}(T)$. Then $p(T)v = p(\lambda)v$, so $p(\lambda) \in \operatorname{sp}(p(T))$.*

Proof. Exercise. □

Proposition 5.1.9. *Let $T \in L(V)$, and suppose $\lambda_1, \ldots, \lambda_n$ are eigenvalues of T. Suppose $v_i \in E_{\lambda_i}$ for all $i \in [n]$ such that $\displaystyle\sum_{i=1}^{n} v_i = 0$. Then $v_i = 0$ for all $i \in [n]$. Thus*

$$\bigoplus_{i=1}^{n} E_{\lambda_i}$$

is a well-defined subspace of V.

Proof. For $i \in [n]$ define $f_i \in \mathbb{F}[z]$ by

$$f_i(z) = \frac{\prod\limits_{j \neq i} (z - \lambda_j)}{\prod\limits_{j \neq i} (\lambda_i - \lambda_j)}.$$

Then

$$f_i(\lambda_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else.} \end{cases}$$

By 5.1.8, for $i \in [n]$,

$$0 = f_i(T)0 = f_i(T)\sum_{j=1}^{n} v_j = \sum_{j=1}^{n} f_i(T)v_j = \sum_{j=1}^{n} f_i(\lambda_j)v_j = v_i.$$

□

Remark 5.1.10. Eigenvectors corresponding to distinct eigenvalues are linearly independent, so T can have at most $\dim(V)$ many distinct eigenvalues.

Exercises

Exercise 5.1.11. We call two operators in $S, T \in L(V)$ similar, denoted $S \sim T$, if there is an invertible $J \in L(V)$ such that $J^{-1}SJ = T$. The similarity class of S is $\{T \in L(V) \big| T \sim S\}$. Show

(1) $\sim$ defines a relation on $L(V)$ which is reflexive, symmetric, and transitive (see 1.1.6),

(2) distinct similarity classes are disjoint, and

(3) if $S \sim T$, then $\mathrm{sp}(S) = \mathrm{sp}(T)$. Thus the spectrum is an invariant of a similarity class.

Exercise 5.1.12 (Finite Shift Operators)**.** Let $B = \{v_1, \dots, v_n\}$ be a basis of V. Find the spectrum of the shift operator $T \in L(V)$ given by $Tv_i = v_{i+1}$ for $i = 1, \dots, n-1$ and $Tv_n = v_1$.

Exercise 5.1.13 (Infinite Shift Operators)**.** Let

$$\begin{aligned}
\ell^1(\mathbb{N}, \mathbb{R}) &= \left\{ (a_n) \middle| a_{\in}\mathbb{R} \text{ for all } n \in \mathbb{N} \text{ and } \sum_{n=1}^{n} |a_n| < \infty \right\}, \\
\ell^\infty(\mathbb{N}, \mathbb{R}) &= \left\{ (a_n) \big| |a_n| \in [-M, M] \text{ for all } n \in \mathbb{N} \text{ for some } M \in \mathbb{N} \right\}, \text{ and} \\
\mathbb{R}^\infty &= \left\{ (a_n) \big| a_n \in \mathbb{R} \text{ for all } n \in \mathbb{N} \right\}.
\end{aligned}$$

In other words, $\ell^1(\mathbb{N}, \mathbb{R})$ is the set of absolutely convergent sequences of real numbers, $\ell^\infty(\mathbb{N}, \mathbb{R})$ is the set of bounded sequences of real numbers, and $\mathbb{R}^\infty$ is the set of all sequences of real numbers.

(1) Show that $\ell^1(\mathbb{N}, \mathbb{R})$, $\ell^\infty(\mathbb{N}, \mathbb{R})$, and $\mathbb{R}^\infty$ are vector spaces over $\mathbb{R}$.

Hint: First show that $\mathbb{R}^\infty$ is a vector space over $\mathbb{R}$, and then show $\ell^1(\mathbb{N}, \mathbb{R})$ and $\ell^\infty(\mathbb{N}, \mathbb{R})$ are subspaces. To show $\ell^1(\mathbb{N}, \mathbb{R})$ is closed under addition, use the fact that if (a_n) is absolutely convergent, then we can add up the terms $|a_n|$ in any order that we want and we will still get the same number.

(2) Define $S_1 \in L(\ell^1(\mathbb{N}, \mathbb{R}))$, $S_\infty \in L(\ell^\infty(\mathbb{N}, \mathbb{R}))$ and $S_0 \in L(\mathbb{R}^\infty)$ by

$$S_i(a_1, a_2, a_3, \dots) = (a_2, a_3, a_4, \dots) \text{ for } i = 0, 1, \infty.$$

(a) Find the set of eigenvalues, denoted E_{S_i} for $i = 0, 1, \infty$.

(b) Why is part (a) asking you to find the set of eigenvalues of S_i and not the spectrum of S_i for $i = 0, 1, \infty$?

Exercise 5.1.14. Let $\mathbb{F}$ be a real vector space. $\lambda \in \mathbb{C} \setminus \mathbb{R}$ is called a psuedoeigenvalue of $T \in L(V)$ if λ is an eigenvalue of $T_{\mathbb{C}} \in L(V_{\mathbb{C}})$. Suppose $\lambda = a + ib$ is a psuedoeigenvalue of $T \in L(V)$, and suppose $u + iv \in V_{\mathbb{C}} \setminus \{0\}$ such that $T_{\mathbb{C}}(u + iv) = \lambda(u + iv)$. Set $B = \{u, v\}$. Show

(1) $W = \operatorname{span}(B) \subset V$ is invariant for T,

(2) B is a basis for W, and

(3) $\min_{T|_W} = \operatorname{Irr}_{\mathbb{R},\lambda}$.

(4) Find $[T|_W]_B$.

5.2 Determinants

In this section, we discuss how to take the determinant of a square matrix. The determinant is a crude tool for calculating the spectrum of L_A for $A \in M_n(\mathbb{F})$ via the characteristic polynomial which is defined in the next section. This method is only recommended for small matrices, and it is generally useless for large matrices without the use of a computer. Even then, one can only get numerical approximations to the eigenvalues.

Definition 5.2.1. A permutation on $[n]$ is a bijection $\sigma\colon [n] \to [n]$. The set of all permutations on $[n]$ is denoted S_n. Permutations are denoted

$$\begin{pmatrix} 1 & \cdots & n \\ \sigma(1) & \cdots & \sigma(n) \end{pmatrix}.$$

A permutation $\sigma \in S_n$ is called a transposition if $\sigma(i) = i$ for all but two distinct elements $j, k \in [n]$ and $\sigma(j) = k$ and $\sigma(k) = j$. Note that the composite of two permutations in S_n is a permutation in S_n.

Examples 5.2.2.

(1)

(2)

Definition 5.2.3. An inversion in $\sigma \in S_n$ is a pair $(i, j) \in [n] \times [n]$ such that $i < j$ and $\sigma(i) > \sigma(j)$. We call $\sigma \in S_n$ odd if there are an odd number of inversions in σ, and we call σ even if there are an even number of inversions in σ. Define the sign function $\operatorname{sgn}\colon S_n \to \{\pm 1\}$ by $\operatorname{sgn}(\sigma) = 1$ if σ is even and -1 if σ is odd.

Examples 5.2.4.

(1) The identity permutation has sign 1 since it has no inversions, and a transposition has sign -1 as it has only one inversion.

(2)

Definition 5.2.5. The determinant of $A \in M_n(\mathbb{F})$ is

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i,\sigma(i)}.$$

Example 5.2.6. We calculate the determinant of the 2 × 2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{F}).$$

There are two permutations in S_2: the identity permutation and the transposition switching 1 and 2. Call these $\operatorname{id}, \sigma$ respectively. It is clear that id has no inversions and σ has one inversion, so $\operatorname{sgn}(\operatorname{id}) = 1$ and $\operatorname{sgn}(\sigma) = -1$. Hence

$$\det(A) = \left(\operatorname{sgn}(\operatorname{id}) \prod_{i=1}^{n} A_{i,\operatorname{id}(i)}\right) + \left(\operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{i,\sigma(i)}\right) = A_{1,1}A_{2,2} + (-1)A_{1,2}A_{2,1} = ad - bc \in \mathbb{F}.$$

Proposition 5.2.7. *For $A \in M_n(\mathbb{F})$, we can calculate $\det(A)$ recursively. Let $A^{i,j} \in M_{n-1}(\mathbb{F})$ be the matrix obtained from A by deleting the i^{th} row and j^{th} column. Then fixing $i \in [n]$, we have*

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

This procedure is commonly referred to as taking the determinant of A by expanding along the i^{th} row of A. We may also take the determinant by expanding along the j^{th} column of A. Fixing $j \in [n]$, we get

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

Proof.

FINISH: We proceed by induction on n.

$\underline{n = 1:}$ Obvious.

$\underline{n - 1 \Rightarrow n:}$ Expanding along the i^{th} row and using the induction hypothesis, we see

$$\begin{aligned} \sum_{j=1}^{n} (-1)^{i+j} A_{i,j} \det(A^{i,j}) &= \sum_{j=1}^{n} (-1)^{i+j} A_{i,j} \sum_{\sigma \in S_{n-1}} \operatorname{sgn}(\sigma) \prod_{k=1}^{n-1} (A^{i,j})_{k,\sigma(k)} \\ &= \sum_{j=1}^{n} \sum_{\sigma \in S_{n-1}} (-1)^{i+j} A_{i,j} \left(\operatorname{sgn}(\sigma) \prod_{k=1}^{n-1} (A^{i,j})_{k,\sigma(k)} \right) \end{aligned}$$

where $[n-1]$ is identified with

□

Corollary 5.2.8. *If $A \in M_n(\mathbb{F})$ has a row or column of zeroes, then* $\det(A) = 0$.

Lemma 5.2.9. *Let $A \in M_n(\mathbb{F})$.*

(1) $\det(A) = \det(A^T)$.

(2) If A is block upper triangular, i.e., there are square matrices $A_1, \ldots, A_m$ such that

$$A = \begin{pmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

then

$$\det(A) = \prod_{i=1}^{m} \det(A_i).$$

Proof.

(1) This is obvious by 5.2.7 as taking the determinant of A along the i^{th} row of A is the same as taking the determinant of A^T along the i^{th} column of A^T.

(2) We proceed by cases.

<u>Case 1:</u> Suppose $m = 2$. We proceed by induction on k where $A_1 \in M_k(\mathbb{F})$.

<u>$k = 1$:</u> $A_1 \in M_1(\mathbb{F})$, the result is trivial by taking the determinant along the first column as $A^{1,1} = A_2$ and $A_{2,1} = 0$.

<u>$k - 1 \Rightarrow k$:</u> Suppose $A_1 \in M_k(\mathbb{F})$ with $k > 1$. Then taking the determinant along the first column, we have

$$A^{i,1} = \begin{pmatrix} (A_1)^{i,1} & * \\ 0 & A_2 \end{pmatrix} \text{ for all } i \leq k,$$

so applying the induction hypothesis, we have $\det(A^{i,1}) = \det((A_1)^{i,1})\det(A_2)$ for all $i \leq k$. Then as $A_{i,1} = 0$ for all $i > k$, we have

$$\det(A) = \sum_{i=1}^{k} (-1)^{1+i} A_{i,1} \det(A^{i,1}) = \sum_{i=1}^{k} (-1)^{i+1} A_{i,1} \det((A_1)^{i,1}) \det(A_2) = \det(A_1)\det(A_2).$$

<u>Case 2:</u> Suppose $m > 2$, and set

$$B_1 = \begin{pmatrix} A_2 & & * \\ & \ddots & \\ 0 & & A_m \end{pmatrix} \Longrightarrow A = \begin{pmatrix} A_1 & * \\ 0 & B_1 \end{pmatrix}.$$

Applying case 1 gives $\det(A) = \det(A_1)\det(B_1)$. We may repeat this trick to peel off the A_i's to get

$$\det(A) = \prod_{i=1}^{m} A_i.$$

□

Corollary 5.2.10. *If A is block lower triangular or block diagonal, then* $\det(A)$ *is the product of the determinants of the blocks on the diagonal. If A is upper triangular, lower triangular, or diagonal, then* $\det(A)$ *is the product of the diagonal entries.*

Proposition 5.2.11. *Suppose $A \in M_n(\mathbb{F})$, and let $E \in M_n(\mathbb{F})$ be an elementary matrix. Then* $\det(EA) = \det(E)\det(A)$.

Proof. There are four cases depending on the type of elementary matrix.

Case 1: If $E = I$, the result is trivial.

Case 2: We must show

(a) the determinant of the matrix E obtained by switching two rows of I is -1, and

(b) switching two rows of A switches the sign of $\det(A)$.

Note that the result is trivial if the two rows are adjacent by 5.2.7. To get the result if the rows are not adjacent, we note that the interchanging of two rows can be accomplished by an odd number of adjacent switches. If we want to switch rows i and j with $i < j$, we switch j with $j-1$, then $j-1$ with $j-2$, all the way to $i+1$ with i for a total of $j-i$ switches. We then switch $i+1$ with $i+2$ as the old i^{th} row is now in the $(i+1)^{\text{th}}$ place, we switch $i+2$ with $i+3$, all the way up to switching $j-1$ with j for a total of $j-i-1$ switches. Hence, we switch a total of $2j-2i-1$ adjacent rows, which is always an odd number, and the result holds.

Case 3: We must show

(a) the determinant of the matrix E obtained from the identity by multiplying a row by a nonzero constant λ is λ, and

(b) multiplying a row of A by a nonzero constant λ changes the determinant by multiplying by λ.

We see (a) immediately holds from 5.2.9 as E is diagonal, and (b) immediately holds by 5.2.7 by expanding along the row multiplied by λ. If the i^{th} row is multiplied by λ, then

$$\det(EA) = \sum_{j=1}^{n}(-1)^{i+j}(EA)_{i,j}\det((EA)^{i,j}) = \sum_{j=1}^{n}(-1)^{i+j}\lambda A_{i,j}\det(A^{i,j}) = \det(E)\det(A)$$

as $\lambda = \det(E)$ and $(EA)^{i,j} = A^{i,j}$ as only the i^{th} row differs between the two matrices.

Case 4: We must show

(a) the determinant of the matrix E obtained from the identity by adding a constant multiple of one row to another row is 1, and

(b) adding a constant multiple of the i^{th} row of A to the k^{th} row of A with $k \neq i$ does not change the determinant of A.

To prove this result, we need a lemma:

Lemma 5.2.12. *Suppose $A \in M_n(\mathbb{F})$ has two identical rows. Then* $\det(A) = 0$.

Proof. We see from Case 2 that if we switch two rows of A, the determinant changes sign. As A has two identical rows, switching these rows does not change the sign of the determinant, so the determinant must be zero. □

Once again, note (a) is trivial by 5.2.9 as E is either upper or lower triangular. To show (b), we note that $(EA)^{k,j} = A^{k,j}$ for all $j \in [n]$, so by 5.2.7

$$\begin{aligned}\det(EA) &= \sum_{j=1}^{n}(-1)^{k+j}(EA)_{k,j}\det((EA)^{k,j}) = \sum_{j=1}^{n}(-1)^{k+j}(A_{i,j}+A_{k,j})\det(A^{k,j})\\ &= \underbrace{\sum_{j=1}^{n}(-1)^{i+j}A_{i,j}\det(A^{k,j})}_{\det(B)} + \underbrace{\sum_{j=1}^{n}(-1)^{k+j}A_{k,j}\det(A^{k,j})}_{\det(A)} = \det(A)\end{aligned}$$

by 5.2.12 as B is the matrix obtained from A by replacing the k^{th} row with the i^{th} row, so two rows of B are the same, and $\det(B) = 0$. □

Theorem 5.2.13. *$A \in M_n(\mathbb{F})$ is invertible if and only if* $\det(A) \neq 0$.

Proof. There is a unique matrix U in reduced row echelon form such that $A = E_n \cdots E_1 U$ for elementary matrices $E_1, \dots, E_n$. By iterating 5.2.11, we have

$$\det(A) = \det(E_n) \cdots \det(E_1)\det(U),$$

which is nonzero if and only if $\det(U) \neq 0$. Now $\det(U) \neq 0$ if and only if $U = I$ as U is in reduced row echelon form, so $\det(A) \neq 0$ if and only if A is row equivalent to I if and only if A is invertible. □

Proposition 5.2.14. *Suppose $A, B \in M_n(\mathbb{F})$. Then* $\det(AB) = \det(A)\det(B)$.

Proof. If A is not invertible, then AB is not invertible by 3.2.18, so $\det(AB) = \det(A)\det(B) = 0$ by 5.2.13.

Now suppose A is invertible. Then there are elementary matrices $E_1, \dots, E_n$ such that $A = E_1 \cdots E_n$ as A is row equivalent to I. By repeatedly applying 5.2.11, we get

$$\det(AB) = \det(E_1 \cdots E_n B) = \det(E_1) \cdots \det(E_n)\det(B) = \det(A)\det(B).$$

□

Corollary 5.2.15. *For $A, B \in M_n(\mathbb{F})$,* $\det(AB) = \det(BA)$.

Proposition 5.2.16.

(1) Suppose $S \in M_n(\mathbb{F})$ is invertible. Then $\det(S^{-1}) = \det(S)^{-1}$

(2) Suppose $A, B \in M_n(\mathbb{F})$ *and* $A \sim B$. *Then* $\det(A) = \det(B)$.

Proof.

(1) By 5.2.14, we have

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

(2) By 5.2.14 and (1),

$$\det(A) = \det(S^{-1}BS) = \det(S^{-1})\det(B)\det(S) = \det(S^{-1})\det(S)\det(B) = \det(B).$$

□

Proposition 5.2.17. *Suppose* $A, B | in M_n(\mathbb{F})$.

(1) $\operatorname{trace}(AB) = \operatorname{trace}(BA)$.

(2) If $A \sim B$, *then* $\operatorname{trace}(A) = \operatorname{trace}(B)$.

Proof. Exercise. □

Exercises

V will denote a finite dimensional vector space over $\mathbb{F}$.

Exercise 5.2.18 (A Faithful Representation of S_n)**.** Show that the symmetric group S_n can be embedded into $L(V)$ where $\dim(V) = n$, i.e. there is an injective function $\Phi\colon S_n \to L(V)$ such that $\Phi(\sigma\tau) = \Phi(\sigma)\Phi(\tau)$ for all $\sigma, \tau \in S_n$.

Hint: Use operators like T defined in 5.1.12.

Note: *If G is a group, a function $\Phi\colon G \to L(V)$ such that $\Phi(gh) = \Phi(g)\Phi(h)$ is called a representation of G. An injective representation is usually called a faithful representation.*

5.3 The Characteristic Polynomial

For this section, V will denote a finite dimensional vector space over $\mathbb{F}$.

Definition 5.3.1.

(1) For $A \in M_n(\mathbb{F})$, define the characteristic polynomial of A, denoted $\operatorname{char}_A \in \mathbb{F}[z]$, by $\operatorname{char}_A(z) = \det(zI - A)$. Note that $\operatorname{char}_A \in \mathbb{F}[z]$ is a monic polynomial of degree n.

(2) Let $T \in L(V)$. Define the characteristic polynomial of T, denoted $\operatorname{char}_T \in \mathbb{F}[z]$, by

$$\operatorname{char}_T(z) = \det(zI - [T]_B)$$

where B is some basis for V. Note that this is well defined as if C is another basis of V, then $[T]_B \sim [T]_C$ by 3.4.13, so $zI - [T]_B \sim zI - [T]_C$. Now we apply 5.2.16.

Remark 5.3.2. Note that $\operatorname{char}_A = \operatorname{char}_{L_A}$ if $A \in M_n(\mathbb{F})$.

Examples 5.3.3.

(1) If $A_\pm \in M_2(\mathbb{F})$ is given by

$$A_\pm = \begin{pmatrix} 0 & \pm 1 \\ 1 & 0 \end{pmatrix},$$

we see that $\operatorname{char}_{A_\pm}(z) = \det(zI - A_\pm) = z^2 \mp 1$.

(2)

Remark 5.3.4. If $A \sim B$, then $\operatorname{char}_A = \operatorname{char}_B$.

Proposition 5.3.5. *Let $T \in L(V)$. Then $\lambda \in \operatorname{sp}(T)$ if and only if λ is a root of* char_T *and $\lambda \in \mathbb{F}$.*

Proof. This is immediate from 5.2.13 and 3.4.12. □

Remark 5.3.6. 5.3.5 tells us that one way to find $\operatorname{sp}(T)$ is to first pick a basis B of V, find $[T]_B$, and then compute the roots of the polynomial

$$\operatorname{char}_T(z) = \det(zI - [T]_B).$$

The problem with this technique is that it is usually very difficult to factor polynomials of high degree. For example, there are quadratic, cubic, and quartic equations for calculating roots of polynomials of degree less than or equal to 4, but there is no formula for finding roots of polynomials with degree greater than or equal to 5. Hence, if $\dim(V) \geq 5$, there is no good way known to factor the characteristic polynomial.

Another problem is that it is very hard to calculate determinants without the use of computers. Even with a computer, we can only calculate determinants to a certain degree of accuracy, so we are still unsure of the spectrum of the matrix if the characteristic polynomial obtained in the fashion can be factored.

Examples 5.3.7. We will calculate the spectrum for some operators.

(1)

(2)

Exercises

V will denote a finite dimensional vector space over $\mathbb{F}$.

Exercise 5.3.8 (Roots of Unity)**.** Suppose V is a finite dimensional vector space over $\mathbb{C}$ with ordered basis $B = (v_1, \dots, v_n)$, and let $T \in L(V)$ be the finite shift operator defined in 5.1.12. Compute $\operatorname{char}_T(z) = \det(zI - [T]_B)$, and relate your answer to 5.1.12.

Exercise 5.3.9. Compute the characteristic polynomial of

(1) the operator $L_B \in L(\mathbb{F}^n)$ where $B = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix} \in M_n(\mathbb{F})$ and $\lambda \in \mathbb{F}$, and

(2) the operator $L_C \in L(\mathbb{F}^n)$ where $C = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ & \ddots & \ddots & \vdots & \vdots \\ & & \ddots & 0 & -a_{n-2} \\ 0 & & & 1 & -a_{n-1} \end{pmatrix} \in M_n(\mathbb{F})$ and $a_i \in \mathbb{F}$ for all $i \in [n-1]$.

5.4 The Minimal Polynomial

For this section, V will denote a finite dimensional vector space over $\mathbb{F}$.

Definition-Proposition 5.4.1. *Recall that if V is finite dimensional, then $L(V)$ is finite dimensional. For $T \in L(V)$, the set*

$$\{T^n \big| n \in \mathbb{Z}_{\geq 0}\}$$

where $T^0 = I$ cannot be linearly independent, so there is a unique monic polynomial $p \in \mathbb{C}[z]$ of minimal degree ≥ 1 such that $p(T) = 0$. This polynomial is called the minimal polynomial, and it is denoted $\min_T$.

Proof. We must show the polynomial p is unique. If $q \in \mathbb{C}[z]$ is monic with $\deg(p) = \deg(q)$, by the Euclidean Algorithm 4.2.1, there are polynomials unique $k, r \in \mathbb{C}[z]$ such that $p = kq + r$ and $\deg(r) < \deg(q)$. Since $p(T) = q(T) = 0$, we must also have that $r(T) = 0$, so $r = 0$ as p was chosen of minimal degree. Thus $p = kq$, and since $\deg(p) = \deg(q)$, k must be constant. Since p and q are both monic, the constant k must be 1. □

Examples 5.4.2.

(1) If $T = 0$, we have that $\min_T(z) = z$ as this polynomial is a linear polynomial which gives the zero operator when evaluated at T.

(2) If $T = I$, we have that $\min_T(z) = z - 1$ for similar reasoning as above (recall that $1(T) = I$, i.e. the constant polynomial 1 evaluated at T is I).

(3) We have that $\min_T$ is linear if and only if $T = \lambda I$ for some $\lambda \in \mathbb{F}$.

Proposition 5.4.3. *Let $p \in \mathbb{F}[z]$. Then $p(T) = 0$ if and only if $\min_T \mid p$.*

Proof. It is obvious that if $\min_T \mid p$, then $p(T) = 0$. Suppose $p(T) = 0$. Since $\min_T$ has minimal degree such that $\min_T(T) = 0$, we have $\deg(p) \geq \deg(\min_T)$. By 4.2.1, there are unique $k, r \in \mathbb{C}[z]$ with $\deg(r) < \deg(\min_T)$ and $p = k \min_T + r$. Then since $p(T) = \min_T(T) = 0$, we have $r(T) = 0$, so $r = 0$ as $\min_T$ was chosen of minimal degree. Hence $\min_T \mid p$. □

Proposition 5.4.4. *For $T \in L(V)$, $\lambda \in \operatorname{sp}(T)$ if and only if λ is a root of* $\min_T$.

Proof. First, suppose $\lambda \in \operatorname{sp}(T)$ corresponding to eigenvector v. Then

$$\min_T(T)v = \min_T(\lambda)v = 0,$$

so λ is a root of $\min_T$. Now suppose λ is a root of $\min_T$. Then $(z - \lambda)$ divides $\min_T(z)$, so there is a $p \in \mathbb{C}[z]$ with $\min_T(z) = (z - \lambda)^n p(z)$ for some $n \in \mathbb{N}$ such that λ is not a root of $p(z)$. By 5.4.3, $p(T) \neq 0$ as $\min_T \nmid p$, so there is a v such that $w = p(T)v \neq 0$. Now we must have tha

$$0 = \min_T(T)v = (T - \lambda I)^n p(T)v = (T - \lambda I)^n w.$$

Hence $(T - \lambda I)^k w$ is an eigenvector for T corresponding to the eigenvalue λ for some $k \in \{0, \dots, n-1\}$. □

Corollary 5.4.5 (Existence of Eigenvalues)**.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in L(V)$. Then T has an eigenvalue.*

Proof. We know that $\min_T \in \mathbb{C}[z]$ has a root by 4.5.4. The result follows immediately by 5.4.4. □

Remark 5.4.6. More generally, $T \in L(V)$ has an eigenvalue if V is a vector space over an algebraically closed field.

Exercises

V will denote a finite dimensional vector space over $\mathbb{F}$.

Exercise 5.4.7. Find the minimal polynomial of the following operators:

(1) The operator $L_A \in L(\mathbb{R}^4)$ where $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \\ 0 & 0 & 0 & 10 \end{pmatrix}$.

(2) The operator $L_B \in L(\mathbb{F}^n)$ where $B = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix} \in M_n(\mathbb{F})$ and $\lambda \in \mathbb{F}$.

(3) The operator $L_C \in L(\mathbb{F}^n)$ where $C = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ & \ddots & \ddots & \vdots & \vdots \\ & & \ddots & 0 & -a_{n-2} \\ 0 & & & 1 & -a_{n-1} \end{pmatrix} \in M_n(\mathbb{F})$ and $a_i \in \mathbb{F}$ for all $i \in [n-1]$.

(4) The finite shift operator T in 5.1.12.

Chapter 6

Operator Decompositions

We begin with the easiest type of decomposition, which is matrix decomposition via idempotents. We then discuss diagonalization and a generalization of diagonalization called primary decomposition.

6.1 Idempotents

For this section, V will denote a vector space over $\mathbb{F}$, and U, W will denote subspaces of V such that $V = U \oplus W$. Please note that the results of this section are highly dependent on the specific direct sum decomposition $V = U \oplus W$.

Definition 6.1.1. An operator $E \in L(V)$ is called idempotent if $E^2 = E \circ E = E$.

Examples 6.1.2.

(1) The zero operator and the identity operator are idempotents.

(2) Let $A \in M_n(\mathbb{F})$ be given by

$$A = \frac{1}{n}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix},$$

and let $E = L_A$. Then $E^2 = E$ as $A^2 = A$.

Facts 6.1.3. *Suppose V is finite dimensional and $E \in L(V)$ is an idempotent.*

(1) As $E^2 = E$, so we know $E^2 - E = 0$. If $E = 0$, then we know $\min_E(z) = p_1(z) = z$, and if $E = I$, then $\min_E(z) = p_2(z) = z - 1$, both of which divide $p_3(z) = z^2 - z$. If $E \neq 0, I$, then we have that $p_1(E) \neq 0$ and $p_2(E) \neq 0$, but $p_3(E) = 0$. As $E \neq \lambda I$ for any $\lambda \in \mathbb{F}$, we have $\deg(\min_E) \geq 2$, so $\min_E = p_3$.

(2) We see that if E is a nontrivial idempotent, i.e. $E \neq 0, I$, then $\operatorname{sp}(E) = \{0, 1\}$ as these are the only roots of $z^2 - z$.

Definition-Proposition 6.1.4. *Define $E_U\colon V\to V$ by $E_U(v)=u$ if $v=u+w$ with $u\in U$ and $w\in W$. Note that E_U is well defined since by 2.2.8, for each $v\in V$ there are unique $u\in U$ and $w\in W$ such that $v=u+w$. Then*

(1) E_U is a linear operator,

(2) $E_U^2=E_U$,

(3) $\ker(E_U)=W$, and

(4) $\operatorname{im}(E_U)=\{v\in V \big| E_U(v)=v\}=U$.

Thus E_U is an idempotent called the idempotent onto U along W. Note that we also have $E_W\in L(V)$, the idempotent onto W along U which satisfies conditions (1)-(4) above after switching U and W. The operators $E_U,E_W\in L(V)$ satisfy

(5) $E_WE_U=0=E_UE_W$, and

(6) $I=E_U+E_W$.

Proof.

(1) Let $v=u_1+w_1$, $v_2=u_2+w_2$, and $\lambda\in\mathbb{F}$ where $u_i\in U$ and $w_i\in W$ for $i=1,2$. Then

$$E_U(\lambda v_1+v_2)=E_U(\underbrace{(\lambda u_1+u_2)}_{\in U}+\underbrace{(\lambda w_1+w_2)}_{\in W})=\lambda u_1+u_2=\lambda E_U(v_1)+E_U(v_2).$$

(2) If $v=u+w$, then $E_U^2v=E_UE_U(v)=E_Uu=u=E_Uv$, so $E_U^2=E_U$.

(3) $E_U(v)=0$ if and only if the U-component of V is zero if and only if $v\in W$.

(4) $E_U(v)=v$ if and only if the W-component of v is zero if and only if $v\in U$. This shows $U=\{v\in V \big| E_U(v)=v\}\subseteq\operatorname{im}(E_U)$. Suppose now that $v\in\operatorname{im}(E_U)$. Then there is an $x\in V$ such that $E_U(x)=v$. By (2), $v=E_Ux=E_U^2x=E_Uv$, so $v\in\{v\in V \big| E_U(v)=v\}=U$.

(5) If $v=u+w$, we have

$$E_WE_U(v)=E_WE_U(u+w)=E_W(u)=0=E_U(w)=E_UE_W(u+w)=E_UE_Wv.$$

Hence $E_WE_U=0=E_UE_W$.

(6) If $v=u+w$, we have

$$(E_U+E_W)v=E_U(u+w)+E_W(u+w)=u+w=v.$$

Hence $E_U+E_W=I$. □

Remark 6.1.5. It is very important to note that E_U is dependent on the complementary subspace W. For example, if we set

$$U=\operatorname{span}\left\{\begin{pmatrix}1\\0\end{pmatrix}\right\}\subset\mathbb{F}^2,\ W_1=\operatorname{span}\left\{\begin{pmatrix}0\\1\end{pmatrix}\right\}\subset\mathbb{F}^2,\ \text{and } W_2=\operatorname{span}\left\{\begin{pmatrix}1\\1\end{pmatrix}\right\}\subset\mathbb{F}^2,$$

then $U \oplus W_1 = V = U \oplus W_2$. Applying 6.1.4 using W_1, we have that

$$E_U \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

but using W_2, we would have

$$E_U \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Proposition 6.1.6. *Suppose $E \in L(V)$ with $E = E^2$. Then*

(1) $I - E \in L(V)$ satisfies $(I - E)^2 = (I - E)$,

(2) $\operatorname{im}(E) = \{v \in V \big| Ev = v\}$,

(3) $V = \ker(E) \oplus \operatorname{im}(E)$,

(4) E is the idempotent onto $\operatorname{im}(E)$ along $\ker(E)$, and

(5) $I - E$ is the idempotent onto $\ker(E)$ along $\operatorname{im}(E)$.

Proof.

(1) $(I - E)^2 = I - 2E + E^2 = I - 2E + E = I - E$.

(2) It is clear $\{v \in V \big| Ev = v\} \subseteq \operatorname{im}(E)$. If $v \in \operatorname{im}(E)$, then there is a $u \in V$ such that $Eu = v$. Then $v = Eu = E^2u = Ev$.

(3) Suppose $v \in \ker(E) \cap \operatorname{im}(E)$. Then by (2), $v = Ev = 0$, so $v = 0$, and $\ker(E) \cap \operatorname{im}(E) = (0)$. Let $v \in V$, and set $u = Ev$ and $w = v - u$. Then $u \in \operatorname{im}(E)$ and $Ew = Ev - Eu = u - u = 0$, so $w \in \ker(E)$, and $v = w + u \in \ker(E) + \operatorname{im}(E)$.

(4) This follows immediately from 6.1.4.

(5) We have that $\ker(I - E) = \operatorname{im}(E)$ and $\operatorname{im}(I - E) = \ker(E)$, so the result follows from (4) (or 6.1.4). □

Exercises

Exercise 6.1.7 (Idempotents and Direct Sum)**.** Show that

$$V = \bigoplus_{i=1}^{n} W_i$$

for nontrivial subspaces $W_i \subset V$ for $i \in [n]$ if and only if there are nontrivial idempotents $E_i \in L(V)$ for $i \in [n]$ such that

(1) $\displaystyle\sum_{i=1}^{n} E_i = I$ and

(2) $E_iE_j = 0$ for all $i \neq j$.

Note: *In this case, we can define $T_{i,j} \in L(W_j, W_i)$ by $T_{i,j} = E_iTE_j$, and note that*

$$\widetilde{T} = (T_{i,j}) \in L(W_1 \overline{\oplus} \cdots \overline{\oplus} W_n) \cong L(V).$$

6.2 Matrix Decomposition

Definition 6.2.1. Let U, W be vector spaces. The external direct sum of U and W is the vector space

$$U\overline{\oplus}W = \left\{ \begin{pmatrix} u \\ w \end{pmatrix} \middle| u \in U \text{ and } w \in W \right\}$$

where addition and scalar multiplication are defined in the obvious way:

$$\begin{pmatrix} u_1 \\ w_1 \end{pmatrix} + \begin{pmatrix} u_2 \\ w_2 \end{pmatrix} = \begin{pmatrix} u_1 + u_2 \\ w_1 + w_2 \end{pmatrix} \text{ and } \lambda \begin{pmatrix} u \\ w \end{pmatrix} = \begin{pmatrix} \lambda u \\ \lambda w \end{pmatrix}$$

for all $u, u_1, u_2 \in U$, $w, w_1, w_2 \in W$, and $\lambda \in \mathbb{F}$.

Notation 6.2.2 (Matrix Decomposition)**.** Since $V = U \oplus W$, there is a canonical isomorphism $V = U \oplus W \cong U\overline{\oplus}W$ given by

$$u + w \longmapsto \begin{pmatrix} u \\ w \end{pmatrix}$$

with obvious inverse.

If $T \in L(V)$, then by 6.1.4,

$$T = (E_U + E_W)T(E_U + E_W) = E_U T E_U + E_U T E_W + E_W T E_U + E_W T E_W.$$

If we set $T_1 = E_U T E_U \in L(U)$, $T_2 = E_U T E_W \in L(W, U)$, $T_3 = E_W T E_U \in L(U, W)$, and $T_4 = E_W T E_W \in L(W)$, define the operator

$$\widetilde{T} = \begin{pmatrix} T_1 & T_2 \\ T_3 & T_4 \end{pmatrix} : U\overline{\oplus}W \to U\overline{\oplus}W.$$

by matrix multiplication, i.e. if $u \in U$ and $w \in W$, define

$$\begin{pmatrix} T_1 & T_2 \\ T_3 & T_4 \end{pmatrix} \begin{pmatrix} u \\ w \end{pmatrix} = \begin{pmatrix} T_1 u + T_2 w \\ T_3 u + T_4 w \end{pmatrix}$$

The map given by $T \mapsto \widetilde{T}$ is an isomorphism

$$L(V) \cong \left\{ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \middle| A \in L(U),\ B \in L(W, U),\ C \in L(U, W), \text{ and } D \in L(W) \right\}$$

where addition and scalar multiplication are defined in the obvious way:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} + \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} = \begin{pmatrix} A + A' & B + B' \\ C + C' & D + D' \end{pmatrix} \text{ and } \lambda \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} \lambda A & \lambda B \\ \lambda C & \lambda D \end{pmatrix}$$

for all $A, A' \in L(U)$, $B, B' \in L(W, U)$, $C, C' \in L(U, W)$, $D, D' \in L(W)$, and $\lambda \in \mathbb{F}$.

This decomposition can be extended to the case when

$$V = \bigoplus_{i=1}^{n} W_i$$

for subspaces $W_i \subset V$ for $i = 1, \ldots, n$.

Exercises

Exercise 6.2.3. Suppose $V = U \oplus W$. Verify the claims made in 6.2.2, i.e.

1. $V \cong U \overline{\oplus} W$ and
2. $L(V) \cong \left\{ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \middle| A \in L(U),\ B \in L(W, U),\ C \in L(U, W),\ \text{and } D \in L(W) \right\}$.

Exercise 6.2.4.

Exercise 6.2.5. Suppose $T \in L(V)$, $V = U \oplus W$ for subspaces U, W, and

$$\widetilde{T} = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

where $A \in L(U)$, $B \in L(W, U)$, and $C \in L(W)$ as in 6.2.2. Show that $\operatorname{sp}(T) = \operatorname{sp}(A) \cup \operatorname{sp}(C)$.

6.3 Diagonalization

For this section, V will denote a finite dimensional vector space over $\mathbb{F}$. The main results of this section are characterization of when an operator $T \in L(V)$ is diagonalizable in terms of its associated eigenspaces and in terms of its minimal polynomial.

Definition 6.3.1. Let $T \in L(V)$. T is diagonalizable if there is a basis of V consisting of eigenvectors of T. A matrix $A \in M_n(\mathbb{F})$ is called diagonalizable if L_A is diagonalizable.

Examples 6.3.2.

(1) Every diagonal matrix is diagonalizable.

(2) There are matrices which are not diagonalizable. For example, consider

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

We see that zero is the only eigenvalue of L_A, but $\dim(E_0) = 1$. Hence there is not a basis of $\mathbb{R}^2$ consisting of eigenvectors of L_A.

Remark 6.3.3. $T \in L(V)$ is diagonalizable if and only if there is a basis B of V such that $[T]_B$ is diagonal.

Theorem 6.3.4. *$T \in L(H)$ is diagonalizable if and only if*

$$V = \bigoplus_{i=1}^{m} E_{\lambda_i}$$

where $\operatorname{sp}(T) = \{\lambda_1, \ldots, \lambda_m\}$.

Proof. Certainly

$$W = \bigoplus_{i=1}^{m} E_{\lambda_i}$$

is a subspace of V by 5.1.9.

Suppose T is diagonalizable. Then there is a basis of V consisting of eigenvectors of T. Thus $W = V$ as every element of V can be written as a linear combination of eigenvectors of T.

Suppose now that $V = W$. Let B_i be a basis for E_{λ_i} for $i = 1, \dots, m$. Then by 2.4.13,

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

is a basis for B. Since B consists of eigenvectors of T, T is diagonalizable. □

Corollary 6.3.5. *Let* $\operatorname{sp}(T) = \{\lambda_1, \dots, \lambda_n\}$ *and let* $n_i = \dim(E_{\lambda_i})$. *Then* $T \in L(V)$ *is diagonalizable if and only if* $\displaystyle\sum_{i=1}^{n} n_i = \dim(V)$.

Remark 6.3.6. Suppose $A \in M_n(\mathbb{F})$ is diagonalizable. Then there is a basis $\{v_1, \dots, v_n\}$ of $\mathbb{F}^n$ consisting of eigenvectors of L_A corresponding to eigenvalues $\lambda_1, \dots, \lambda_n$. Form the matrix

$$S = [v_1|v_2|\cdots|v_n],$$

which is invertible by 3.2.15, and note that

$$S^{-1}AS = S^{-1}A[v_1|\cdots|v_n] = S^{-1}[\lambda_1 v_1|\cdots|\lambda_n v_n] = S^{-1}S\operatorname{diag}(\lambda_1, \dots, \lambda_n) = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

where $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ is the diagonal matrix in $M_n(\mathbb{F})$ whose $(i,i)^{\text{th}}$ entry is λ_i. Hence, there is an invertible matrix S such that $S^{-1}AS = D$, a diagonal matrix. This is the usual way diagonalizability is presented in a matrix theory course. We show the converse holds, so that this definition of diagonalizability is equivalent to the one given in these notes.

Suppose $S^{-1}AS = D$, a diagonal matrix for some invertible $S \in M_n(\mathbb{F})$. Then we know $CS(S) = \mathbb{F}^n$ by 3.2.18, and if $B = \{v_1, \dots, v_n\}$ are the columns of S, we have

$$[Av_1|\cdots|Av_n] = AS = SD = [D_{1,1}v_1|\dots|D_{n,n}\lambda_n],$$

so B is a basis of $\mathbb{F}^n$ consisting of eigenvectors of L_A, and A is diagonalizable.

Exercises

Exercise 6.3.7. Suppose $T \in L(V)$ is diagonalizable and $W \subset V$ is a T-invariant subspace. Show $T|_W$ is diagonalizable.

Exercise 6.3.8. Suppose V is finite dimensional. Let $S, T \in L(V)$ be diagonalizable. Show that $S, T \in L(V)$ commute if and only if S, T are simultaneously diagonalizable, i.e. there is a basis of V consisting of eigenvectors of both S and T.

Hint: First show that the eigenspaces E_λ of T are S-invariant. Then show $S_\lambda = S|_{E_\lambda} \in L(E_\lambda)$ is diagonalizable for all $\lambda \in \operatorname{sp}(T)$.

6.4 Nilpotents

Definition 6.4.1.

(1) An operator $N \in L(V)$ is called nilpotent if $N^n = 0$ for some $n \in \mathbb{N}$. We say N is nilpotent of order n if n is the smallest $n \in \mathbb{N}$ such that $N^n = 0$.

Examples 6.4.2.

(1) The matrix

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

is nilpotent of order 2. In fact,

$$A_n = \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix} \in M_n(\mathbb{F})$$

is nilpotent of order n for all $n \in \mathbb{N}$.

(2) The block matrix

$$N = \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix}$$

where $0, B \in M_n(\mathbb{F})$ for some $n \in \mathbb{N}$ is nilpotent of order 2 in $M_{2n}(\mathbb{F})$.

Exercises

6.5 Generalized Eigenvectors

For this section, V will denote a finite dimensional vector space over $\mathbb{F}$. This main result of this section is Theorem 12 of Section 6.8 in [3]

Definition 6.5.1. Let $T \in L(V)$, and suppose $p \in \mathbb{F}[z]$ is a monic irreducible factor of $\min_T$.

(1) There is a maximal $m \in \mathbb{N}$ such that $p^m \mid \min_T$. Define $K_p = \ker(p(T)^m)$.

(2) Define $G_p = \big\{v \in V \big| p(T)^n v = 0 \text{ for some } n \in \mathbb{N}\big\}$.

(3) If $\deg(p) = 1$, then $p(z) = z - \lambda$ for some $\lambda \in \operatorname{sp}(T)$ by 5.4.4, and we set $G_\lambda = G_p$. In this case, elements of $G_\lambda \setminus \{0\}$ are called generalized eigenvectors of T corresponding the the eigenvalue λ. We say the multiplicity of λ is $M_\lambda = \dim(G_\lambda)$.

Remarks 6.5.2.

(1) Note that G_p and K_p are both T-invariant subspaces of V.

(2) For $\lambda \in \operatorname{sp}(T)$, there can be most M_λ linearly independent eigenvectors corresponding to the eigenvalue λ as $E_\lambda \subset G_\lambda$.

Examples 6.5.3.

(1) Every eigenvector of $T \in L(V)$ is a generalized eigenvector of T.

(2) Suppose

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

We saw earlier that A has a one eigenvalue, namely zero, and that $\dim(E_0) = 1$ so A is not diagonalizable. However, we see that $A^2 = 0$, so every $v \in \mathbb{F}^2$ is a generalized eigenvector corresponding to the eigenvalue 0.

(3) Let $B \in M_4(\mathbb{R})$ be given by

$$B = \begin{pmatrix} 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

We calculate G_p for all monic irreducible $p \mid \min_{L_B}$. First, we calculate char_B. We see that $B - \lambda I$ is block upper triangular, so

$$\operatorname{char}_B(z) = \det(zI - B) = \begin{vmatrix} z & 1 & -1 & 0 \\ -1 & z & 0 & -1 \\ 0 & 0 & z & 1 \\ 0 & 0 & -1 & z \end{vmatrix} = \begin{vmatrix} z & 1 \\ -1 & z \end{vmatrix} \begin{vmatrix} z & 1 \\ -1 & z \end{vmatrix} = (z^2+1)^2.$$

We claim $\operatorname{char}_B = \min_{L_B}$. Note that

$$(z^2+1)|_B = B^2 + I = \begin{pmatrix} -1 & 0 & 0 & -2 \\ 0 & -1 & 2 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -2 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

so we see immediately that $(B^2 + I)^2 = 0$. We now have a candidate for $\min_{L_B}$. Set $p(z) = (z^2+1)^2$. As $p(B) = 0$, we have that $\min_{L_B} \mid p$, but as the only monic irreducible factor of p is $q(z) = z^2+1$ and $q(B) \neq 0$, we must have that $\min_{L_B} = p$. Hence $K_p = G_p = V$ as $p(B)^2 v = 0v = 0$ for all $v \in V$.

Remark 6.5.4. If $T \in L(V)$ is nilpotent, then every $v \in V \setminus \{0\}$ is a generalized eigenvector of T corresponding to the eigenvalue zero.

Lemma 6.5.5. *Let $T \in L(V)$, and suppose $p, q \in \mathbb{F}[z]$ are two distinct monic irreducible factors of $\min_T$.*

(1) $G_p \cap G_q = (0)$.

(2) $q(T)|_{G_p}$ is bijective.

Proof.

(1) Suppose $v \in G_p \cap G_q$. Then there are $n, m \in \mathbb{N}$ such that $p(T)^n v = 0 = q(T)^m v$. But as p^n, q^m are relatively prime, by 4.3.8, there are $f, g \in \mathbb{F}[z]$ such that $fp^n + gq^m = 1$. Thus,

$$v = 1(T)v = f(T)p(T)^n v + g(T)q(T)^m v = 0,$$

and we are finished.

(2) First, we note that G_p is T-invariant, so it is $q(T)$-invariant. Next, suppose $q(T)v = 0$ for some $v \in G_p$. Then $v \in G_p \cap G_q = (0)$ by (1), so $v = 0$. Thus $q(T)|_{G_p}$ is injective and thus bijective by 3.2.15. □

Proposition 6.5.6. *Let $T \in L(V)$, and suppose $p \in \mathbb{F}[z]$ is a monic irreducible factor of $\min_T$. Then $K_p = G_p$.*

Proof. Certainly $K_p \subset G_p$. Let m be the multiplicity of p, and let f be the unique monic polynomial such that $\min_T = fp^m$. We know G_p is $f(T)$-invariant, and as $f = q_1 \cdots q_m$ for monic irreducible $q_i \neq p$ which divide $\min_T$ for all $i \in [m]$, by 6.5.5, $q_i(T)$ is bijective, and thus so is

$$f(T)|_{G_p} = (q_1 \cdots q_m)(T)|_{G_p} = (q_1(T)|_{G_p}) \cdots (q_m(T)|_{G_p}).$$

Thus, if $x \in G_p$, then there is a $y \in G_p$ such that $f(T)y = x$, so

$$0 = \min_T(T)y = (fp^m)(T)y = p(T)^m f(T)y = p(T)^m x,$$

and $x \in K_p$. □

Exercises

6.6 Primary Decomposition

Theorem 6.6.1 (Primary Decomposition)**.** *Suppose $T \in L(V)$, and suppose*

$$\min_T = p_1^{r_1} \cdots p_n^{r_n}$$

where the $p_j \in \mathbb{F}[z]$ are distinct irreducible polynomials for $i \in [n]$. Then

(1) $V = \bigoplus_{i=1}^{n} K_{p_i} = \bigoplus_{i=1}^{n} G_{p_i}$ and

(2) If $T_i = T|_{K_{p_i}} \in L(K_{p_i})$ for $i \in [n]$, $\min_{T_i} = p_j^{r_i}$.

Proof.

(1) For $i \in [n]$, set

$$f_i = \frac{\min_T}{p_i^{r_i}} = \prod_{j \neq i} p_j^{r_j},$$

and note that $\min_T | f_i f_j$ for all $i \neq j$. The f_i's are relatively prime, so there are polynomials $g_i \in \mathbb{F}[z]$ such that

$$\sum_{i=1}^n f_i g_i = 1.$$

Set $h_i = f_i g_i$ for $i \in [n]$ and $E_i = h_i(T)$. First note that

$$\sum_{i=1}^n E_i = \sum_{i=1}^n h_i(T) = \left(\sum_{i=1}^n h_i \right)(T) = q(T) = I$$

where $q \in \mathbb{F}[z]$ is the constant polynomial given by $q(z) = 1$. Moreover,

$$E_i E_j = h_i(T) h_j(T) = (h_i h_j)(T) = (g_i g_j f_i f_j)(T) = 0$$

by 5.4.3 as $\min_T | f_i f_j$. Hence the E_i's are idempotents:

$$E_j = E_j I = E_j \left(\sum_{i=1}^n E_i \right) = \sum_{i=1}^n E_j E_i = E_j^2.$$

Since they sum to I, they corrspond to a direct sum decomposition of V:

$$V = \bigoplus_{i=1}^n \operatorname{im}(E_i).$$

It remains to show $\operatorname{im}(E_i) = K_{p_i}$. If $v \in \operatorname{im}(E_i)$, then $v = E_i v$, so

$$p_i(T)^{r_i} v = p_i(T)^{r_i} E_i v = p_i(T)^{r_i} h_i(T) v = (p_i^{r_i} f_i g_i)(T) v = (\min_T g_i)(T) v = 0.$$

Hence $v \in K_{p_i}$ and $\operatorname{im}(E_i) \subset K_{p_i}$. Now suppose $v \in K_{p_i}$. If $j \neq i$, then $f_j g_j(T) v = 0$ as $p_i^{r_i} | f_j$. Thus $E_j v = h_j(T) v = (f_j g_j) v = 0$, and

$$E_i v = \left(\sum_{j=1}^n E_j v \right) = \left(\sum_{j=1}^n E_j \right) v = I v = v.$$

Hence $v \in \operatorname{im}(E_i)$, and $K_{p_i} \subseteq \operatorname{im}(E_i)$.

(2) Since $p_i(T_i)^{r_i} \in L(K_{p_i})$ is the zero operator, we have that $\min_{T_i} | (p_i)^{r_i}$. Conversely, if $g \in \mathbb{F}[z]$ such that $g(T_i) = 0$, then $g(T) f_i(T) = 0$ as $f_i(T)$ is only nonzero on K_{p_i}, but $g(T) = 0$ on K_{p_i}. That is, if $v \in V$, then by (1), we can write

$$v = \sum_{j=1}^n v_j \text{ where } v_j \in K_{p_j} \text{ for all } j \in [n],$$

and we have that

$$g(T)f_i(T)v = g(T)f_i(T)\sum_{j=1}^{n} v_j = g(T)\sum_{j=1}^{n} f_i(T)v_j = g(T)v_i = 0.$$

Hence $\min_T|gf_i$, and $p_i^{r_i}f_i$ divides gf_i. This implies $p_i^{r_i}|g$, so $\min_{T_i} = p_i^{r_i}$. □

Corollary 6.6.2. *$V = \bigoplus\limits_{\lambda\in\operatorname{sp}(T)} G_\lambda$ if and only if $\min_T$ splits (into linear factors).*

Corollary 6.6.3. *$T \in L(V)$ is diagonalizable if and only if $\min_T$ splits into distinct linear factors in $\mathbb{F}[z]$, i.e.*

$$\min_T(z) = \prod_{\lambda\in\operatorname{sp}(T)} (z-\lambda).$$

Proof. Suppose T is diagonalizable, and set

$$p(z) = \prod_{\lambda\in\operatorname{sp}(T)} (z-\lambda).$$

We claim $p = \min_T$, which will imply $\min_T$ splits into distinct linear factors in $\mathbb{F}[z]$. By 6.3.4,

$$V = \bigoplus_{\lambda\in\operatorname{sp}(T)} E_\lambda.$$

Let $v \in V$. Then by 2.2.8, there are unique $v_\lambda \in E_\lambda$ for each $\lambda \in \operatorname{sp}(T)$ such that

$$v = \sum_{\lambda\in\operatorname{sp}(T)} v_\lambda,$$

so it is clear that $p(T)v = 0$, and $p(T) = 0$. Furthermore, as $(z-\lambda)|\min_T(z)$ for all $\lambda \in \operatorname{sp}(T)$ by 5.4.4, $p = \min_T$.

Suppose now that

$$\min_T(z) = \prod_{\lambda\in\operatorname{sp}(T)} (z-\lambda).$$

Then by 6.6.1, we have that

$$V = \bigoplus_{\lambda\in\operatorname{sp}(T)} \ker(T-\lambda I) = \bigoplus_{\lambda\in\operatorname{sp}(T)} E_\lambda.$$

Hence T is diagonalizable by 6.3.4. □

Exercises

Chapter 7

Canonical Forms

For this chapter, V will denote a finite dimensional vector space over $\mathbb{F}$. We discuss the rational canonical form and the Jordan canonical form. Along the way, we prove the Cayley-Hamilton Theorem which relates the characteristic and minimal polynomials of an operator. We then will show to what extent these canonical forms are unique. Finally, we discuss the holomorphic functional calculus which is an application of the Jordan canonical form.

7.1 Cyclic Subspaces

Definition 7.1.1. Let $p \in \mathbb{F}[z]$ be the monic polynomial given by

$$p(z) = t^n + \sum_{i=0}^{n-1} a_i z^i.$$

Then the companion matrix for p is the matrix given by

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ & \ddots & \ddots & \vdots & \vdots \\ & & \ddots & 0 & -a_{n-2} \\ 0 & & & 1 & -a_{n-1} \end{pmatrix}$$

Examples 7.1.2.

(1) Suppose $p(z) = z - \lambda$. Then the companion matrix for p is $(\lambda) \in M_1(\mathbb{F})$.

(2) Suppose $p(z) = z^2 + 1 \in M_2(\mathbb{F})$. Then the companion matrix for p is

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

(3) If $p(z) = z^n$, then the companion matrix for p is

$$\begin{pmatrix} 0 & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{pmatrix}$$

Proposition 7.1.3. *Let $A \in M_n(\mathbb{F})$ be the companion matrix of $p \in \mathbb{F}[z]$. Then* $\operatorname{char}_A = \min_A = p$.

Proof. The routine exercise of checking that $p = \operatorname{char}_A$ and $p(A) = 0$ are left to the reader. Note that if $\{e_1, \dots, e_n\}$ is the standard basis of $\mathbb{F}^n$, we have that $Ae_i = e_{i+1}$ for $i \in [n-1]$, so $\left\{A^i e_1 \middle| i = 0, \dots, n-1\right\}$ is linearly independent. Thus $\deg(\min_A) \geq n$ as $q(A)e_1 \neq 0$ for all nonzero $q \in \mathbb{F}[z]$ with $\deg(q) < n$. As $\deg(p) = n$, p is a monic polynomial of least degree such that $p(A) = 0$, and thus $p = \min_A$. □

Definition 7.1.4. Suppose $T \in L(V)$.

(1) For $v \in V$, the T-cyclic subspace generated by v is $Z_{T,v} = \operatorname{span}\left\{T^n v \middle| n \in \mathbb{Z}_{\geq 0}\right\}$.

(2) A subspace $W \subset V$ is called T-cyclic if there is a $w \in W$ such that $W = Z_{T,w}$. This w is called a T-cyclic vector for W.

Examples 7.1.5.

(1) If $A \in M_n(\mathbb{F})$ is a companion matrix for $p \in \mathbb{F}[z]$, then $\mathbb{F}^n$ is L_A-cyclic with cyclic vector e_1 as $Ae_i = e_{i+1}$ for all $i \in [n-1]$.

(2)

Remarks 7.1.6.

(1) Every T-cyclic subspace is T-invariant. In fact, if $w \in Z_{T,v}$, then there is an $n \in \mathbb{N}$ and there are scalars $\lambda_0, \dots, \lambda_n \in \mathbb{F}$ such that

$$w = \sum_{i=0}^{n} \lambda_i T^i v, \text{ so } Tw = \sum_{i=0}^{n} \lambda_i T^{i+1} v \in Z_{T,v}.$$

(2) Suppose W is a T-invariant subspace and $w \in W$. Then $Z_{T,w} \subset W$ as $T^i w \in W$ for all $i \in \mathbb{N}$.

Proposition 7.1.7. *Suppose $v \in V \setminus \{0\}$. There there is an $n \in \mathbb{Z}_{\geq 0}$ such that $\left\{T^i v \middle| i = 0, \dots, n\right\}$ is a basis for $Z_{T,v}$.*

Proof. As V is finite dimensional, the set $\left\{T^i v \middle| i \in \mathbb{Z}_{\geq 0}\right\}$ is linearly dependent, so we can pick a maximal $n \in \mathbb{Z}_{\geq 0}$ such that $B = \left\{T^i v \middle| i = 0, \dots, n\right\}$ is linearly independent. We claim that $T^m v \in \operatorname{span}(B)$ for all $m > n$ so B is the desired basis. We prove this claim by induction on m. The base case is $m = n + 1$.

$\underline{m = n+1}$: We already know $T^{n+1}v \in \operatorname{span}(B)$ as n was chosen to be maximal. Hence there are scalars $\lambda_0, \dots, \lambda_n \in \mathbb{F}$ such that

$$T^{n+1}v = \sum_{i=0}^{n} \lambda_i T^i v.$$

$\underline{m \Rightarrow m+1}$: Suppose now that $T^i v \in \operatorname{span}(B)$ for all $i = 0, \dots, m$. Then

$$T^{m+1}v = T^{m-n}T^{n+1}v = T^{m-n} \sum_{i=0}^{n} \lambda_i T^i v = \sum_{i=0}^{n} \lambda_i T^{m-n+i} v \in \operatorname{span}(B)$$

by the induction hypothesis. □

Definition 7.1.8. Suppose $W \subset V$ is a nontrivial T-cyclic subspace of V. Then a T-cyclic basis of W is a basis of the form $\{w, Tw, \dots, T^n w\}$ for some $n \in \mathbb{Z}_{\geq 0}$. We know such a basis exists by 7.1.7. If $W = Z_{T,v}$, then the T-cyclic basis associated to v is denoted $B_{T,v}$.

Examples 7.1.9.

(1) If $A \in M_n(\mathbb{F})$ is a companion matrix for $p \in \mathbb{F}[z]$, then the standard basis for $\mathbb{F}^n$ is an L_A-cyclic basis. In fact, $\mathbb{F}^n = Z_{L_A, e_1}$, and the standard basis is equal to B_{L_A, e_1}.

(2)

Proposition 7.1.10. *Suppose $T \in L(V)$, and suppose V is T-cyclic.*

(1) Let B be a T-cyclic basis for V. Then $[T]_B$ is the companion matrix for $\min_T$.

(2) $\deg(\min_T) = \dim(V)$.

Proof.

(1) Let $n = \dim(V)$. We know that $[T]_B$ is a companion matrix for some polynomial $p \in \mathbb{F}[z]$ as $B = \{v, Tv, \dots, T^{n-1}v\}$, and

$$[T]_B = \Big[[Tv]_B \Big| [T^2 v]_B \Big| \cdots \Big| [T^n v]_B \Big] = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ & \ddots & \ddots & \vdots & \vdots \\ & & \ddots & 0 & -a_{n-2} \\ 0 & & & 1 & -a_{n-1} \end{pmatrix}$$

where

$$T^n v = -\sum_{i=0}^{n-1} a_i T^i v \text{ and } p(z) = z^n + \sum_{i=1}^{n-1} a_i z^i.$$

Now $\min_{L_{[T]_B}} = \min_T = p$ by 7.1.3.

(2) This follows immediately from (1). □

Lemma 7.1.11. *Suppose p is a monic irreducible factor of $\min_T$ for $T \in L(V)$, and $d = \deg(p)$. Suppose that $v \in V \setminus \{0\}$ and $m \in \mathbb{N}$ is minimal with $p(T)^m v = 0$. Then $B_{T,v} = \left\{T^i v \middle| i = 0, \dots, dm-1\right\}$ (so $Z_{T,v}$ has dimension d).*

Proof. Let $W = Z_{T,v}$, and note that W is a T-invariant subspace. Set $S = T|_W$, and note that $p(T)^m v = 0$ implies that $p(T)^m w = 0$ for all $w \in W$. Hence $p(S)^m = 0$, and $\min_S \mid p^m$ by 5.4.3. But p is irreducible, so $\min_S = p^k$ for some $k \leq m$, but as $p(T)^k v \neq 0$ for all $k < m$, we must have $\min_S = p^m$. As W is S-cyclic, we must have that

$$dm = \deg(p^m) = \deg(\min_S) = \dim(W)$$

by 7.1.10, so $|B_{T,v}| = dm$. The result now follows by 7.1.7. □

Exercises

Exercise 7.1.12. Let $T \in L(V)$ with $\min_T = p^m$ for some monic irreducible $p \in \mathbb{F}[z]$ and some $m \in \mathbb{N}$. Show the following are equivalent:

(1) V is T-cyclic,

(2) $\deg(\min_T) = \dim(V)$, and

(3) $\min_T = \operatorname{char}_T$.

7.2 Rational Canonical Form

The proof of the main theorem in this section is adapted from [2].

Definition 7.2.1. Let $T \in L(V)$.

(1) Subspaces $Z_1, \dots, Z_n$ are called a rational canonical decomposition of V for T if

$$V = \bigoplus_{i=1}^{n} Z_i,$$

Z_i is T-cyclic for all $i \in [n]$, and $Z_i \subset K_{p_i}$ for some monic irreducible $p_i \mid \min_T$ for all $i \in [n]$.

(2) A basis B of V is called a rational canonical basis for T if B is the disjoint union of nonempty sets B_i, denoted

$$B = \coprod_{i=1}^{n} B_i,$$

such that each B_i is a T-cyclic basis for $\operatorname{span}(B_i) \subset K_{p_i}$ for some monic irreducible $p_i \mid \min_T$ for all $i \in [n]$.

(3) The matrix $[T]_B$ is called a rational canonical form of T if B is a rational canonical basis for T.

(4) The matrix $A \in M_n(\mathbb{F})$ is said to be in rational canonical form if the standard basis of $\mathbb{F}^n$ is a rational canonical basis for L_A.

Remark 7.2.2. Note that if B is a rational canonical basis for T, then $[T]_B$ is a block diagonal matrix such that each block is a companion matrix for a polynomial $q \in \mathbb{F}[z]$ of the form p^m where $p \in \mathbb{F}[z]$ is monic and irreducible and $m \in \mathbb{N}$.

Proposition 7.2.3. *Suppose $T \in L(V)$ such that $\min_T = p^m$ for some monic, irreducible $p \in \mathbb{F}[z]$ and some $m \in \mathbb{N}$. Suppose $v_1, \dots, v_n \in V$ such that*

$$S_1 = \coprod_{i=1}^n B_{T,w_i}$$

is linearly independent so that

$$W = \bigoplus_{i=1}^n Z_{T,w_i}$$

is a subspace of V. For each $i \in [n]$, Suppose there is $v_i \in V$ such that $p(T)v_i = w_i$, i.e. $w_i \in \operatorname{im}(p(T))$ for all $i \in [n]$. Then

$$S_2 = \coprod_{i=1}^n B_{T,v_i}$$

is linearly independent.

Proof. Set $Z_i = Z_{T,w_i}$, $T_i = T|_{Z_i}$, and $m_i = |B_{T,v_i}|$ for each $i \in [n]$. Suppose

$$\sum_{i=1}^n \sum_{j=1}^{m_i} \lambda_{i,j} T^j v_i = 0.$$

For $i \in [n]$, let

$$f_i(z) = \sum_{j=1}^{m_i} \lambda_{i,j} z^j \text{ so that } 0 = \sum_{i=1}^n f_i(T) v_i.$$

Applying $p(T)$, we get

$$0 = p(T) \sum_{i=1}^n f_i(T) v_i = \sum_{i=1}^n f_i(T) p(T) v_i = \sum_{i=1}^n f_i(T) w_i.$$

Now as $f_i(T) w_i \in Z_i$ and W is a direct sum of the Z_i's, we must have $f_i(T) w_i = 0$ for all $i \in [n]$. Hence $f_i(T_i) = 0$ on Z_i, and $\min_{T_i} \mid f_i$. Since $\min_T(T_i) = 0$, we must have that $\min_{T_i} \mid \min_T$, so $\min_{T_i} = p^k$ for some $k \leq m$. Hence $p|f_i$ for all $i \in [n]$. Let $g_i \in \mathbb{F}[z]$ such that $pg_i = f_i$ for all $i \in [n]$. Then

$$\sum_{i=1}^n f_i(T) v_i = \sum_{i=1}^n g_i(T) p(T) v_i = \sum_{i=1}^n g_i(T) w_i = 0,$$

so once again, $g_i(T)w_i = 0$ for all $i \in [n]$ as W is the direct sum of the Z_i's and $g_i(T)w_i \in Z_i$ for all $i \in [n]$. But then

$$0 = g_i(T)w_i = f_i(T)v_i = \sum_{j=1}^{m_i} \lambda_{i,j} T^j v_i,$$

so we must have that $\lambda^i_j = 0$ for all $i \in [n]$ and $j \in [m_i]$ as B_{T,v_i} is linearly independent. □

Proposition 7.2.4. *Suppose $T \in L(V)$ with $\min_T = p^m$ for some monic, irreducible $p \in \mathbb{F}[z]$ and some $m \in \mathbb{N}$. Suppose that W is a T-invariant subspace of V with basis B.*

(1) Suppose $v \in \ker(p(T)) \setminus W$. Then $B \cup B_{T,v}$ is linearly independent.

(2) There are $v_1, \dots, v_n \in \ker(p(T))$ such that

$$B' = B \cup \bigcup_{i=1}^{n} B_{T,v_i}$$

is linearly independent and contains $\ker(p(T))$.

Proof.

(1) Let $d = \deg(\min_T)$, and note that $B_{T,v} = \left\{T^i v \middle| i = 0, \dots, d-1\right\}$ by 7.1.11. Suppose $B = \{v_1, \dots, v_n\}$ and

$$\sum_{i=1}^{n} \lambda_i v_i + w = 0 \text{ where } w = \sum_{j=0}^{d-1} \mu_j T^j v.$$

Then $w \in \operatorname{span}(B) = W$ and $w \in Z_{T,v}$, both of which are T-invariant subspaces. Thus we have $Z_{T,w} \subset W$ and $Z_{T,w} \subset Z_{T,v}$, so we have

$$Z_{T,w} \subset Z_{T,v} \cap W \subset Z_{T,v}.$$

If $w \neq 0$, then $p(T)w = 0$ as $p(T)|_{Z_{T,v}} = 0$, so by 7.1.11,

$$d = \dim(Z_{T,w}) \leq \dim(Z_{T,v} \cap W) \leq \dim(Z_{T,v}) = d,$$

so equality holds. But then $Z_{T,v} = Z_{T,v} \cap W$, which is a subset of W, a contradiction as $v \notin W$. Thus $w = 0$, and $\mu_j = 0$ for all $j = 0, \dots, d-1$. But as B is a basis, we have $\lambda_i = 0$ for all $i \in [n]$.

(2) If $\ker(p(T)) \not\subseteq W$, then pick $v_1 \in W \setminus \ker(p(T))$. By (1), $B \cup B_{T,v_1}$ is linearly independent, and $W_1 = \operatorname{span}(B \cup B_{T,v_t})$ is T-invariant. If $\ker(p(T)) \not\subseteq W_1$, pick $v_2 \in W_1 \setminus \ker(p(T))$. By (1), $B \cup B_{T,v_1} \cup B_{T,v_2}$ is linearly independent, and $W_2 = \operatorname{span}(B \cup B_{T,v_1} \cup B_{T,v_2})$ is T-invariant. We can repeat this process until W_k contains $\ker(p(T))$ for some $k \in \mathbb{N}$. □

Lemma 7.2.5. *Suppose $T \in L(V)$ such that $\min_T = p^m$ where $p \in \mathbb{F}[z]$ is monic and irreducible and $m \in \mathbb{N}$. Then there is a rational canonical basis B for T.*

Proof. We proceed by induction on $m \in \mathbb{N}$.

$\underline{m=1}$: Then $V = \ker(p(T))$, so we may apply 7.2.4 with $W = (0)$ to get a basis for $\ker(p(T))$.

$\underline{m-1 \Rightarrow m}$: We know $W = \operatorname{im}(p(T))$ is T-invariant, and $T|_W$ has minimal polynomial p^{m-1}. By the induction hypothesis, we have a rational canonical basis

$$B = \coprod_{i=1}^{n} B_{T,w_i}$$

for $T|_W$. As $W = \operatorname{im}(p(T))$, there are $v_i \in V$ for $i \in [n]$ such that $p(T)v_i = w_i$, and

$$C = \coprod_{i=1}^{n} B_{T,v_i}$$

is linearly independent by 7.2.3. By 7.2.4, there are $u_1, \ldots, u_k \in \ker(p(T))$ such that

$$D = C \cup \coprod_{i=1}^{k} B_{T,u_i}$$

is linearly independent and contains $\ker(p(T))$. Let $U = \operatorname{span}(D)$. We claim that $U = V$. First, note that U is T-invariant, so it is $p(T)$ invariant. Let $S = p(T)|_U$. Then

$$\operatorname{im}(S) = SU = S\operatorname{span}(C) \supseteq W = \operatorname{im}(p(T)), \text{ and } \ker(S) \supseteq \ker(p(T))$$

as U contains $\ker(p(T))$. By 3.2.13

$$\dim(U) = \dim(\operatorname{im}(S)) + \dim(\ker(S)) \geq \dim(\operatorname{im}(p(T))) + \dim(\ker(p(T))) = \dim(V),$$

so $U = V$. Thus D is a rational canonical basis for T. □

Theorem 7.2.6. *Every $T \in L(V)$ has a rational canonical decomposition.*

Proof. By 4.3.11, we can factor $\min_T$ uniquely:

$$\min_T = \prod_{i=1}^{n} p_i^{m_i}$$

where $p_i \in \mathbb{F}[z]$ are distinct monic irreducible factors of $\min_T$ and $m_i \in \mathbb{N}$ for all $i \in [n]$. By 6.6.1, we have

$$V = \bigoplus_{i=1}^{n} K_{p_i}$$

where $K_{p_i} = \ker(p_i(T)^{m_i})$ are T-invariant subspaces, and if $T_i = T|_{K_{p_i}}$, then $\min_{T_i} = p_i^{m_i}$ for all $i \in [n]$. By 7.2.5, we know that there is a rational canonical basis B_i for T_i on K_{p_i}, so

$$B = \coprod_{i=1}^{n} B_i$$

is the desired rational canonical basis of T. □

Corollary 7.2.7. *For $T \in L(V)$, there is a rational canonical decomposition $V = \bigoplus_{i=1}^{n} Z_{T,v_i}$ into cyclic subspaces.*

Proof. By 7.2.6 there is a rational canonical basis

$$B = \coprod_{i=1}^{n} B_{T,v_i}$$

for $v_1, \ldots, v_n \in V$. Setting $Z_{T,v_i} = \operatorname{span}(B_{T,v_i})$, we get the desired result. □

Exercises

Exercise 7.2.8. Classify all pairs of polynomials $(m, p) \in \mathbb{F}[z]^2$ such that there is an operator $T \in L(V)$ where V is a finite dimensional vector space over $\mathbb{F}$ such that $\min_T = m$ and $\operatorname{char}_T = p$.

Exercise 7.2.9. Show that two matrices $A, B \in M_n(\mathbb{F})$ are similar if and only if they share a rational canonical form, i.e. there are bases C, D of $\mathbb{F}^n$ such that $[L_A]_C = [L_B]_D$ is in rational canonical form.

Exercise 7.2.10. Prove or disprove: A square matrix $A \in M_n(\mathbb{F})$ is similar to its transpose A^T. If the statement is false, find a condition which makes it true.

7.3 The Cayley-Hamilton Theorem

Theorem 7.3.1 (Cayley-Hamilton)**.** *For $T \in L(V)$, $\operatorname{char}_T(T) = 0$.*

Proof. Factor $\min_T$ into irreducibles as in 4.3.11:

$$\min_T = \prod_{i=1}^{n} p_i^{m_i}.$$

Let B be a rational canonical basis for T as in 7.2.6. By 7.1.10, $[T]_B$ is block diagonal, so let $A_1, \ldots, A_k$ be the blocks. We know that A_j is the companion matrix for $p_{i_j}^{n_j^i}$ where $i_j \in [n]$ and $n_j \in \mathbb{N}$ for $j \in [k]$. As

$$0 = [\min_T(T)]_B = \min_T([T]_B) = \min_T \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_k \end{pmatrix} = \begin{pmatrix} \min_T(A_1) & & \\ & \ddots & \\ & & \min_T(A_k) \end{pmatrix},$$

we must have that $n_j^i \leq m_i$ for all $i \in [n]$. But as $\min_T$ is the minimal polynomial of $[T]_B$, we must have that $n_j^i = m_i$ for some $i \in [n]$. By 7.1.3, we see that $\operatorname{char}_T(z) = \det(zI - [T]_B)$ is a product

$$\operatorname{char}_T(z) = \det(zI - [T]_B) = \prod_{i=1}^{n} p_i(z)^{r_i}$$

where $r_i \geq m_i$ for all $i \in [n]$. Thus,

$$\operatorname{char}_T(T) = \prod_{i=1}^{n} p_i(z)^{r_i}(T) = \min_T(T) \prod_{i=1}^{n} p_i(T)^{r_i - m_i} = 0.$$

□

Corollary 7.3.2. $\min_T \mid \operatorname{char}_T$ *for all* $T \in L(V)$.

Proposition 7.3.3. *Let* $A \in M_n(\mathbb{F})$.

(1) The coefficient of the z^{n-1} *term in* $\operatorname{char}_A(z)$ *is equal to* $-\operatorname{trace}(A)$.

(2) The constant term in $\operatorname{char}_A(z)$ *is equal to* $(-1)^n \det(A)$.

Proof. First note that the result is trivial if A is the companion matrix to a polynomial $p(z) = z^n + \sum_{i=0}^{n-1} a_i z^i \in \mathbb{F}[z]$ as $\operatorname{char}_A(z) = p(z)$, $\operatorname{trace}(A) = -a_{n-1}$, and $\det(A) = (-1)^n a_0$. For the general case, let B be a rational canonical basis for L_A so that $[A]_B$ is a block diagonal matrix

$$[A]_B = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}$$

where each block A_i is the companion matrix of some polynomial $p_i(z) = z^{n_i} + \sum_{j=0}^{n_i - 1} a_j^i z^j \in \mathbb{F}[z]$. As $A \sim [A]_B$, we have that

$$\operatorname{char}_A(z) = \det(zI - [A]_B) = \prod_{i=1}^{n} p_i(z),$$

$\operatorname{trace}(A) = \operatorname{trace}([A]_B)$, and $\det(A) = \det([A]_B)$.

(1) It is easy to see that the trace of $[A]_B$ is the sum of the traces of the A_i, i.e.

$$\operatorname{trace}(A) = \operatorname{trace}([A]_B) = \sum_{i=1}^{m} \operatorname{trace}(A_i) = \sum_{i=1}^{m} -a_{n_i - 1}^i.$$

Now one sees that the $(n-1)^{\text{th}}$ coefficient of char_A is exactly the negative of the right hand side as

$$\operatorname{char}_A(z) = \prod_{i=1}^{n} \left(z^{n_i} + \sum_{j=0}^{n_i} a_j^i z^j \right) = z^n + \left(\sum_{i=1}^{m} a_{n_i - 1}^i \right) z^{n-1} + \cdots + \sum_{i=1}^{m} a_0^i$$

since the $(n-1)^{\text{th}}$ terms in the product above are obtained by taking the $(n_i - 1)^{\text{th}}$ term of p_i and multiplying by the leading term (the z^{n_j} term) for $j \neq i$. Similarly, the constant term of char_A is obtained by taking the product of the constant terms of the p_i's.

(2) In (1), we calculated that the constant term of char_A is $\sum_{i=1}^{m} a_0^i$. But this is $(-1)^n \det(A)$ as

$$\det(A) = \det([A]_B) = \prod_{i=1}^{m} \det(A_i) = \prod_{i=1}^{m} (-1)^{n_i} a_0^i = (-1)^n \prod_{i=1}^{m} a_0^i.$$

□

Exercises

7.4 Jordan Canonical Form

Definition 7.4.1. A Jordan block of size n associated to $\lambda \in \mathbb{F}$ is a matrix $J \in M_n(\mathbb{F})$ such that

$$J_{i,j} = \begin{cases} \lambda & \text{if } i = j \\ 1 & \text{if } j = i + 1 \\ 0 & \text{else,} \end{cases}$$

i.e. J looks like

$$\begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix}.$$

Proposition 7.4.2. *Suppose $J \in M_n(\mathbb{F})$ is a Jordan block associated to λ. Then* $\operatorname{char}_J(z) = \min_T(z) = (z - \lambda)^n$.

Proof. Obvious. □

Remark 7.4.3. Let $T \in L(V)$ and $\lambda \in \operatorname{sp}(T)$. If $m \in \mathbb{N}$ is maximal such that $(z - \lambda)^m \mid \min_T$ and $S = T|_{G_\lambda}$, then $(z - \lambda)^m = \min_S$ by 6.6.1. This means that the operator $N = (T - \lambda I)|_{G_\lambda}$ is nilpotent of order m. Now by 7.2.6, there is a rational canonical decomposition for N

$$G_\lambda = \bigoplus_{i=1}^{k} Z_{N,v_i}$$

for some $k \in \mathbb{N}$ where $v_i \in G_\lambda$ for all $i \in [n]$. Set $n_i = |B_{N,v_i}|$ for each $i \in [n]$. Then

$$B_{N,v_i} = \left\{ (T - \lambda I)^j v_i \middle| j = 0, \dots, n_i - 1 \right\},$$

and note that $[N|_{B_{N,v_i}}]_{B_{N,v_i}}$ is a matrix in $M_{n_i}(\mathbb{F})$ of the form

$$\begin{pmatrix} 0 & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{pmatrix}$$

as the minimal polynomial of $N|_{B_{N,v_i}}$ is z^{n_i}. Setting $C_{N,v_i} = \big\{(T-\lambda I)^j v_i \big| j = n_i - 1, \dots, 0\big\}$, i.e. reordering B_{N,v_i}, we have that $[N|_{C_{N,v_i}}]_{C_{N,v_i}}$ is a matrix in $M_{n_i}(\mathbb{F})$ of the form

$$\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix},$$

i.e. it is a Jordan block associated to 0 of size n_i. Thus, we see that if we set

$$B = \coprod_{i=1}^{k} B_{N,v_i},$$

we have that $[N]_B$ is a block diagonal matrix where the i^{th} block is the companion matrix of the polynomial z^{n_i}, and if we set

$$C = \coprod_{i=1}^{k} C_{N,v_i},$$

we have that $[N]_C$ is a block diagonal matrix where the i^{th} block is a Jordan block of size n_i associated to 0. Thus $[T|_{G_\lambda}]_C = [N]_C + \lambda I$ is a block diagonal matrix where the i^{th} block is a Jordan block of size n_i associated to λ.

Theorem 7.4.4. *Suppose $T \in L(V)$ and $\min_T$ splits in $\mathbb{F}[z]$. Then there is a basis B of V such that $[T]_B$ is a block diagonal matrix where each block is a Jordan block. The diagonal elements of $[T]_B$ are precisely the eigenvalues of T, and if $\lambda \in \operatorname{sp}(T)$, λ appears exactly M_λ times on the diagonal of $[T]_B$.*

Proof. By 6.6.2, we have $\min_T$ splits if and only if

$$V = \bigoplus_{\lambda \in \operatorname{sp}(T)} G_\lambda.$$

by 7.4.3, for each $\lambda \in \operatorname{sp}(T)$, there is a basis B_λ of G_λ such that $T|_{B_\lambda}$ is a block diagonal matrix in $M_{M_\lambda}(\mathbb{F})$ where each block is a Jordan block associated to λ. Setting

$$B = \coprod_{\lambda \in \operatorname{sp}(T)} B_\lambda$$

gives the desired result. □

Definition 7.4.5.

(1) Let $T \in L(V)$. A basis B for V as in 7.4.4 is called a Jordan canonical basis for T, and the matrix $[T]_B$ is called a Jordan canonical form of T.

(2) A matrix $A \in M_n(\mathbb{F})$ is said to be in Jordan canonical form if the standard basis is a Jordan canonical basis of $\mathbb{F}^n$ for L_A.

Examples 7.4.6.

(1)

(2)

Definition 7.4.7. Let $T \in L(V)$. Recall that if $v \in G_\lambda \setminus \{0\}$ for some $\lambda \in \operatorname{sp}(T)$, then there is a minimal n such that $(T-\lambda I)^n v = 0$. This means that $B_{T-\lambda I,v} = \big\{(T-\lambda I)^i v \big| i = 0, \dots, n-1\big\}$ is a basis for $Z_{T-\lambda I,v}$.

(1) The set $C_{T-\lambda I,v} = ((T-\lambda I)^{n-1}v, \dots, (T-\lambda I)v, v)$ is called a Jordan chain for T associated to the eigenvalue λ of length n. The vector $(T-\lambda I)^{n-1}v$ is called the lead vector of the Jordan chain $C_{T-\lambda I,v}$, and note that it is an eigenvector corresponding to the eigenvalue λ.

Examples 7.4.8.

(1) Suppose $J \in M_n(\mathbb{F})$ is a Jordan block associated to $\lambda \in \mathbb{F}$

$$J = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix}.$$

One can easily see that $(J-\lambda I)e_i = e_{i-1}$ for all $1 < i \leq n$ and $(J-\lambda I)e_1 = 0$, so the standard basis of $\mathbb{F}^n$ is a Jordan chain for L_J associated to λ with lead vector e_1.

(2) If A is a block diagonal matrix

$$A = \begin{pmatrix} J_1 & & 0 \\ & \dots & \\ 0 & & J_n \end{pmatrix}$$

where J_i is a Jordan block associated to $\lambda \in \mathbb{F}$ for all $i \in [n]$, then we see by (1) that the standard basis is a disjoint union of Jordan chains associated to λ.

Corollary 7.4.9. *Suppose B is a Jordan canonical basis for $T \in L(V)$. Then B is a disjoint union of Jordan chains associated to the eigenvalues of λ.*

Proof. Let $\operatorname{sp}(T) = \{\lambda_1, \dots, \lambda_n\}$. We have that $[T]_B$ is a block diagonal matrix

$$[T]_B = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_n \end{pmatrix}$$

where each A_i is a block diagonal matrix corresponding to $\lambda_i \in \operatorname{sp}(T)$ composed of Jordan blocks:

$$A_i = \begin{pmatrix} J_1^i & & 0 \\ & \ddots & \\ 0 & & J_{n_i}^i \end{pmatrix}.$$

Now set $B_i = B \cap G_{\lambda_i}$ for each $i \in [n]$ so that

$$B = \coprod_{i=1}^{n} B_i,$$

and set $T_i = T|_{G_{\lambda_i}}$. Then $[T_i]_{B_i} = A_i$, and it is easy to check that B_i is the disjoint union of Jordan chains associated to λ_i for all $i \in [n]$. Hence B is a disjoint union of Jordan chains associated to the eigenvalues of T. □

Exercises

Exercise 7.4.10. Classify all pairs of polynomials $(m, p) \in \mathbb{C}[z]^2$ such that there is an operator $T \in L(V)$ where V is a finite dimensional vector space over $\mathbb{C}$ such that $\min_T = m$ and $\operatorname{char}_T = p$.

Exercise 7.4.11. Show that two matrices $A, B \in M_n(\mathbb{C})$ are similar if and only if they share a Jordan canonical form.

7.5 Uniqueness of Canonical Forms

In this section, we discuss to what extent a rational canonical form of $T \in L(V)$ is unique, and to what extend the rational canonical form is unique if $\min_T$ splits. The information from this section is adapted from sections 7.2 and 7.4 of [2].

Notation 7.5.1 (Jordan Canonical Dot Diagrams)**.**

Ordering Jordan Blocks: Suppose $T \in L(V)$ such that $\min_T$ splits, and let B is a Jordan canonical basis for T. If $\lambda \in \operatorname{sp}(T)$ and $m \in \mathbb{N}$ is the largest integer such that $(z-\lambda)^m \mid \min_T$, then we have that $S = T|_{G_\lambda} \in L(G_\lambda)$ has minimal polynomial z^m, and there is a subset $C \subset B$ that is a Jordan canonical basis for S. This means that $[S]_C$ is a block diagonal matrix, so let $A_1, \ldots, A_n$ be these blocks, i.e

$$[S]_C = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_n \end{pmatrix}.$$

We know that A_i is a Jordan block of size m_i where $m_i \leq m$ for all $i \in [n]$. We now impose the condition that in order for $[S]_C$ to be in Jordan canonical form, we must have that

$m_i \geq m_{i+1}$ for all $i \in [n-1]$. This means that to find a Jordan canonical basis for $T \in L(V)$, we first decompose G_λ into T-cyclic subspaces for all $\lambda \in \text{sp}(T)$, then we take bases of these T-cyclic subspaces, and then we order these bases by how many elements they have.

Nilpotent Operators: Suppose $\min_T = z^m$ for some $\lambda \in \mathbb{F}$. Let

$$B = \coprod_{i=1}^{n} C_{T,v_i}$$

be a Jordan canonical basis of V consisting of disjoint Jordan chains C_{T,v_i} for $i \in [n]$. We can picture B as an array of dots called a Jordan canonical dot diagram for T relative to the basis B. For $i \in [n]$, let $m_i = |C_{T,v_i}|$, and note that the ordering of Jordan blocks requires that $m_i \geq m_{i+1}$ for all $i \in [n-1]$. Write B as an array of dots such that

(1) the array contains n columns starting at the top,

(2) the i^{th} column contains m_i dots, and

(3) the $(i,j)^{\text{th}}$ dot is labelled $T^{m_i - j} v_i$.

$$\begin{array}{llcllcll}
\bullet & T^{m_1-1}v_1 & \cdots & \bullet & T^{m_2-1}v_2 & \cdots & \bullet & T^{m_n-1}v_n \\
\bullet & T^{m_1-2}v_1 & \cdots & \bullet & T^{m_2-2}v_2 & \cdots & \bullet & T^{m_n-2}v_n \\
\vdots & & & \vdots & & & \vdots & \\
\vdots & & & \vdots & & & \bullet & Tv_n \\
\vdots & & & \vdots & & & \bullet & v_n \\
\vdots & & & \bullet & Tv_2 & & & \\
\vdots & & & \bullet & v_2 & & & \\
\bullet & Tv_1 & & & & & & \\
\bullet & v_1 & & & & & &
\end{array}$$

Operators: Suppose that $T \in L(V)$, and let B be a Jordan canonical basis for T. Let $B_\lambda = B \cap G_\lambda$ for $\lambda \in \text{sp}(T)$ and $T_\lambda = T|_{G_\lambda}$. Then $T_\lambda - \lambda I$ is nilpotent, so we may apply (2) to get a dot diagram for $T_\lambda - \lambda I$ relative to B_λ for each $\lambda \in \text{sp}(T)$. The dot diagram for T relative to B is the disjoint union of the dot diagrams for T_λ relative to the B_λ for $\lambda \in \text{sp}(T)$.

Lemma 7.5.2. *Suppose $T \in L(V)$ is nilpotent of order m, i.e. $\min_T(z) = z^m$. Let B be a Jordan canonical basis for T, and let r_i denote the number of dots in the i^{th} row of the dot diagram for T relative to B as in 7.5.1 for $i \in [n]$. Then*

(1) If $R_k = \displaystyle\sum_{j=1}^{k} r_j$ for $k \in [n]$, then $R_k = \text{nullity}(T^k) = \dim(\ker(T^k))$ for all $k \in [n]$,

(2) $r_1 = \dim(V) - \text{rank}(T)$, and

(3) $r_k = \text{rank}(T^{k-1}) - \text{rank}(T^k)$ for all $k > 1$.

Hence the dot diagram for T is independent of the choice of Jordan canonical basis B for T as each r_i is completely determined by T, and the Jordan canonical form $[T]_B$ is unique.

Proof.

(1) We see that there are at least R_k linearly independent vectors in $\ker(T^k)$, namely $T^{m_i-j}v_i$ for all $i \in [n]$ and $j \in [k]$, so $\operatorname{nullity}(T^k) \geq R_k$. Moreover, $\{T^k(T^{m_i-j}v_i) \big| i \in [n] \text{ and } j > k\}$ is a linearly independent subset of B, so $\operatorname{rank}(T^k) \geq |B| - R_k$. By the rank-nullity theorem, we have
$$|B| = \operatorname{rank}(T^k) + \operatorname{nullity}(T^k) \geq |B| - R_k + R_k = |B|,$$
so equality holds, and we must have that $\operatorname{nullity}(T^k) = R_k$.

(2) This is immediate from (1) and the rank-nullity theorem.

(3) For $k > 1$, by (1) we have that
$$r_k = \sum_{i=1}^{k} r_i - \sum_{i=1}^{k-1} r_i = (\dim(V) - \operatorname{nullity}(T^k)) - (\dim(V) - \operatorname{nullity}(T^{k-1})) = \operatorname{rank}(T^k) - \operatorname{rank}(T^{k-1}).$$

□

Theorem 7.5.3. *Suppose $T \in L(V)$ and $\min_T$ splits in $\mathbb{F}[z]$. Two Jordan canonical forms (using the convention of 7.5.1) of $T \in L(V)$ differ only by a permutation of the eigenvalues of T.*

Proof. If B is a Jordan canonical basis of T as in 7.5.1, we set $B_\lambda = B \cap G_\lambda$ and $T_\lambda = T|_{G_\lambda}$ for all $\lambda \in \operatorname{sp}(T)$, and we note that
$$B = \coprod_{\lambda \in \operatorname{sp}(T)} B_\lambda.$$

Note further that $T_\lambda - \lambda I$ is nilpotent, and $[T_\lambda - \lambda I]_{B_\lambda}$ is in Jordan canonical form, which is unique by 7.5.2. Hence $[T_\lambda]_{B_\lambda} = [T_\lambda - \lambda I]_{B_\lambda} + \lambda I$ is unique, so is unique up to the ordering of the $\lambda \in \operatorname{sp}(T)$. In fact, if $\{\lambda_1, \ldots, \lambda_n\}$ is an enumeration of $\operatorname{sp}(T)$ and
$$B = \coprod_{i=1}^{n} B_{\lambda_i},$$

then setting $T_i = T_{\lambda_i}$ and $B_i = B_{\lambda_i}$, we have
$$[T]_B = \begin{pmatrix} [T_1]_{B_1} & & 0 \\ & \ddots & \\ 0 & & [T_n]_{B_n} \end{pmatrix}.$$

□

Notation 7.5.4 (Rational Canonical Dot Diagrams)**.**

Ordering Companion Blocks: Let B is a rational canonical basis for $T \in L(V)$. If $p \in \mathbb{F}[z]$ is a monic irreducible factor of $\min_T$ and $m \in \mathbb{N}$ is the largest integer such that $p^m \mid \min_T$, then we have that $S = T|_{K_p} \in L(K_p)$ has minimal polynomial p^m, and there is a subset $C \subset B$ that is a rational canonical basis for S. This means that $[S]_C$ is a block diagonal matrix, so let $A_1, \ldots, A_n$ be these blocks, i.e

$$[S]_C = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_n \end{pmatrix}.$$

We know that A_i is the companion matrix for p^{m_i} where $m_i \leq m$ for all $i \in [n]$. We now impose the condition that in order for $[S]_C$ to be in rational canonical form, we must have that $m_i \geq m_{i+1}$ for all $i \in [n-1]$. This means that to find a rational canonical basis for $T \in L(V)$, we first decompose K_p into T-cyclic subspaces for all monic irreducible $p \in \mathbb{F}[z]$ dividing $\min_T$, then we take bases of these T-cyclic subspaces, and then we order these bases by how many elements they have.

Case 1: Suppose $\min_T = p^m$ for some monic, irreducible $p \in \mathbb{F}[z]$ and some $m \in \mathbb{N}$. Let

$$B = \coprod_{i=1}^{n} B_{T,v_i}$$

be a rational canonical basis of V for T. B can be represented as an array of dots called a rational canonical dot diagram for T relative to B. For $i \in [n]$, let $B_i = B_{T,v_i}$, let $Z_i = Z_{T,v_i}$, let $T_i = T|_{Z_i}$, and let $m_i \in \mathbb{N}$ be the minimal number such that $p(T)^{m_i} v_i = 0$, i.e. $\min_{T_i} = p^{m_i}$ and $m_i \leq m$ for all $i \in [n]$. Set $d = \deg(p)$, and note that $|B_i| = dm_i$ by 7.1.11. Now by the ordering of the companion blocks, we must have that $m_i \geq m_{i+1}$ for all $i \in [n-1]$. The dot diagram for T relative to B is given by an array of dots such that

1. the array contains n columns starting at the top,
2. the i^{th} column has m_i dots, and
3. the $(i,j)^{\text{th}}$ dot is labelled $p(T)^{m_i - j} v_i$.

Note that there are exactly $|B|/d$ dots in the dot diagram.

Operators: As before, for a general $T \in L(V)$, we let B be a Rational canonical basis for T, and we let $B_p = B \cap G_p$ and $T_p = T|_{G_p}$ for each distinct monic irreducible $p \mid \min_T$. The dot diagram for T is then the disjoint union of the dot diagrams for the T_p's relative to the B_p's.

Lemma 7.5.5. *Suppose $T \in L(V)$ with $\min_T = p^m$ for a monic irreducible $p \in \mathbb{F}[z]$ and some $m \in \mathbb{N}$. Let $d = \deg(p)$, let*

$$B = \coprod_{i=1}^{n} B_{T,v_i}$$

be a rational canonical basis of V for T, and let $m_i \in \mathbb{N}$ be minimal such that $p(T)^{m_i} v_i = 0$. Let r_j be the number of dots in the j^{th} row of the dot diagram for T relative to B for $j \in [l]$. Then

(1) Set $C_l^i = \left\{ (p(T)^{m_i-h} T^l v_i \middle| h \in [m_i] \right\}$ is a $p(T)$-cyclic basis for $Z_{p(T),T^l v_i}$ for $0 \leq l < d$ and $i \in [n]$. Then $C_i = \displaystyle\coprod_{l=0}^{d-1} C_l^i$ is a basis for $Z_i = Z_{T,v_i}$, so it is a Jordan canonical basis of Z_i for $p(T)|_{Z_i}$. Hence $C = \displaystyle\coprod_{i=1}^{n} C_i$ is a Jordan canonical basis of V for $p(T)$.

(2) $r_1 = \dfrac{1}{d}(\dim(V) - \operatorname{rank}(p(T)))$, and

(3) $r_j = \dfrac{1}{d}(\operatorname{rank}(p(T)^{j-1}) - \operatorname{rank}(p(T)^j))$ for $j > 1$.

Hence the dot diagram for T is independent of the choice of rational canonical basis B for T as each r_j is completely determined by $p(T)$, and the rational canonical form $[T]_B$ is unique.

Proof. We have that (2) and (3) are immediate from (1) and 7.5.2 as the Jordan canonical form of $p(T)$, a nilpotent operator of order m, is unique. It suffices to prove (1).

We may suppose $p(z) \neq z$ as this case would be similar to 7.5.3 as in this case, $\min_T$ would split. The fact that C_l^i is linearly independent for all i, l comes from 7.1.11 as $T^l v_i \neq 0$ and m_i is minimal such that $p(T)^{m_i} T^l v_i = 0$ (in fact the restriction of T^l to Z_i is invertible as if $q(z) = z^l$, then $q(T)$ is bijective on $G_p = V$ by the proof of 6.5.5 as p, q are relatively prime).

We show C_i is a basis for Z_i for a fixed $i \in [n]$. Suppose

$$\sum_{l=0}^{d-1} \sum_{h=1}^{m_i} \lambda_{h,l} p(T)^{m_i-h} T^l v_i = 0,$$

and set

$$q_h(z) = \sum_{l=0}^{d-1} \lambda_{h,l} z^l \text{ for } h \in [m_i].$$

Then we see that $\deg(q_h) < d$ for all $h \in [m_i]$, and

$$\sum_{h=1}^{m_i} p(T)^{m_i-h} q_h(T) v_i = 0.$$

Now setting

$$q(z) = \sum_{h=1}^{m_i} p(z)^{m_i-h} q_h(z),$$

we have that $\deg(q) < \deg(p^{m_i})$ and $q(T)v_i = 0$. If $p \nmid q$, then p, q are relatively prime, so $q(T)$ is invertible on $G_p = V$ by the proof of 6.5.5, which is a contradiction as $q(T)v_i = 0$. Hence $p \mid q$. Let $s \in \mathbb{N}$ be maximal such that $p^s \mid q$, and let $f \in \mathbb{F}[z]$ such that $fp^s = q$. If $f \neq 0$, then p, f are relatively prime, so $f(T)$ is invertible on $G_p = V$. But then

$$0 = f(T)^{-1}0 = f(T)^{-1}f(T)p(T)^s v_i = p(T)^s v_i,$$

so we must have that $s \geq m_i$, a contradiction as $\deg(q) < \deg(p^{m_i})$. Hence $f = 0$, so $q = 0$. As $\deg(q_h) < d = \deg(p)$ for all h, we must have that

$$0 = \operatorname{LC}(q) = \operatorname{LC}(p^{m_i-1}q_1) = \operatorname{LC}(p^{m_i-1})\operatorname{LC}(q_1) \Longrightarrow q_1 = 0.$$

Once more, as $\deg(q_n) < d$ for all h, we now have that

$$0 = \operatorname{LC}(q) = \operatorname{LC}(p^{m_i-2}q_2) = \operatorname{LC}(p^{m_i-2})\operatorname{LC}(q_2) \Longrightarrow q_2 = 0.$$

We repeat this process to see that $q_h = 0$ for all $h \in [m_i]$. Hence $\lambda_{h,l} = 0$ for all h, l, and C_i is linearly independent. Now we see that $|C_i| = dm_i = \dim(Z_i) = dm_i$ by 7.1.11, and we are finished by the extension theorem.

It follows immediately that C is a basis for V as $V = \bigoplus\limits_{i=1}^{n} Z_i$. □

Theorem 7.5.6. *Suppose $T \in L(V)$. Two rational canonical forms of $T \in L(V)$ differ only by a permutation of the irreducible monic factor powers of* $\min_T$.

Proof. This follows immediately from 7.5.5 and 6.6.1. □

Exercises

V will denote a finite dimensional vector space over $\mathbb{F}$.

Exercise 7.5.7. Recall that spectrum is an invariant of similarity class by 5.1.11. In this sense, we may define the spectrum of a similarity class of $L(V)$ to be the spectrum of one of its elements. Given distinct $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{C}$, how many similarity classes of matrices in $M_7(\mathbb{C})$ have spectrum $\{\lambda_1, \lambda_2, \lambda_3\}$?

7.6 Holomorphic Functional Calculus

For this section, V will define a finite dimensional vector space over $\mathbb{C}$.

Definition 7.6.1.

(1) The open ball of radius $r > 0$ centered at $z_0 \in \mathbb{C}$ is

$$B_r(z_0) = \left\{z \in \mathbb{C} \middle| |z - z_0| < r\right\}.$$

(2) A subset $U \subset \mathbb{C}$ is called open if for each $z \in U$, there is an $\varepsilon > 0$ such that $B_\varepsilon(z) \subset U$.

Examples 7.6.2.

(1)

(2)

Definition 7.6.3. Suppose $U \subset \mathbb{C}$ is open. A function $f\colon U \to \mathbb{C}$ is called holomorphic (on U) if for each $z_0 \in U$, there is a power series representation of f on $B_\varepsilon(z_0) \subset U$ for some $\varepsilon > 0$, i.e., for each $z_0 \in U$, there is an $\varepsilon > 0$ and a sequence $(\lambda_k)_{k \in \mathbb{Z}_{geq0}}$ such that

$$f(z) = \sum_{k=0}^{\infty} \lambda_k (z - z_0)^k$$

converges for all $z \in B_\varepsilon(z_0)$. Note that if f is holomorphic on U, then f is infinitely many times differentiable at each point in U. We have

$$f'(z) = \sum_{k=1}^{\infty} \lambda_k k (z - z_0)^{k-1}, \;\; f''(z) = \sum_{k=2}^{\infty} \lambda_k k(k-1)(z - z_0)^{k-2}, \;\text{ etc.}$$

In particular,

$$f^{(n)}(z_0) = \lambda_n n \Longrightarrow \lambda_n = \frac{f^{(n)}(z_0)}{n!}.$$

This is Taylor's Formula.

Examples 7.6.4.

(1)

(2)

Definition 7.6.5 (Holomorphic Functional Calculus).

Jordan Blocks: Suppose $A \in M_n(\mathbb{C})$ is a Jordan block, i.e. there is a $\lambda \in \mathbb{C}$ such that

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

We see that we can write A canonically as $A = D + N$ with $D \in M_n(\mathbb{C})$ diagonal ($D = \lambda I$) and $N \in M_n(\mathbb{C})$ nilpotent of order n:

$$A = \underbrace{\begin{pmatrix} \lambda & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & \lambda \end{pmatrix}}_{D} + \underbrace{\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}}_{N}.$$

Now $\operatorname{sp}(T) = \{\lambda\}$, so if $\varepsilon > 0$ and $f\colon B_\varepsilon(\lambda) \to \mathbb{C}$ is holomorphic, then there is a sequence $(\mu_n)_{n\in\mathbb{Z}_{geq0}}$ such that

$$f(z) = \sum_{k=0}^{\infty} \mu_k (z-\lambda)^k$$

converges for all $z \in B_\varepsilon(\lambda)$, so we may define

$$f(A) = f(D+N) = \sum_{k=0}^{\infty} \mu_k (D+N-\lambda I)^k = \sum_{k=0}^{n-1} \mu_k N^k = \begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_{n-1} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \mu_1 \\ 0 & & & \mu_0 \end{pmatrix}$$

$$= \begin{pmatrix} f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & f'(\lambda) \\ 0 & & & f(\lambda) \end{pmatrix}$$

Matrices: Suppose $A \in M_n(\mathbb{C})$. Then there is an invertible $S \in M_n(\mathbb{C})$ and a matrix $J \in M_n(\mathbb{C})$ in Jordan canonical form such that $J = S^{-1}AS$. Let $K_1, \dots, K_n$ be the Jordan blocks of J, and let $\lambda_i \in \operatorname{sp}(A)$ be the diagonal entry of K_i for $i \in [n]$. Let $U \subset \mathbb{C}$ be open such that $\operatorname{sp}(A) \subset U$, and suppose $f\colon U \to \mathbb{C}$ is holomorphic. By the above discussion, we know how to define $f(K_i)$ for all $i \in [n]$. We define

$$f(J) = \begin{pmatrix} f(K_1) & & 0 \\ & \ddots & \\ 0 & & f(K_n) \end{pmatrix}.$$

We then define $f(A) = Sf(J)S^{-1}$. We must check that $f(A)$ is well defined. If J_1, J_2 are two Jordan canonical forms of A, then there are invertible $S_1, S_2 \in M_n(\mathbb{F})$ such that $J_i = S_i^{-1}AS_i$ for $i = 1, 2$. We must show $S_1 f(J_1) S_1^{-1} = S_2 f(J_2) S_2^{-1}$. By 7.5.3, we know J_1 and J_2 differ only by a permutation of the Jordan blocks, and since $J_1 = S_1^{-1}S_2J_2S_2^{-1}S_1$, we must have that $S = S_1^{-1}S_2$ is a generalized permutation matrix. It is easy to see that $f(J_2) = Sf(J_1)S^{-1}$ as the Jordan blocks do not interact under multiplication by the generalized permutation matrix. Thus $S_1 f(J_1) S_1^{-1} = S_2 f(J_2) S_2^{-1}$, and we are finished.

Operators: Suppose $T \in L(V)$. By 7.4.4, there is a Jordan canonical basis B for T, so $[T]_B$ is in Jordan canonical form. Thus, we define

$$f(T) = [f([T]_B)]_B^{-1}.$$

We must check that if B' is another Jordan canonical basis for T, then $[f([T]_B)]_B^{-1} = [f([T]_{B'})]_{B'}^{-1}$. This follows directly from the above discussion and 3.4.13.

Examples 7.6.6.

(1) If $p \in \mathbb{F}[z]$, we have that the $p(T)$ defined in 4.6.2 agrees with the $p(T)$ defined in 7.6.5. Hence the holomorphic functional calculus is a generalization of the polynomial functional calculus when V is a complex vector space.

(2) We will compute $\cos(A)$ for

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

We have that

$$U = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \in M_2(\mathbb{C})$$

is a unitary matrix that maps the standard orthonormal basis to the orthonormal basis consisting of eigenvectors of

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

We then see that

$$D = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}^* \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} = U^*AU.$$

Since $A = UDU^*$, by the entire functional calculus, we have that

$$\begin{aligned} \cos(A) = \cos(UDU^*) = U\cos(D)U^* &= U \cos \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} U^* \\ &= U \begin{pmatrix} \cos(i) & 0 \\ 0 & \cos(-i) \end{pmatrix} U^* = U\cosh(1)IU^* = \cosh(1)I. \end{aligned}$$

Theorem 7.6.7 (Spectral Mapping)**.** *Suppose $T \in L(V)$ and $U \subset \mathbb{C}$ is open such that* $\operatorname{sp}(T) \subset U$. *Then* $\operatorname{sp}(f(T)) = f(\operatorname{sp}(T))$.

Proof. Exercise. □

Proposition 7.6.8. *Let $T \in L(V)$, let $f, g\colon U \to \mathbb{C}$ be holomorphic such that* $\operatorname{sp}(T) \subset U$, *and let $h\colon V \to \mathbb{C}$ be holomorphic such that* $\operatorname{sp}(f(T)) \subset V$. *The holomorphic functional calculus satisfies*

(1) $(f+g)(T) = f(T) + g(T)$,

(2) $(fg)(T) = f(T)g(T)$,

(3) $(\lambda f)(T) = \lambda f(T)$ *for all* $\lambda \in \mathbb{C}$, *and*

(4) $(h \circ f)(T) = h(f(T))$.

Proof. Exercise. □

Exercises

V will denote a finite dimensional vector space over $\mathbb{C}$.

Exercise 7.6.9. Compute

$$\sin\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} \text{ and } \exp\begin{pmatrix}0 & -\pi\\ \pi & 0\end{pmatrix}.$$

Exercise 7.6.10. Suppose V is a finite dimensional vector space over $\mathbb{C}$ and $T \in L(V)$ with $\operatorname{sp}(T) \subset B_1(1) \subset \mathbb{C}$. Use the holomorphic functional calculus to show T is invertible.

Hint: Look at $f(z) = \dfrac{1}{z} = \dfrac{1}{1-(1-z)}$. When is f holomorphic?

Exercise 7.6.11 (Square Roots)**.** Determine which matrices in $M_n(\mathbb{C})$ have square roots, i.e. all $A \in M_n(\mathbb{C})$ such that there is a $B \in M_n(\mathbb{C})$ with $B^2 = A$.

Hint: The function $g\colon \mathbb{C}\setminus\{0\} \to \mathbb{C}$ given by $g(z) = \sqrt{z}$ is holomorphic. First look at the case where A is a single Jordan block associated to $\lambda \in \mathbb{C}$. In particular, when does a Jordan block associated to $\lambda = 0$ have a square root?

Chapter 8

Sesquilinear Forms and Inner Product Spaces

For this chapter, $\mathbb{F}$ will denote either $\mathbb{R}$ or $\mathbb{C}$, and V will denote a vector space over $\mathbb{F}$.

8.1 Sesquilinear Forms

Definition 8.1.1. A sesquilinear form on V is a function $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{F}$ such that

(i) $\langle \cdot, \cdot \rangle$ is linear in the first variable, i.e. for each $v \in V$, the function $\langle \cdot, v \rangle \colon V \to \mathbb{F}$ given by $u \mapsto \langle u, v \rangle$ is a linear transformation, and

(ii) $\langle \cdot, \cdot \rangle$ is conjugate linear in the second variable, i.e. for each $v \in V$, the function $\overline{\langle v, \cdot \rangle} \colon V \to \mathbb{F}$ given by $u \mapsto \overline{\langle v, u \rangle}$ is a linear transformation.

The sesquilinear form $\langle \cdot, \cdot \rangle$ is called

(1) self adjoint if $\langle u, v \rangle = \overline{\langle v, u \rangle}$ for all $u, v \in V$,

(2) positive if $\langle v, v \rangle \geq 0$ for all $v \in V$, and

(3) definite if $\langle v, v \rangle = 0$ implies $v = 0$.

(4) an inner product it is a positive definite self-adjoint sesquilinear form.

An inner product space over $\mathbb{F}$ is a vector space V over $\mathbb{F}$ together with an inner product on V.

Remarks 8.1.2.

(1) Note that linearity in the first variable and self adjointness of a sesquilinear form $\langle \cdot, \cdot \rangle$ implies that $\langle \cdot, \cdot \rangle$ is conjugate linear in the second variable.

(2) If V is a vector space over $\mathbb{R}$, then a sesquilinear form on V is usually called a bilinear form, and the adjective “self adjoint” is replaced by “symmetric.” Note that in this case, conjugate linear in the second variable means linear in the second variable. Note further that if $\langle \cdot, \cdot \rangle$ is linear in the first variable and symmetric, then it is also linear in the second variable.

Examples 8.1.3.

(1) The function

$$\langle u, v\rangle = \sum_{i=1}^{n} e_i^*(u)\overline{e_i^*(v)}$$

is the standard inner product on $\mathbb{F}^n$.

(2) Let $x_0, x_1, \dots, x_n$ be $n+1$ points in $\mathbb{F}$. Then

$$\langle p, q\rangle = \sum_{i=0}^{n} p(x_i)\overline{q(x_i)}$$

is an inner product on P_m if $m \leq n$, but it is not definite if $m > n$ (including $m = \infty$).

(3) The function

$$\langle f, g\rangle = \int\limits_a^b f(x)\overline{g(x)}\ dx$$

is the standard inner product on $C([a, b], \mathbb{F})$

(4) trace: $M_n(\mathbb{F}) \to \mathbb{F}$ given by

$$\operatorname{trace}(A) = \sum_{i=1}^{n} A_{ii}$$

induces an inner product on $M_{m\times n}(\mathbb{F})$ by $\langle A, B\rangle = \operatorname{tr}(B^*A)$.

(5) Let $a, b \in \mathbb{R}$. Then the function

$$\langle p, q\rangle = \int\limits_a^b p(x)\overline{q(x)}\ dx$$

is an inner product on $\mathbb{F}[x]$.

(6) Suppose V is a real inner product space. Then the complexification $V_{\mathbb{C}}$ defined in 2.1.12 is a complex inner product space with inner product given by

$$\langle u_1 + iv_1, u_2 + iv_2\rangle_{\mathbb{C}} = \langle u_1, v_1\rangle + \langle v_1, v_2\rangle + i(\langle u_2, v_1\rangle - \langle u_1, v_2\rangle).$$

In particular, note that $\langle\cdot,\cdot\rangle_{\mathbb{C}}$ is definite.

Proposition 8.1.4 (Polarization Identity)**.**

(1) If V is a vector space space over $\mathbb{R}$ and $\langle\cdot,\cdot\rangle$ is a symmetric bilinear form, then

$$4\langle u, v\rangle = \langle u + v, u + v\rangle - \langle u - v, u - v\rangle.$$

(2) If V is a vector space space over $\mathbb{C}$ and $\langle\cdot,\cdot\rangle$ is a self adjoint sesquilinear form, then

$$4\langle u,v\rangle = \sum_{k=0}^{3} i^k\langle u+i^kv, u+i^kv\rangle.$$

Proof. This is immediate from the definition of a symmetric bilinear form or self adjoint sesquilinear form respectively. □

Proposition 8.1.5 (Cauchy-Schwartz Inequality)**.** *Let $\langle\cdot,\cdot\rangle$ be a positive, self adjoint sesquilinear form on V. Then*

$$|\langle u,v\rangle|^2 \leq \langle u,u\rangle\langle v,v\rangle \text{ for all } u,v\in V.$$

Proof. Let $u,v\in V$. Since $\langle\cdot,\cdot\rangle$ is positive, we know that for all $\lambda\in\mathbb{F}$,

$$0 \leq \langle u-\lambda v, u-\lambda v\rangle = \langle u,u\rangle - 2\operatorname{Re}\overline{\lambda}\langle u,v\rangle + |\lambda|^2\langle v,v\rangle. \qquad (*)$$

Case 1: If $\langle v,v\rangle = 0$, then $2\operatorname{Re}\overline{\lambda}\langle u,v\rangle \leq \langle u,u\rangle$. Since this holds for all λ, we must have $\langle u,v\rangle = 0$.

Case 2: If $\langle v,v\rangle \neq 0$, then in particular, equation $(*)$ holds for

$$\overline{\lambda} = \frac{\langle v,u\rangle}{\langle v,v\rangle}$$

Hence,

$$0 \leq \langle u,u\rangle - 2\operatorname{Re}\frac{\langle v,u\rangle}{\langle v,v\rangle}\langle u,v\rangle + \left|\frac{\langle v,u\rangle}{\langle v,v\rangle}\right|^2\langle v,v\rangle = \langle u,u\rangle - 2\frac{|\langle u,v\rangle|^2}{\langle v,v\rangle} + \frac{|\langle u,v\rangle|^2}{\langle v,v\rangle} = \langle u,u\rangle - \frac{|\langle u,v\rangle|^2}{\langle v,v\rangle}.$$

Thus,

$$\frac{|\langle u,v\rangle|^2}{\langle v,v\rangle} \leq \langle u,u\rangle \Longrightarrow |\langle u,v\rangle|^2 \leq \langle u,u\rangle\langle v,v\rangle.$$

□

Exercises

Exercise 8.1.6. Show that the sesquilinear form

$$\langle f,g\rangle = \int_a^b f(x)\overline{g(x)}\ dx$$

on $C([a,b],\mathbb{F})$ is an inner product.

Hint: Show that if $|f(x)| > 0$ for some $x\in[a,b]$, then by the continuity of f, there is some $\delta > 0$ such that $|f(y)| > 0$ for all $y\in(x-\delta,x+\delta)\cap(a,b)$.

Exercise 8.1.7. Show that the function $\langle\cdot,\cdot\rangle_2\colon M_n(\mathbb{F})\times M_n(\mathbb{F})\to\mathbb{F}$ given by $\langle A,B\rangle_2 = \operatorname{trace}(B^*A)$ is an inner product on $M_n(\mathbb{F})$.

8.2 Inner Products and Norms

From this point on, $(V, \langle\cdot,\cdot\rangle)$ will denote an inner product space over $\mathbb{F}$.

We illustrate an important proof technique called "true for all, then true for a specific one."

Proposition 8.2.1.

(1) Suppose $u, v \in V$ such that $\langle u, w\rangle = \langle v, w\rangle$ for all $w \in W$. Then $u = v$.

(2) Suppose $S, T \in L(V)$ such that $\langle Sx, y\rangle = \langle Tx, y\rangle$ for all $x, y \in V$. Then $S = T$.

(3) Suppose $S, T \in L(V)$ such that $\langle x, Sy\rangle = \langle x, Ty\rangle$ for all $x, y \in V$. Then $S = T$.

Proof.

(1) We have that $\langle u - v, w\rangle = 0$ for all $w \in V$. In particular, this holds for $w = u - v$. Hence $\langle u - v, u - v\rangle = 0$, so $u - v - 0$ by definiteness. Hence $u = v$.

(2) Let $x \in V$, and set $u = Sx$ and $v = Tx$. Applying (1), we see $Sx = Tx$. Since this is true for all $x \in V$, we have $S = T$.

(3) This follows immediately from self-adjointness of an inner product and (2). □

Definition 8.2.2. A norm on V is a function $\|\cdot\|\colon V \to \mathbb{R}_{\geq 0}$ such that

(i) (definiteness) $\|v\| = 0$ implies $v = 0$,

(ii) (homogeneity) $\|\lambda v\| = |\lambda| \cdot \|v\|$ for all $\lambda \in \mathbb{F}$ and $v \in V$, and

(iii) (triangle inequality) $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in V$.

Examples 8.2.3.

(1) The Euclidean norm on $\mathbb{F}^n$ is given by

$$\|v\| = \sqrt{\sum_{i=1}^{n} e_i^*(v)}.$$

We will see in 8.2.4 that it is the norm induced by the standard inner product on $\mathbb{F}^n$.

(2) The 1-norm on $C([a, b], \mathbb{F})$ is given by

$$\|f\|_1 = \int_a^b |f| \ dx,$$

and the 2-norm is given by

$$\|f\|_2 = \left(\int_a^b |f|^2 \ dx\right)^{1/2}.$$

We will see in 8.2.4 that the 2-norm is the norm induced from the standard inner product on $C([a,b],\mathbb{F})$. The ∞-norm is given by

$$\|f\|_\infty = \max\left\{|f(x)|\big|x \in [a,b]\right\}.$$

It exists by the Extreme Value Theorem, which says that a continuous, real-valued function, namely $|f|$, achieves its maximum on a closed, bounded interval, namely $[a,b]$. These norms are all different. For example, if $[a,b]=[0,2\pi]$ and $f\colon [0,2\pi]\to\mathbb{R}$ is given by $f(x)=\sin(x)$, then $\|f\|_1 = 4$, $\|f\|_2=\pi$, and $\|f\|_\infty = 1$.

(3) The norms defined in (2) can all be defined for $\mathbb{F}[x]$ as well.

Proposition 8.2.4. *The function $\|\cdot\|\colon V\to\mathbb{R}_{\geq 0}$ given by $\|v\| = \sqrt{\langle v,v\rangle}$ is a norm. It is usually called the induced norm on V.*

Proof. Clearly $\|\cdot\|$ is definite by definition. It is homogeneous since

$$\|\lambda v\| = \sqrt{\langle \lambda v, \lambda v\rangle} = \sqrt{|\lambda|^2\langle v,v\rangle} = |\lambda|\cdot\|v\|.$$

Now the Cauchy-Schwartz inequality can be written as $|\langle u,v\rangle| \leq \|u\|\|v\|$ after taking square roots, so we have

$$\begin{aligned}\|u+v\|^2 &= \langle u+v,u+v\rangle = \langle u,u\rangle + 2\operatorname{Re}\langle u,v\rangle + \langle v,v\rangle\\ &\leq \langle u,u\rangle + 2|\langle u,v\rangle| + \langle v,v\rangle \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\|+\|v\|)^2.\end{aligned}$$

Taking square roots now gives the desired result. □

Remark 8.2.5. In many books, the Cauchy-Scwartz inequality is proved for inner products (not positive, self adjoint sesquilinear forms as was done in 8.1.5), and it is usually in the form used in the proof of 8.2.4:

$$|\langle u,v\rangle| \leq \|u\|\|v\|.$$

Proposition 8.2.6 (Parallelogram Identity)**.** *The induced norm $\|\cdot\|$ on V satisfies*

$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2 \text{ for all } u,v\in V.$$

Proof. This is immediate from the definition of $\|\cdot\|$. □

Definition 8.2.7.

(1) If $u,v\in V$ and $\langle u,v\rangle = 0$, then we say u is perpendicular to v. Sometimes this is denoted as $u\perp v$.

(2) Let $S\subset V$ be a subset. Then $S^\perp = \left\{v\in V\big|v\perp s \text{ for all } s\in S\right\}$ is a subspace of V.

(3) We say the set S_1 is orthogonal to the set S_2, denoted $S_1\perp S_2$ if $v\in S_1$ and $w\in S_2$ implies $v\perp w$.

Examples 8.2.8.

(1) The standard basis vectors in $\mathbb{F}^n$ are all pairwise orthogonal. If we pick $v \in \mathbb{F}^n$, then $\{v\}^\perp \cong \mathbb{F}^{n-1}$.

(2) Suppose we have the inner product on $\mathbb{R}[x]$ given by

$$\langle p, q \rangle = \int_0^1 p(x)\overline{q(x)}\ dx.$$

Then $1 \perp x - 1/2$. Moreover, $\{1\}^\perp$ is infinite dimensional as $x^n - 1/(n+1) \in \{1\}^\perp$ for all $n \in \mathbb{N}$.

Proposition 8.2.9 (Pythagorean Theorem)**.** *Suppose $u, v \in V$ such that $u \perp v$. Then*

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Proof. $\|u+v\|^2 = \langle u+v, u+v \rangle = \langle u, u \rangle + 2\operatorname{Re}\langle u, v \rangle + \langle v, v \rangle = \|u\|^2 + \|v\|^2$. □

Exercises

Exercise 8.2.10.

8.3 Orthonormality

Definition 8.3.1.

(1) A subset $S \subset V$ is called an orthogonal set if $u, v \in S$ with $u \neq v$ implies that $u \perp v$.

(2) A subset $S \subset V$ is called an orthonormal set if S is an orthogonal set and $v \in S$ implies $\|v\| = 1$.

Examples 8.3.2.

(1) Zero is never in an orthonromal set, but can be in an orthogonal set.

Proposition 8.3.3. *Let S be an orthogonal set such that $0 \notin S$. Then S is linearly independent. Hence all orthonormal sets are linearly independent.*

Proof. Let $\{v_1, \ldots, v_n\}$ be a finite subset of S, and suppose there are scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ such that

$$\sum_{i=1}^n \lambda_i v_i = 0.$$

Then we have a linear functional $v_j^* = \langle \cdot, v_j \rangle \in V^*$ given by $v_j^*(u) = \langle u, v_j \rangle$ for all $u \in V$. We apply φ_j to the above expression ("hit is with v_j^*") to get

$$0 = v_j\left(\sum_{i=1}^n \lambda_i v_i\right) = \sum_{i=1}^n \lambda_i v_j^*(v_i) = \sum_{i=1}^n \lambda_i \langle v_i, v_j \rangle = \lambda_j \langle v_j, v_j \rangle = \lambda_j \|v_j\|^2.$$

as $\langle v_i, v_j \rangle = 0$ if $i \neq j$. Since $v_j \neq 0$, we have $\|v_j\|^2 \neq 0$, so $\lambda_j = 0$. □

Theorem 8.3.4 (Gram-Schmidt Orthonormalization)**.** *Let $S = \{v_1, \dots, v_n\}$ be a linearly independent subset of V. Then there is an orthonormal set $\{u_1, \dots, u_n\}$ such that for each $k \in \{1, \dots, n\}$,* $\operatorname{span}\{u_1, \dots, u_k\} = \operatorname{span}\{v_1, \dots, v_k\}$.

Proof. The proof is by induction on n.

$\underline{n = 1:}$ Setting $u_1 = v_1/\|v_1\|$ works.

$\underline{n - 1 \Rightarrow n:}$ Suppose we have a linearly independent set $S = \{v_1, \dots, v_n\}$. By the induction hypothesis, there is an orthonormal set $\{u_1, \dots, u_{n-1}\}$ such that $\operatorname{span}\{u_1, \dots, u_k\} = \operatorname{span}\{v_1, \dots, v_k\}$ for each $k \in \{1, \dots, n-1\}$. Set

$$w_n = v_n - \sum_{i=1}^{n-1} \langle v_{n+1}, u_i\rangle u_i \text{ and } u_n = \frac{w_n}{\|w_n\|}.$$

It is clear that $w_{n+1} \perp u_i$ for all $j = 1, \dots, n-1$ by applying the linear functional $\langle \cdot, u_j\rangle$:

$$\langle w_{n+1}, u_j\rangle = \langle v_{n+1}, u_j\rangle - \sum_{i=1}^{n-1} \langle v_{n+1}, u_i\rangle\langle u_i, u_j\rangle = \langle v_{n+1}, u_j\rangle - \langle v_{n+1}, u_j\rangle\langle u_j, u_j\rangle = 0.$$

Hence $u_n \perp u_j$ for all $j = 1, \dots, n-1$, and $R = \{u_1, \dots, u_{n+1}\}$ is an orthonormal set. It remains to show that $\operatorname{span}(S) = \operatorname{span}(R)$. Recall that $\operatorname{span}(S \setminus \{u_n\}) = \operatorname{span}(R \setminus \{v_n\})$. It is clear that $\operatorname{span}(R) \subseteq \operatorname{span}(S)$ since $v_1, \dots, v_{n-1} \in \operatorname{span}(S)$ and v_n is a linear combination of $u_1, \dots, u_n$. But we immediately see $\operatorname{span}(S) \subseteq \operatorname{span}(R)$ as $u_1, \dots, u_{n-1} \in \operatorname{span}(R)$ and u_n is a linear combination of $u_1, \dots, u_{n-1}$ and v_n. □

Definition 8.3.5. If V is a finite dimensional inner product space over $\mathbb{F}$, a subset $B \subset V$ is called an orthonormal basis of V if B is an orthonormal set and B is a basis of V.

Examples 8.3.6.

(1) The standard basis of $\mathbb{F}^n$ is an orthonormal basis.

(2) The set $\{1, x, x^2, \dots, x^n\}$ is a basis, but not an orthonormal basis of P_n with the inner product given by

$$\langle p, q\rangle = \int_0^1 p(x)\overline{q(x)}\, dx.$$

Theorem 8.3.7 (Existence of Orthonormal Bases)**.** *Let V be a finite dimensional inner product space over $\mathbb{F}$. Then V has an orthonormal basis.*

Proof. Let B be a basis of V. By 8.3.4, there is an orthonormal set C such that $\operatorname{span}(B) = \operatorname{span}(C)$. Moreover, C is linearly independent by 8.3.3. Hence C is an orthonormal basis for V. □

Proposition 8.3.8. *Let $B = \{v_1, \dots, v_n\} \subset V$ be an orthogonal set that is also basis of V. Then if $w \in V$,*

$$w = \sum_{i=1}^{n} \frac{\langle w, v_i \rangle}{\langle v_i, v_i \rangle} v_i.$$

If B is an orthonormal basis for V, then this expression simplifies to

$$w = \sum_{i=1}^{n} \langle w, v_i \rangle v_i,$$

and the $\langle w, v_i \rangle$ are called Fourier coefficients of w with respect to B.

Proof. Since B spans V, there are scalars $\lambda_1, \dots, \lambda_n \in \mathbb{F}$ such that

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

Now we apply $\langle \cdot, v_j \rangle$ to see that

$$\langle w, v_j \rangle = \sum_{i=1}^{n} \lambda_i \langle v_i, v_j \rangle = \lambda_j \langle v_j, v_j \rangle$$

as $\langle v_i, v_j \rangle = 0$ if $i \neq j$. Dividing by $\langle v_j, v_j \rangle$ gives the desired formula. □

Proposition 8.3.9 (Parseval's Identity). *Suppose $B = \{v_1, \dots, v_n\}$ be an orthonormal basis of V. Then*

$$\|v\|^2 = \sum_{i=1}^{n} |\langle v, v_i \rangle|^2 \text{ for all } v \in V.$$

Proof. The reader may check that this identity follows immediately from 8.3.8 using induction and the Pythagorean Theorem (8.2.9). □

Fact 8.3.10. *Suppose $B = (v_1, \dots, v_n)$ is an ordered basis of the vector space V. We may impose an inner product on V by setting*

$$\langle u, v \rangle_B = \langle [u]_B, [v]_B \rangle_{\mathbb{F}^n},$$

and we check that

$$\langle v_i, v_j \rangle_B = \langle [v_i]_B, [v_j]_B \rangle_{\mathbb{F}^n} = \langle e_i, e_j \rangle_{\mathbb{F}^n} = \begin{cases} 1 & \text{if i=j} \\ 0 & \text{else,} \end{cases}$$

so B is an orthonormal basis of V with inner product $\langle \cdot, \cdot \rangle_B$. Moreover, we see that the linear functional $v_j^* = \langle \cdot, v_j \rangle$ for all $j \in [n]$. If

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

then we have that

$$\langle v, v_j \rangle_B = \left\langle \sum_{i=1}^{n} \lambda_i v_i, v_j \right\rangle_B = \sum_{i=1}^{n} \lambda_i \langle v_i, v_j \rangle_B = \lambda_j = v_j^*(v).$$

Exercises

Exercise 8.3.11. Use Gram-Schmidt orthonormalization on $\{1, z, z^2, z^3\}$ to find an orthonormal basis for $P_3(\mathbb{F})$ with inner product

$$\langle p, q\rangle = \int_0^1 p(z)\overline{q(z)}\ dz.$$

Exercise 8.3.12. Suppose $\{v_1, \ldots, v_n\}$ is an orthonormal basis for V. Show that the linear functionals $\langle \cdot v_i, v_j\rangle \colon L(V) \to \mathbb{F}$ given by $T \mapsto \langle Tv_i, v_j\rangle$ are a basis for $L(V)^*$.

8.4 Finite Dimensional Inner Product Subspaces

Lemma 8.4.1. *Let W be a finite dimensional subspace of V, and let $\{w_1, \ldots, w_n\}$ be an orthonormal basis for W. Then $x \in W^\perp$ if $x \perp w_i$ for all $i = 1, \ldots, n$.*

Proof. Let $w \in W$. Then by 8.3.8,

$$w = \sum_{i=1}^{n} \langle w, w_i\rangle w_i.$$

Then we have

$$\langle x, w\rangle = \sum_{i=1}^{n} \langle w, w_i\rangle \langle x, w_i\rangle = 0,$$

so $x \in W^\perp$. □

Proposition 8.4.2. *Let $W \subset V$ be a finite dimensional subspace, and let $v \in V$. Then there is a unique $w \in W$ minimizing $\|v - w\|$, i.e. $\|v - w\| \leq \|v - w'\|$ for all $w' \in W$.*

Proof. Let $\{w_1, \ldots, w_n\}$ be an orthonormal basis for W. Set

$$w = \sum_{i=1}^{n} \langle v, w_i\rangle w_i \in W$$

and $u = v - w$. We have $u \in W^\perp$ by 8.4.1 as $\langle u, w_j\rangle = 0$ for all $j = 1, \ldots, n$:

$$\langle u, w_j\rangle = \langle v, w_j\rangle - \sum_{i=1}^{n} \langle v, w_i\rangle \langle w_i, w_j\rangle = \langle v, w_j\rangle - \langle v, w_j\rangle \langle w_j, w_j\rangle = 0.$$

We show $w \in W$ is the unique vector minimizing the distance to v. Suppose $w' \in W$ such that $\|v - w'\| \leq \|v - w\|$. Since $v = u + w$ and $u \perp (w - w')$, by 8.2.9,

$$\|u\|^2 + \|w - w'\|^2 = \|u + (w - w')\|^2 = \|v - w'\|^2 \leq \|v - w\|^2 = \|u\|^2.$$

Hence $\|w - w'\|^2 = 0$ and $w = w'$. □

Definition 8.4.3. Let $W \subset V$ be a finite dimensional subspace. Then P_W, the projection onto W, is the operator in $L(V)$ defined by $P_W(v) = w$ where w is the unique vector in W closest to v which exists by 8.4.2.

Examples 8.4.4.

(1) Let $v \in V$. The projection onto $\operatorname{span}\{v\}$ denoted by P_v, is is the operator in $L(V)$ given by

$$u \longmapsto \frac{\langle u, v\rangle}{\|v\|^2} v.$$

The corresponding component operator is the linear functional $C_v \in V^*$ given by

$$u \longmapsto \frac{\langle u, v\rangle}{\|v\|^2}.$$

Note that if $\|v\| = 1$, the formulas simplify to $P_v(u) = \langle u, v\rangle v$ and $C_v(u) = \langle u, v\rangle$.

(2) Suppose $u, v \in V$ with $u, v \neq 0$ and $u = \lambda v$ for some $\lambda \in \mathbb{F}$. Then $P_u = P_v$.

(3) Let W be a finite dimensional subspace of V, and let $\{w_1, \dots, w_n\}$ be an orthonormal basis of W. Then the projection onto W is the operator in $L(V)$ given by

$$u \longmapsto \sum_{i=1}^{n} \langle u, w_i\rangle w_i.$$

In particular, this definition is independent of the choice of orthonormal basis for W.

Lemma 8.4.5. *Let $W \subset V$ be a finite dimensional subspace. Then*

(1) $P_W^2 = P_W$,

(2) $\operatorname{im}(P_W) = \big\{v \in V \big| P_W(v) = v\big\} = W$,

(3) $\ker(P_W) = W^{\perp}$, *and*

(4) $V = W \oplus W^{\perp}$.

Proof. We have that (2)-(4) follow immediately from 6.1.6 if (1) holds.

(1) Let $v \in V$. Then there is a unique $w \in W$ closest to v by 8.4.2. Then by the definition of P_W, we have $P_W(v) = w = P_W(w)$. Hence $P_W^2(v) = P_W(w) = w = P_W(v)$, and $P_W^2 = P_W$. □

Corollary 8.4.6. $\langle P_W u, v\rangle = \langle u, P_W v\rangle$ *for all* $u, v \in V$.

Proof. Let $u, v \in V$. Then by 2.2.8, there are unique $w_1, w_2 \in W$ and $x_1, x_2 \in W^{\perp}$ such that $u = w_1 + x_1$ and $v = w_2 + x_2$. Then

$$\begin{aligned}\langle P_W u, v\rangle &= \langle P_W(w_1 + x_1), w_2 + x_2\rangle = \langle w_1, w_2 + x_2\rangle = \langle w_1, w_2\rangle \\ &= \langle w_1 + x_1, w_2\rangle = \langle w_1 + x_1, P_W(w_2 + x_2)\rangle = \langle u, P_W v\rangle.\end{aligned}$$

□

Exercises

V will denote an inner product space over $\mathbb{F}$.

Exercise 8.4.7 (Invariant Subspaces)**.** Let $T \in L(V)$, and let $W \subset V$ be a finite dimensional subspace.

(1) Show W is T-invariant if and only if $P_W T P_W = T P_W$.

(2) Show W and $W^\perp$ are T-invariant if and only if $T P_W = P_W T$.

Chapter 9

Operators on Hilbert Space

9.1 Hilbert Spaces

The preceding discussion brings up a few questions. What can we say about subspaces of an inner product space V that are not finite dimensional? Is there a projection P_W onto an infinite dimensional subspace? Is it still true that $V = W \oplus W^\perp$ if W is infinite dimensional? As we are not assuming a knowledge of basic topology and analysis, there is not much we can say about these questions. Hence, we will need to restrict our attention to finite dimensional inner product spaces, which are particular examples of Hilbert spaces.

Definition 9.1.1. For these notes, a Hilbert space over $\mathbb{F}$ will mean a finite dimensional inner product space over $\mathbb{F}$.

Remark 9.1.2. The study of operators on infinite dimensional Hilbert spaces is a vast area of research that is widely popular today. One of the biggest differences between an undergraduate course on linear algebra and a graduate course in functional analysis is that in the undergraduate course, one only studies finite dimensional Hilbert spaces.

For this section H will denote a Hilbert space over $\mathbb{F}$. In this section, we discuss various types of operators in $L(H)$.

Proposition 9.1.3. *Let $K \subset H$ be a subspace. Then $(K^\perp)^\perp = K$.*

Proof. It is obvious that $K \subset (K^\perp)^\perp$. We know $H = K \oplus K^\perp$ by 8.4.5. By 8.3.7, choose orthonormal bases $B = \{v_1, \dots, v_n\}$ and $C = \{u_1, \dots, u_m\}$ for K and $K^\perp$ respectively. Then $B \cup C$ is an orthonormal basis for H by 2.4.13. Suppose $w \in (K^\perp)^\perp$. Then by 8.3.8,

$$w = \sum_{i=1}^{n} \langle w, v_i \rangle v_i + \sum_{i=1}^{m} \langle w, u_i \rangle u_i,$$

but $w \perp u_i$ for all $i = 1, \dots, m$, so $w \in \operatorname{span}(B) = K$. Hence $K \subset (K^\perp)^\perp$. □

Lemma 9.1.4. *Suppose $T \in L(H, V)$ where V is a vector space. Then $T = TP_{\ker(T)^\perp}$.*

Proof. We know that $H = \ker(T) \oplus \ker(T)^\perp$ by 8.4.5. Let $v \in \ker(T)$ and $w \in \ker(T)^\perp$. Then $T(v+w) = Tv + Tw = 0 + Tw = Tw = TP_{\ker(T)^\perp}(v+w)$. Thus, $T = TP_{\ker(T)^\perp}$. □

Theorem 9.1.5 (Reisz Representation)**.** *The map* $\Phi\colon H \to H^*$ *given by* $v \mapsto \langle \cdot, v\rangle$ *is a conjugate-linear isomorphism, i.e.* $\Phi(\lambda u + v) = \overline{\lambda}\Phi(u) + \Phi(v)$ *for all* $\lambda \in \mathbb{F}$ *and* $u, v \in H$*, and* Φ *is bijective.*

Proof. It is obvious that the map is conjugate linear. We show Φ is injective. Suppose $\langle \cdot, v\rangle = 0$, i.e. $\langle u, v\rangle = 0$ for all $u \in V$. Then in particular, $\langle v, v\rangle = 0$, so $v = 0$ by definiteness. Note that the proof of 3.2.4 still works for conjugate-linear transformations, so Φ is injective. We show Φ is surjective. It is clear that $\Phi(0) = 0$. Suppose that $\varphi \in H^*$ with $\varphi \neq 0$. Then $\ker(\varphi) \neq H$, so $\ker(\varphi)^\perp \neq (0)$. Pick $v \in \ker(\varphi)^\perp$ such that $\varphi(v) \neq 0$. Now consider the functional

$$\psi = \varphi(v)\frac{\langle \cdot, v\rangle}{\|v\|^2} = \Phi\left(\frac{\varphi(v)}{\|v\|^2}v\right).$$

Since $\operatorname{span}\{v\} = \ker(\varphi)^\perp$, by 9.1.4 and 3.3.2 we have

$$\psi(u) = \varphi(v)\frac{\langle u, v\rangle}{\|v\|^2} = \varphi\left(\frac{\langle u, v\rangle}{\|v\|^2}v\right) = \varphi(P_v(u)) = \varphi(u).$$

Hence $\psi = \varphi$, and Φ is surjective. □

Exercises

Exercise 9.1.6 (Sesquilinear Forms and Operators)**.** Show that there is a bijective correspondence $\Psi\colon L(H) \to \{\text{sesquilinear forms on } H\}$.

9.2 Adjoints

Definition 9.2.1. Let $T \in L(H)$. Then if $v \in H$, $u \mapsto \langle Tu, v\rangle = (\Phi(v) \circ T)(u)$ defines a linear operator on H. By the Reisz Representation Theorem, 9.1.5, there is a vector in H, which we will denote T^*v, such that

$$\langle Tu, v\rangle = \langle u, T^*v\rangle \text{ for all } u \in H.$$

We show the map $T^*\colon H \to H$ given by $v \mapsto T^*v$ is linear. Suppose $\lambda \in \mathbb{F}$ and $w \in H$. Then

$$\langle Tu, \lambda v + w\rangle = \overline{\lambda}\langle Tu, v\rangle + \langle Tu, w\rangle = \overline{\lambda}\langle u, T^*v\rangle + \langle u, T^*w\rangle = \langle u, \lambda T^*v + T^*w\rangle$$

for all $u \in H$. Hence $T^*(\lambda v + w) = \lambda T^*v + T^*w$ by 8.2.1. The map T^* is called the adjoint of T.

Examples 9.2.2.

(1) For $\lambda I \in L(H)$, $(\lambda I)^* = \overline{\lambda}I$.

(2) For $A \in M_n(\mathbb{F})$, we have $(L_A)^* = L_{A^*}$, multiplication by the adjoint matrix.

(3) Suppose H is a real Hilbert space and $T \in L(H)$. Then $(T_\mathbb{C})^* \in L(H_\mathbb{C})$ is defined by the following formula:

$$(T_\mathbb{C})^*(u + iv) = T^*u + iT^*v.$$

In other words, $(T_\mathbb{C})^* = (T^*)_\mathbb{C}$.

Proposition 9.2.3. *Let $S, T \in L(H)$ and $\lambda \in \mathbb{F}$.*

(1) The map $*\colon L(H) \to L(H)$ is conjugate-linear, i.e. $(\lambda T + S)^* = \overline{\lambda} T^* + S$,

(2) $(ST)^* = T^*S^*$, and

(3) $T^{**} = (T^*)^* = T$.

In short, $*$ is a conjugate-linear, anti-automorphism of period two.

Proof.

(1) For all $u, v \in H$, we have

$$\langle u, (\lambda T+S)^*v\rangle = \langle (\lambda T+S)u, v\rangle = \lambda\langle Tu, v\rangle + \langle Su, v\rangle = \langle u, \overline{\lambda}T^*v\rangle + \langle u, S^*v\rangle = \langle u, (\overline{\lambda}T^*+S^*)v\rangle.$$

By 8.2.1 we get the desired result.

(2) For all $u, v \in H$, we have

$$\langle u, (ST)^*v\rangle = \langle STu, v\rangle = \langle u, T^*S^*v\rangle.$$

By 8.2.1 we get the desired result.

(3) For all $u, v \in H$, we have

$$\langle Tu, v\rangle = \langle u, T^*v\rangle = \overline{\langle T^*v, u\rangle} = \overline{\langle v, T^{**}u\rangle} = \langle T^{**}u, v\rangle.$$

Hence, by 8.2.1, we get $T = T^{**}$. □

Definition 9.2.4. An operator $T \in L(H)$ is called

(1) normal if $T^*T = TT^*$,

(2) self adjoint if $T = T^*$ (note that a self adjoint operator is normal),

(3) positive if T is self adjoint and $\langle Tv, v\rangle \geq 0$ for all $v \in H$, and

(4) positive definite if T is positive and $\langle Tv, v\rangle = 0$ implies $v = 0$.

(5) A matrix $A \in M_n(\mathbb{F})$ is called positive (definite) if L_A is positive (definite).

Examples 9.2.5.

(1) The following matrices in $M_2(\mathbb{F})$ are normal:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}.$$

(2)

Definition 9.2.6.

(1) An operator $T \in L(H)$ is called an isometry if $T^*T = I$.

(2) An operator $U \in L(H)$ is called unitary if $U^*U = UU^* = I$.

(3) An operator $P \in L(H)$ is called a projection (or an orthogonal projection) if $P = P^* = P^2$.

(4) An operator $V \in L(H)$ is called a partial isometry if V^*V is a projection.

Examples 9.2.7.

(1) Suppose $A \in M_n(\mathbb{F})$. Then L_A is unitary if and only if $A^*A = AA^* = I$. In fact, L_A is unitary if and only if the columns of A are orthonormal if and only if the rows of A are orthonormal.

(2) The simplest example of a projection is L_A where $A \in M_n(\mathbb{F})$ is a diagonal matrix with only zeroes and ones on the diagonal.

(3) If $T \in L(H)$ is a projection or (partial) isometry, and if $U \in L(H)$ is unitary, then U^*TU is a projection or (partial) isometry respectively.

(4) Let $H = P_n$ with the inner product

$$\langle p, q\rangle = \int\limits_0^1 p(x)\overline{q(x)}\ dx.$$

Then multiplication by $p \in P_n$ is a unitary operator if and only if $|p(x)| = 1$ for all $x \in [0,1]$ if and only if $p(x) = \lambda$ where $|\lambda| = 1$ for all $x \in [0,1]$.

Exercises

Exercise 9.2.8. Let B be an orthonormal basis for H, and let $T \in L(H)$. Show that $[T^*]_B = [T]_B^*$.

Exercise 9.2.9 (Sesquilinear Forms and Operators 2)**.** Let $\Psi\colon L(H) \to \{\text{sesquilinear forms on } H\}$ be the bijective correspondence found in 9.1.6. For $T \in L(H)$, show that the map Ψ satisfies

(1) T is self adjoint if and only if $\Psi(T)$ is self adjoint,

(2) T is positive if and only if $\Psi(T)$ is positive, and

(2) T is positive definite if and only if $\Psi(T)$ is an inner product.

Exercise 9.2.10. Suppose $T \in L(H)$. Show

(1) if T is self adjoint, then $\operatorname{sp}(T) \subset \mathbb{R}$,

(2) if T is positive, then $\operatorname{sp}(T) \subset [0,\infty)$, and

(3) if T is positive definite, then $\operatorname{sp}(T) \subset (0,\infty)$.

9.3 Unitaries

Lemma 9.3.1. *Suppose B is an orthonormal basis of H. Then $[\cdot]_B$ preserves inner products, i.e.*

$$\langle u, v\rangle_H = \langle [u]_B, [v]_B\rangle_{\mathbb{F}^n} \text{ for all } u, v \in H.$$

Proof. Let $B = \{v_1, \dots, v_n\}$, and suppose

$$u = \sum_{i=1}^n \mu_i v_i \text{ and } v = \sum_{i=1}^n \lambda_i v_i.$$

Then we have

$$[u]_B = \sum_{i=1}^n \mu_i e_i \text{ and } [v]_B = \sum_{i=1}^n \lambda_i e_i,$$

so

$$\langle [u]_B, [v]_B\rangle_{\mathbb{F}^n} = \sum_{i=1}^n \mu_i \overline{\lambda_i} = \langle u, v\rangle_H.$$

□

Theorem 9.3.2 (Unitary)**.** *The following are equivalent for $U \in L(H)$:*

(1) U is unitary,

(2) U^* is unitary,

(3) U is an isometry,

(4) $\langle Uv, Uw\rangle = \langle v, w\rangle$ for all $v, w \in H$,

(5) $\|Uv\| = \|v\|$ for all $v \in H$,

(6) If B is an orthonormal basis of H, then UB is an orthonormal basis of H, i.e. U maps orthonormal bases to bases,

(7) If B is an orthonormal basis of H, then the columns of $[U]_B$ form an orthonormal basis of $\mathbb{F}^n$, and

(8) If B is an orthonormal basis of H, then the rows of $[U]_B$ form an orthonormal basis of $\mathbb{F}^n$.

Proof.

(1) ⇔ (2): Obvious.

(1) ⇒ (3): Obvious.

(3) ⇒ (4): Suppose $U^*U = I$, and let $v, w \in H$. Then

$$\langle Uv, Uw\rangle = \langle U^*Uv, w\rangle = \langle Iv, w\rangle = \langle v, w\rangle.$$

(4) ⇒ (5): Suppose $\langle Uv, Uw\rangle = \langle v, w\rangle$ for all $v, w \in H$. Then

$$\|Uv\|^2 = \langle Uv, Uv\rangle = \langle v, v\rangle = \langle v, v\rangle = \|v\|^2.$$

Now take square roots.

(5) ⇒ (1): Suppose $\|Uv\| = \|v\|$ for all $v \in V$, and let $v \in \ker(U)$. Then

$$0 = \|0\| = \|Uv\| = \|v\|,$$

so $v = 0$ and U is injective. Hence U is bijective by 3.2.15, and U is invertible. Thus $UU^* = I$.

(4) ⇒ (6): Let $B = \{v_1, \dots, v_n\}$ be an orthonormal basis of H. Then

$$\langle Uv_i, Uv_j\rangle = \langle v_i, v_j\rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else,} \end{cases}$$

so $UB = \{Uv_1, \dots, Uv_n\}$ is an orthonormal basis of H.

(6) ⇒ (7): Suppose $B = \{v_1, \dots, v_n\}$ is an orthonormal basis of H. Then

$$[U]_B = \left[[Uv_1]_B \middle| \cdots \middle| [Uv_n]_B \right],$$

and we have that

$$\langle [Uv_i]_B, [Uv_j]_B\rangle_{\mathbb{F}^n} = \langle [U]_B^*[U]_B[v_i]_B], [v_j]_b\rangle_{\mathbb{F}^n} = \langle e_i, e_j\rangle_{\mathbb{F}^n} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else,} \end{cases}$$

so the columns $[Uv_i]_B$ of $[U]_B$ form an orthonormal basis of $\mathbb{F}^n$.

(7) ⇒ (8): Suppose B is an orthonormal basis of H and the columns of $[U]_B$ form an orthonormal basis of $\mathbb{F}^n$. Then we see that

$$[U]_B^*[U]_B = I = [U]_B[U]_B^*.$$

Hence if $U_i \in M_{1\times n}(\mathbb{F})$ is the i^{th} row of $[U]_B$ for $i \in [n]$, then we have

$$U_iU_j^* = I_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else,} \end{cases}$$

so the rows of $[U]_B$ form an orthonormal basis of $\mathbb{F}^n$.

(8) ⇒ (1): Suppose B is an orthonormal basis of H and the rows of $[U]_B$ form an orthonormal basis of $\mathbb{F}^n$. Then by 9.2.8

$$[UU^*]_B = [U]_B[U^*]_B = [U]_B[U]_B^* = I,$$

so $UU^* = I$. Thus U^* is injective and invertible by 3.2.15, so $U^*U = I$, and U is unitary. □

Remark 9.3.3. Note that only (5) ⇒ (1) fails if we do not assume that H is finite dimensional.

Exercises

Exercise 9.3.4 (The Trace)**.** Let $\{v_1, \dots, v_n\}$ be an orthonormal basis of H. Define $\operatorname{tr}\colon L(H) \to \mathbb{F}$ by

$$\operatorname{tr}(T) = \sum_{i=1}^{n} \langle Tv_i, v_i\rangle.$$

(1) Show that $\operatorname{tr} \in L(H)^*$.

(2) Show that $\operatorname{tr}(TS) = \operatorname{tr}(ST)$ for all $S, T \in L(H)$. Deduce that $\operatorname{tr}(UTU^*) = \operatorname{tr}(T)$ for all unitary $U \in L(H)$. Deduce further that tr is independent of the choice of orthonormal basis of H.

(3) Let B be an orthonormal basis of H. Show that

$$\operatorname{tr}(T) = \operatorname{trace}([T]_B)$$

for all $T \in L(H)$ where $\operatorname{trace} \in M_n(\mathbb{F})^*$ is given by

$$\operatorname{trace}(A) = \sum_{i=1}^{n} A_{ii}.$$

9.4 Projections

Proposition 9.4.1. *There is a bijective correspondence between the set of projections $P(H) \subset L(H)$ and the set of subspaces of H.*

Proof. We show the map $K \mapsto P_K$ where P_K is as in 8.4.3 is bijective. First, suppose $P_L = P_K$ for subspaces L, K. Suppose $u \in L$. Then $P_K(u) = P_L(u) = u$ by 8.4.5, so $u \in K$. Hence $L \subseteq K$. By symmetry, i.e. switching L and K in the preceding argument, we have that $K \subseteq L$, so $L = K$.

We must show now that all projections are of the form P_K for some subspace K. Let P be a projection, and let $K = \operatorname{im}(P)$. We claim that $P = P_K$. First, note that $P^2 = P$, so if $w \in K$, then $Pw = w$. Second, if $u \in K^\perp$, then for all $w \in K$,

$$0 = \langle w, u\rangle = \langle Pw, u\rangle = \langle w, Pu\rangle.$$

Hence $Pu \in K \cap K^\perp$, so $Pu = 0$ by 8.4.5. Now let $v \in H = K \oplus K^\perp$. Then by 2.2.8 there are unique $y \in K$ and $z \in K^\perp$ such that $v = y + z$. Then

$$Pv = Py + Pz = y = P_K y + P_K z = P_K v,$$

so $P = P_K$. □

Corollary 9.4.2. *Let $P \in L(H)$ be a projection. Then*

(1) $H = \ker(P) \oplus \operatorname{im}(P)$ and

(2) $\operatorname{im}(P) = \{v \in H \mid Pv = v\} = \ker(P)^\perp$*.*

Remark 9.4.3. Let P be a minimal projection, i.e. a projection onto a one dimensional subspace. In light of 9.4.1, we see that a P is minimal in the sense that $\operatorname{im}(P)$ has no nonzero proper subspaces. Hence, there are no nonzero projections that are "smaller" than P.

Definition 9.4.4. Projections $P, Q \in L(H)$ are called orthogonal, denoted $P \perp Q$ if $\operatorname{im}(P) \perp \operatorname{im}(Q)$.

Examples 9.4.5.

(1) Suppose $u, v \in H$ with $u \perp v$. Then $P_u \perp P_v$.

(2) If L, K are two subspaces of H such that $L \perp K$, then $P_L \perp P_K$. In particular, $P_K \perp P_{K^\perp}$ for all subspaces $K \subset H$.

Proposition 9.4.6. *Let $P, Q \in L(H)$ be projections. The following are equivalent:*

(1) $P \perp Q$*,*

(2) $PQ = 0$*, and*

(3) $QP = 0$*.*

Proof.

$\underline{(1) \Rightarrow (2), (3):}$ Suppose $P \perp Q$. For all $u, v \in H$, we have

$$\langle PQu, v\rangle = \langle Qu, Pv\rangle = \langle u, QPv\rangle = 0.$$

Hence $PQ = 0 = QP$ by 8.2.1.

$\underline{(2) \Leftrightarrow (3):}$ We have $PQ = 0$ if and only if $QP = (PQ)^* = 0$.

$\underline{(2) \Rightarrow (1):}$ Now suppose $PQ = 0$ and let $u \in \operatorname{im}(P)$ and $v \in \operatorname{im}(Q)$. Then there are $x, y \in H$ such that $u = Px$ and $v = Qy$, and

$$\langle u, v\rangle = \langle Px, Qy\rangle = \langle PQu, v\rangle = \langle 0, v\rangle = 0.$$

□

Definition 9.4.7.

(1) We say the projection $P \in L(H)$ is larger than the projection $Q \in L(H)$, denoted $P \geq Q$ if $\operatorname{im}(Q) \subset \operatorname{im}(P)$.

(2) If $P, Q \in L(H)$, then we define the sup of P and Q by $P \vee Q = P_{\operatorname{im}(P)+\operatorname{im}(Q)}$ and the inf of P and Q by $P \wedge Q = P_{\operatorname{im}(P)\cap\operatorname{im}(Q)}$.

Proposition 9.4.8.

(1) $P \vee Q$ *is the smallest projection larger than both* P *and* Q*, i.e. if* $P, Q \leq E$ *and* $E \in L(H)$ *is a projection, then* $P \vee Q \leq E$*.*

(2) $P \wedge Q$ is the largest projection smaller than both P and Q, i.e. if $F \leq P, Q$, and $F \in L(H)$ is a projection, then $F \leq P \wedge Q$.

Proof.

(1) It is clear that $\operatorname{im}(P), \operatorname{im}(Q) \subset \operatorname{im}(P) + \operatorname{im}(Q)$, so $P, Q \leq P \vee Q$. Suppose $P, Q \leq E$. Then $\operatorname{im}(P) \subset \operatorname{im}(E)$ and $\operatorname{im}(Q) \subset \operatorname{im}(E)$, so we must have that $\operatorname{im}(P) + \operatorname{im}(Q) \subset \operatorname{im}(E)$, and $\operatorname{im}(P \vee Q) \subset \operatorname{im}(E)$. Thus $P \vee Q \leq E$.

(2) It is clear that $\operatorname{im}(P) \cap \operatorname{im}(Q) \subset \operatorname{im}(P), \operatorname{im}(Q)$, so $P \wedge Q \leq P, Q$. Suppose $F \leq P, Q$. Then $\operatorname{im}(F) \subset \operatorname{im}(P)$ and $\operatorname{im}(F) \subset \operatorname{im}(Q)$, so we must have that $\operatorname{im}(F) \subset \operatorname{im}(P) \cap \operatorname{im}(Q)$, and $\operatorname{im}(F) \subset \operatorname{im}(P \wedge Q)$. Thus $F \leq P \wedge Q$. □

Proposition 9.4.9. *Let $P, Q \in L(H)$ be projections. The following are equivalent:*

(1) $\operatorname{im}(Q) \subset \operatorname{im}(P)$, i.e. $Q \leq P$,

(2) $Q = PQ$, and

(3) $Q = QP$.

Proof.

<u>(1) ⇒ (2):</u> Suppose $Q \leq P$. Then $\{v \in H | Qv = v\} \subset \{v \in H | Pv = v\}$. Let $v \in H$, and note there are unique $x \in \operatorname{im}(Q)$ and $y \in \ker(Q)$ such that $v = x + y$. Then

$$Qv = Qx + Qy = x = Px = PQx = PQv,$$

so $Q = PQ$.

<u>(2) ⇔ (3):</u> We have $Q = PQ$ if and only if $Q = Q^* = (PQ)^* = QP$.

<u>(2) ⇒ (1):</u> Suppose $Q = PQ$, and suppose $v \in \operatorname{im}(Q)$. Then $Qv = v$, so $v = Qv = PQv = Pv$, and $v \in \operatorname{im}(P)$. □

Corollary 9.4.10. *Suppose $P, Q \in L(H)$ are projections with $Q \leq P$. Then $P - Q$ is a projection.*

Proof. We know $P - Q$ is self adjoint. By 9.4.9,

$$(P - Q)^2 = P - QP - PQ + Q = P - Q - Q + Q = P - Q,$$

so $P - Q$ is a projection. □

Exercises

9.5 Partial Isometries

Proposition 9.5.1. *Suppose $V \in L(H)$ is a partial isometry. Set $P = V^*V$ and $Q = VV^*$*

(1) P is the projection onto $\ker(V)^\perp$.

(2) Q is the projection onto $\operatorname{im}(V)$*. In particular, V^* is a partial isometry.*

Proof.

(1) We show $\ker(P) = \ker(V)$, so that $\operatorname{im}(P) = \ker(P)^\perp = \ker(V)^\perp$. It is clear that $\ker(V) \subset \ker(P)$ as $Vv = 0$ implies $Pv = V^*Vv = V^*0 = 0$.

Now suppose $v \in \ker(P)$ so $Pv = 0$. Then

$$0 = \langle Pv, v\rangle = \langle V^*Vv, v\rangle = \langle Vv, Vv\rangle = \|Vv\|^2,$$

so $v \in \ker(V)$.

(2) By 9.1.4 and (1), we see $V = VP$. Suppose $v \in \operatorname{im}(V)$. Then $v = Vu$ for some $u \in H$. Then $Qv = QVu = VV^*Vu = VPu = Vu = v$, so $v \in \operatorname{im}(Q)$.

Now suppose $v \in \operatorname{im}(Q)$. Then $v = Qv = VV^*v$, so $v \in \operatorname{im}(V)$. □

Remark 9.5.2. If V is a partial isometry, then $\ker(V)^\perp$ is called the initial subspace of V and $\operatorname{im}(V)$ is called the final subspace of V.

Theorem 9.5.3. *Let $V \in L(H)$. The following are equivalent:*

(1) V is a partial isometry,

(2) $\|Vv\| = \|v\|$ for all $v \in \ker(V)^\perp$,

(3) $\langle Vu, Vv\rangle = \langle u, v\rangle$ for all $u, v \in \ker(V)^\perp$, and

(4) $V|_{\ker(V)^\perp} \in L(\ker(V)^\perp, \operatorname{im}(V))$ is an isomorphism with inverse $V^*|_{\operatorname{im}(V)} \in L(\operatorname{im}(V), \ker(V)^\perp)$.

Proof.

$\underline{(1) \Rightarrow (2):}$ Suppose V is a partial isometry. Then if $v \in \ker(V)^\perp$,

$$\|Vv\|^2 = \langle Vv, Vv\rangle = \langle V^*Vv, v\rangle = \langle v, v\rangle = \|v\|^2$$

by 9.5.1. Taking square roots gives $\|Vv\| = \|v\|$.

$\underline{(2) \Rightarrow (3):}$ If $u, v \in \ker(V)^\perp$, then assuming H is a complex Hilbert space, by 8.1.4,

$$\begin{aligned} 4\langle Vu, Vv\rangle &= \sum_{k=0}^{3} i^k \langle Vu + i^k Vv, Vu + i^k Vv\rangle = \sum_{k=0}^{3} i^k \|V(u + i^k v)\|^2 = \sum_{k=0}^{3} i^k \|u + i^k v\|^2 \\ &= \sum_{k=0}^{3} i^k \langle u + i^k v, u + i^k v\rangle = 4\langle u, v\rangle. \end{aligned}$$

Now divide by 4. The proof is similar using 8.1.4 if H is a real Hilbert space.

$\underline{(3) \Rightarrow (1):}$ We show V^*V is a projection. Let $w, x \in H$. Then there are unique $y_1, y_2 \in \ker(V)$ and $z_1, z_2 \in \ker(V)^\perp$ such that $w = y_1 + z_1$ and $x = y_2 + z_2$. Then

$$\begin{aligned} \langle V^*Vw, x\rangle &= \langle V^*V(y_1 + z_1), y_2 + z_2\rangle = \langle Vy_1 + Vz_1, Vy_2 + Vz_2\rangle = \langle Vz_1, Vz_2\rangle = \langle z_1, z_2\rangle \\ &= \langle z_1, y_2 + z_2\rangle = \langle P_{\ker(V)^\perp} w, x\rangle. \end{aligned}$$

Since this holds for all $w, x \in H$, by 8.2.1, $V^*V = P_{\ker(V)^\perp}$.

$\underline{(1) \Rightarrow (4)}$: We know by 9.5.1 that V^*V is the projection onto $\ker(V)^\perp$ and VV^* is the projection onto $\operatorname{im}(V)$. Hence if $S = V|_{\ker(V)^\perp} \in L(\ker(V)^\perp, \operatorname{im}(V))$ and $T = V^*|_{\operatorname{im}(V)} \in L(\operatorname{im}(V), \ker(V)^\perp)$, then $ST = I_{\operatorname{im}(V)}$ and $TS = I_{\ker(V)^\perp}$.

$\underline{(4) \Rightarrow (1)}$: Suppose $v \in H$. Then there are unique $x \in \ker(V)$ and $y \in \ker(V)^\perp$ such that $v = x + y$. Then $V^*Vv = V^*V(x+y) = V^*Vy = y = P_{\ker(V)^\perp}v$, so $V^*V = P_{\ker(V)^\perp}$, a projection, and V is a partial isometry. □

Exercises

Exercise 9.5.4 (Equivalence of Projections)**.** Projections $P, Q \in L(H)$ are said to be equivalent if there is a partial isometry $V \in L(H)$ such that $VV^* = P$ and $V^*V = Q$.

(1) Show that $\operatorname{tr}(P) = \dim(\operatorname{im}(P))$ for all projections $P \in L(H)$.

(2) Show that projections $P, Q \in L(H)$ are equivalent if and only if $\operatorname{tr}(P) = \operatorname{tr}(Q)$.

9.6 Dirac Notation and Rank One Operators

Notation 9.6.1 (Dirac)**.** Given the inner product space $(H\langle, \cdot, \cdot\rangle)$, we can define a function $\langle \cdot | \cdot \rangle \colon H \times H \to \mathbb{F}$ by

$$\langle u|v\rangle = \langle u, v\rangle \text{ for all } u, v \in H.$$

In some treatments of linear algebra, the inner products are linear in the second variable and conjugate linear in the first. We can go back and forth between these two notations by using the above convention.

In his work on quantum mechanics, Dirac found a beautiful and powerful notation that is now referred to as "Dirac notation" or "bras and kets." A vector in H is sometimes called a "ket" and is sometiems denoted using the right half of the alternate inner product defined above:

$$v \in H \text{ or } |v\rangle \in H.$$

A linear functional in $L(H, \mathbb{F})$ is sometimes called a "bra" and is sometimes denoted using the left half of the alternate inner product:

$$v^* = \langle \cdot, v\rangle \in H^* \text{ or } \langle v| \in H^*.$$

Now the (alternate) inner product of $u, v \in H$ is nothing more than the "bra" $\langle u|$ applied to the "ket" $|v\rangle$ to get the "braket" $\langle u|v\rangle$.

The power of this notation is that it allows for the opposite composition to get "rank one operators" or "ket-bras." If $u, v \in H$, we define a linear transformation $|u\rangle\langle v| \in L(H)$ by

$$w = |w\rangle \longmapsto \langle v|w\rangle|u\rangle = \langle w, v\rangle u.$$

If we write out the equation naively, we see $(|u\rangle\langle v|)|w\rangle = |u\rangle\langle v|w\rangle$. The term "rank one" refers to the fact that the dimension of the image of a rank one operator is less than or equal to one. The dimension is one if and only if $u, v \neq 0$.

For example, recall that P_v for $v \in H$ was defined to be the projection onto $\operatorname{span}\{v\}$ given by

$$u \longmapsto \frac{\langle u, v\rangle}{\|v\|^2} v.$$

Thus, we see that $P_v = \|v\|^{-2}|v\rangle\langle v|$. This makes it clear that if $u = \lambda v$ and $u, v \neq 0 \in V$ and $\lambda \in S^1$, then $P_u = P_v$:

$$P_u = \frac{|u\rangle\langle u|}{\|u\|^2} = \frac{|\lambda v\rangle\langle \lambda v|}{\|\lambda v\|^2} = \frac{|\lambda|^2|v\rangle\langle v|}{|\lambda|^2\|v\|^2} = \frac{|v\rangle\langle v|}{\|v\|^2} = P_v.$$

In particular, we have that P is a minimal projection if and only if $P = |v\rangle\langle v|$ for some $v \in H$ with $\|v\| = 1$. Sometimes these projections are called "rank one" projections.

One can easily show that composition of rank one operators $|u\rangle\langle v|$ and $|w\rangle\langle x|$ is exactly the naive composition:

$$(|u\rangle\langle v|)(|w\rangle\langle x|) = |u\rangle\langle v|w\rangle\langle x| = \langle v|w\rangle|u\rangle\langle x| \text{ for all } u, v, w, x \in H,$$

and taking adjoints is also easy:

$$(|u\rangle\langle v|)^* = |v\rangle\langle u| \text{ for all } u, v \in H.$$

Furthermore, if $T \in L(H)$, then composition is also naive:

$$T|u\rangle\langle v| = |Tu\rangle\langle v| \text{ and } |u\rangle\langle v|T = |u\rangle\langle T^*v|.$$

Exercises

Exercise 9.6.2. Let $u, v \in H$ and $T \in L(H)$.

(1) Show that $(|u\rangle\langle v|)^* = |v\rangle\langle u|$.

(2) Show that $T|u\rangle\langle v| = |Tu\rangle\langle v|$ and $|u\rangle\langle v|T = |u\rangle\langle T^*v|$.

Chapter 10

The Spectral Theorems and the Functional Calculus

For this section H will denote a Hilbert space over $\mathbb{F}$. The main result of this section is the spectral theorem which says we can decompose H into invariant subspaces for T under a few conditions.

10.1 Spectra of Normal Operators

This section provides some of the main lemmas for the proofs of the spectral theorems.

Proposition 10.1.1. *Let $T \in L(H)$ be a normal operator.*

(1) $\|Tv\| = \|T^*v\|$ for all $v \in H$.

(2) Suppose $v \in H$ is an eigenvector of $T \in L(H)$ with corresponding eigenvalue λ. Then v is an eigenvector of T^* with corresponding eigenvalue $\overline{\lambda}$.

Proof.

(1) Ke have that

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

Now take square roots.

(2) Since T is normal, so is $T_\lambda I$. By 9.2.3, we know $(T - \lambda I)^* = T^* - \overline{\lambda} I$. Now we apply (1) to get

$$0 = \|(T - \lambda I)v\| = (T - \lambda I)^*v\| = \|(T^* - \overline{\lambda} I)v\|.$$

Hence $T^*v = \overline{\lambda} v$. □

Proposition 10.1.2. *Let $T \in L(H)$ be normal.*

(1) If v_1, v_2 are eigenvectors of T corresponding to distinct eigenvalues λ_1, λ_2 respectively, then $v_1 \perp v_2$. Hence $E_{\lambda_1} \perp E_{\lambda_2}$.

(3) If $S = \{v_1, \dots, v_n\}$ is a set of eigenvectors of T corresponding to distinct eigenvalues, then S is linearly independent.

Proof.

(1) By 10.1.1 we have

$$\lambda_1\langle v_1, v_2\rangle = \langle \lambda_1 v_1, v_2\rangle = \langle T v_1 v_2\rangle = \langle v_1, T^* v_2\rangle = \langle v_1, \overline{\lambda_2} v_2\rangle = \lambda_2\langle v_1, v_2\rangle.$$

The only way these two can be equal is if $v_1 \perp v_2$.

(2) By 8.3.3, it suffices to show S is orthogonal, but this follows by (1). □

Exercises

H will denote a Hilbert space over $\mathbb{F}$.

10.2 Unitary Diagonalization

Let V be a finite dimensional vector space over $\mathbb{F}$. In Chapter 4, we proved that $T \in L(V)$ is diagonalizable if and only if

$$V = \bigoplus_{i=1}^{m} E_{\lambda_i} \text{ where } \operatorname{sp}(T) = \{\lambda_1, \dots, \lambda_m\}$$

. Recall that this theorem was independent of an inner product structure of V and merely relies on the finite dimensionality of V. In this section, we will characterize when an operator $T \in L(H)$ is unitarily diagonalizable, which is inherently connected to the inner product structure of H as we need an inner product structure to define a unitary operator.

Definition 10.2.1. Let $T \in L(H)$. T is unitarily diagonalizable if there is an orthonormal basis of H consisting of eigenvectors of T. A matrix $A \in M_n(\mathbb{F})$ is called unitarily diagonalizable if L_A is (unitarily) diagonalizable.

Examples 10.2.2.

(1) Every diagonal matrix is unitarily diagonalizable.

(2) Not every diagonalizable matrix is unitarily diagonalizable. An example is

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

The basis

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

is a basis of $\mathbb{R}^2$ consisting of eigenvectors of L_A, but there is no orthonormal basis of $\mathbb{R}^2$ consisting of eigenvectors of L_A.

Proposition 10.2.3. *$T \in L(H)$ is unitarily diagonalizble if and only if*

$$H = \bigoplus_{i=1}^{n} E_{\lambda_i} \text{ and } E_{\lambda_i} \perp E_{\lambda_j} \text{ for all } i \neq j$$

where $\operatorname{sp}(T) = \{\lambda_1, \ldots, \lambda_n\}$.

Proof. Suppose T is unitarily diagonalizable, and let $B = \{v_1, \ldots, v_m\}$ be an orthonormal basis of H consisting of eigenvectors of T. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T, and let $B_i = B \cap E_{\lambda_i}$ for $i \in [n]$. Then $E_{\lambda_i} = \operatorname{span}(B_i)$ for all $i \in [n]$ and

$$H = \bigoplus_{i=1}^{n} E_{\lambda_i} \text{ as } B = \coprod_{i=1}^{n} B_i.$$

Now $E_{\lambda_i} \perp E_{\lambda_j}$ for $i \neq j$ as $B_i \perp B_j$ for $i \neq j$.

Suppose now that

$$H = \bigoplus_{i=1}^{n} E_{\lambda_i} \text{ and } E_{\lambda_i} \perp E_{\lambda_j} \text{ for all } i \neq j$$

where $\operatorname{sp}(T) = \{\lambda_1, \ldots, \lambda_n\}$. For $i \in [n]$, let B_i be an orthonormal basis for E_{λ_i}, and note that $B_i \perp B_j$ as $E_{\lambda_i} \perp E_{\lambda_j}$ for all $i \neq j$. Then

$$B = \coprod_{i=1}^{n} B_i$$

is an orthonormal basis for H consisting of eigenvectors of T as B_i is an orthonormal basis for E_{λ_i} for all $i \in [n]$ and B is orthogonal. □

Corollary 10.2.4. *Let $T \in L(H)$. T is unitarily diagonalizable if and only if there are mutually orthogonal projections $P_1, \ldots, P_n \in L(H)$ and distinct scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ such that*

$$I = \sum_{i=1}^{n} P_i \text{ and } T = \sum_{i=1}^{n} \lambda_i P_i.$$

Proof. We know by 10.2.3 that T is unitarily diagonalizable if and only if

$$H = \bigoplus_{i=1}^{n} E_{\lambda_i} \text{ and } E_{\lambda_i} \perp E_{\lambda_j} \text{ for all } i \neq j$$

where $\operatorname{sp}(T) = \{\lambda_1, \ldots, \lambda_n\}$.

Suppose that H is the orthogonal direct sum of the eigenspaces of T. By 6.1.7, setting $P_i = P_{E_{\lambda_i}}$ for all $i \in [n]$, we have that

$$I = \sum_{i=1}^{n} P_i \text{ and } P_i P_j = 0 \text{ if } i \neq j.$$

Hence by 9.4.6, the P_i's are mutually orthogonal. Now if $v \in H$, we have v can be written uniquely as a sum of elements of the E_{λ_i}'s by 2.2.8:

$$v = \sum_{i=1}^{n} v_i \text{ where } v_i \in E_{\lambda_i}.$$

Now it is immediate that

$$Tv = \sum_{i=1}^{n} Tv_i = \sum_{i=1}^{n} \lambda_i v_i = \sum_{i=1}^{n} \lambda_i P_i v_i = \sum_{j=1}^{n} \lambda_j P_j \sum_{i=1}^{n} v_i = \left(\sum_{j=1}^{n} \lambda_j P_j\right) v,$$

so we have

$$T = \sum_{i=1}^{n} \lambda_i P_i.$$

Now suppose there are mutually orthogonal projections $P_1, \dots, P_n \in L(H)$ and distinct scalars $\lambda_1, \dots, \lambda_n \in \mathbb{F}$ such that

$$I = \sum_{i=1}^{n} P_i \text{ and } T = \sum_{i=1}^{n} \lambda_i P_i.$$

The reader should check that $\operatorname{sp}(T) = \{\lambda_1, \dots, \lambda_n\}$ and $E_{\lambda_i} = \operatorname{im}(P_i)$. Note that $E_{\lambda_i} \perp E_{\lambda_j}$ for $i \neq j$ as $P_i \perp P_j$ for $i \neq j$. Finally, by 6.1.7, we know that

$$H = \bigoplus_{i=1}^{n} \operatorname{im}(P_i) = \bigoplus_{i=1}^{n} E_{\lambda_i},$$

and we are finished. □

Remark 10.2.5. Note that if $\in L(H)$ with

$$T = \sum_{i=1}^{n} \lambda_i P_i,$$

where $\lambda_i \in \mathbb{F}$ are distinct and the P_i's are mutually orthogonal projections in $L(H)$ that sum to I, we can immediately see that $\operatorname{sp}(T) = \{\lambda_1, \dots, \lambda_n\}$, and the corresponding eigenspaces are $\{\operatorname{im}(P_1), \dots, \operatorname{im}(P_n)\}$.

Exercises

10.3 The Spectral Theorems

The complex, respectively real, spectral theorem is a classification of unitarily diagonalizable operators on complex, respectively real, Hilbert space. The key result for this section is 5.4.5.

Theorem 10.3.1 (Complex Spectral)**.** *Suppose H is a finite dimensional inner product space over $\mathbb{C}$, and let $T \in L(H)$. Then T is unitarily diagonalizable if and only if T is normal.*

Proof. Suppose T is unitarily diagonalizable. Then by 10.2.4, there are $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ and mutually orthogonal projections $P_1, \dots, P_n$ such that

$$I = \sum_{i=1}^{n} P_i \text{ and } T = \sum_{i=1}^{n} \lambda_i P_i.$$

Then

$$TT^* = \sum_{i=1}^{n} \lambda_i P_i \sum_{j=1}^{n} \overline{\lambda_j} P_j = \sum_{i=1}^{n} \lambda_i \overline{\lambda_i} P_i = \sum_{i=1}^{n} \overline{\lambda_i} \lambda_i P_i = \sum_{j=1}^{n} \overline{\lambda_j} P_j \sum_{i=1}^{n} \lambda_i P_i = T^*T.$$

Suppose now that T is normal. By 5.4.5, we know T has an eigenvector. Let $S = \{v_1, \dots, v_k\}$ be a maximal orthonormal set of eigenvectors of T corresponding to eigenvalues $\lambda_1, \dots, \lambda_k$. Let $K = \operatorname{span}(S)$. We need to show $K = H$, or $K^\perp = (0)$. First, we show $TP_K = P_K T$. By the Extension Theorem, we may extend S to an orthonormal basis $B = \{v_1, \dots, v_n\}$ of H. By 10.1.1, for $v \in H$, we have

$$\begin{aligned} P_K(Tv) &= P_K \sum_{i=1}^{n} \langle Tv, v_i \rangle v_i = \sum_{i=1}^{n} \langle Tv, v_i \rangle P_K v_i = \sum_{i=1}^{k} \langle Tv, v_i \rangle v_i = \sum_{i=1}^{k} \langle v, T^* v_i \rangle v_i \\ &= \sum_{i=1}^{k} \langle v, \overline{\lambda_i} v_i \rangle v_i = \sum_{i=1}^{k} \langle v, v_i \rangle \lambda_i v_i = \sum_{i=1}^{k} \langle v, v_i \rangle T v_i = T \left(\sum_{i=1}^{k} \langle v, v_i \rangle v_i \right) \\ &= T \left(\sum_{i=1}^{n} \langle v, v_i \rangle P_K v_i \right) = T(P_K v). \end{aligned}$$

By 8.4.7, we know that K and $K^\perp$ are invariant subspaces for T, so $T|_{K^\perp} = (I - P_K)T(I - P_K)$ is a well defined normal operator in $L(K^\perp)$. Suppose $K^\perp \neq (0)$, by 5.4.5 $T|_{K^\perp}$ has an eigenvector $w \in K^\perp$. We may assume $\|w\| = 1$. But then $S \cup \{w\}$ is an orthonormal set of eigenvectors of T which is strictly larger than S, a contradiction. Hence $K^\perp = (0)$, and $K = H$. □

Lemma 10.3.2. *If $T \in L(H)$ is self adjoint, then all eigenvalues of T are real.*

Proof. Suppose λ is an eigenvalue of T corresponding to the eigenvector $v \in H$. Then

$$\lambda \langle v, v \rangle = \langle Tv, v \rangle = \langle v, Tv \rangle = \overline{\lambda} \langle v, v \rangle.$$

The only way this is possible is if $\lambda \in \mathbb{R}$. □

Lemma 10.3.3. *Suppose H is a finite dimensional inner product space over $\mathbb{R}$, and let $T \in L(H)$ be self adjoint. Then T has a real eigenvalue.*

Proof. Recall by 2.1.12 and 3.1.3 that the complexifcation $H_\mathbb{C}$ of H is a complex Hilbert space and that the complexification of T is given by $T_\mathbb{C}(u+iv) = Tu+iTv$. Note that $(T_\mathbb{C})^*$ is given by

$$(T_\mathbb{C})^*(u+iv) = T^*u + iT^*v,$$

so $T_\mathbb{C}$ is self adjoint, hence unitarily diagonalizable by 10.3.1:

$$(T_\mathbb{C})^*(u+iv) = T^*u + iT^*v = Tu + iTv = T_\mathbb{C}(u+iv).$$

Hence, there is an orthonormal basis $\{w_1, \dots, w_n\}$ of $H_\mathbb{C}$ consisting of eigenvectors of $T_\mathbb{C}$. For $j = 1, \dots, n$, let $w_j = u_j + iv_j$, and let $\lambda_j \in \mathbb{R}$ (by 10.3.2) be the eigenvalue corresponding to w_j. First note that $Tu_j = \lambda_j u_j$ and $Tv_j = \lambda_j v_j$ by 2.1.12:

$$Tu_j + iTv_j = T_\mathbb{C}(u_j + iv_j) = \lambda_j(u_j + iv_j) = \lambda_j u_j + i\lambda_j v_j.$$

Hence one of u_j, v_j must be nonzero as $w_j \neq 0$, and is thus an eigenvector of T with corresponding real eigenvalue λ_j. □

Theorem 10.3.4 (Real Spectral)**.** *Suppose H is a finite dimensional inner product space over $\mathbb{R}$, and let $T \in L(H)$. Then T is unitarily diagonalizable if and only if T is self adjoint.*

Proof. Suppose T is unitarily diagonalizable. Then by 10.2.4, there are distinct $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ and mutually orthogonal projections $P_1, \dots, P_n$ such that

$$I = \sum_{i=1}^n P_i \text{ and } T = \sum_{i=1}^n \lambda_i P_i.$$

This immediately implies

$$T = \sum_{i=1}^n \lambda_i P_i = \sum_{i=1}^n \overline{\lambda_i} P_i = T^*.$$

Now suppose T is self adjoint. Then by 10.3.3, T has an eigenvector. Let $S = \{v_1, \dots, v_m\}$ be a maximal orthonormal set of eigenvectors, and let $K = \operatorname{span}(S)$. Then as in the proof of 10.3.1, we have $P_K T = TP_K$. Note that $T|_{K^\perp} = (I - P_K)T(I - P_K)$ is a well defined self adjoint operator in $L(K^\perp)$. The rest of the argument is exactly the same as in 10.3.1. □

Exercises

Exercise 10.3.5. Let $S, T \in L(H)$ be normal operators. Show that $S, T \in L(H)$ commute if and only if S, T are simultaneously unitarily diagonalizable, i.e. there is an orthonormal basis of H consisting of eigenvectors of both S and T.

Exercise 10.3.6 (Rayleigh's Principle)**.** Suppose $T \in L(H)$ with $T = T^*$, and let

$$\operatorname{sp}(T) = \{\lambda_{\min} = \lambda_1 < \lambda_2 < \dots < \lambda_n = \lambda_{\max}\}$$

(see 9.2.10). Show that

$$\lambda_{\min} \leq \frac{\langle Tv, v\rangle}{\|v\|^2} \leq \lambda_{\max}$$

for all $v \in H \setminus \{0\}$ with equality at each side if and only if v is an eigenvector with the corresponding eigenvalue.

10.4 The Functional Calculus

Lemma 10.4.1. *Suppose $P_1, \dots, P_n \in L(H)$ are mutually orthogonal projections. Then*

$$\sum_{i=1}^{n} \lambda_i P_i = \sum_{i=1}^{n} \mu_i P_i$$

if and only if $\lambda_i = \mu_i$ for all $i \in [n]$.

Proof. For $i \in [n]$, let $v_i \in \operatorname{im}(P_i) \setminus \{0\}$. Then

$$\lambda_i v_i = \left(\sum_{j=1}^{n} \lambda_j P_j\right) v_i = \left(\sum_{j=1}^{n} \mu_j P_j\right) v_i = \mu_i v_i,$$

so $(\lambda_i - \mu_i)v_i = 0$ and $\lambda_i - \mu_i = 0$ for all $i \in [n]$. □

Proposition 10.4.2. *Let $T \in L(H)$ be normal if $\mathbb{F} = \mathbb{C}$ or self adjoint if $\mathbb{F} = \mathbb{R}$. Then T is*

(1) self adjoint if and only if $\operatorname{sp}(T) \subset \mathbb{R}$*,*

(2) positive if and only if $\operatorname{sp}(T) \subset [0, \infty)$*,*

(3) a projection if and only if $\operatorname{sp}(T) = \{0, 1\}$*,*

(4) a unitary if and only if $\operatorname{sp}(T) \subset S^1 = \big\{\lambda \in \mathbb{C} \big| |\lambda = 1\big\}$*,*

(5) a partial isometry if and only if $\operatorname{sp}(T) \subset S^1 \cup \{0\}$*.*

Proof. We write

$$T = \sum_{i=1}^{n} \lambda_i P_i$$

as in 10.2.4 as T is unitarily diagonalizable by 10.3.1.

(1) Clearly $\overline{\lambda_i} = \lambda_i$ for all i implies that

$$T = \sum_{i=1}^{n} \lambda_i P_i = \sum_{i=1}^{n} \overline{\lambda_i} P_i = T^*.$$

Now if $T = T^*$, then we have by 9.4.6 that $\lambda_j P_j = T P_j = T^* P_j = \overline{\lambda_j} P_j$ for all $j = 1 \dots, n$, which implies $\overline{\lambda_j} = \lambda_j$ for all j.

(2) Suppose $T \geq 0$. Then $\langle Tv, v\rangle \geq 0$ for all $v \in H$, and in particular, for all eigenvectors. Hence $\lambda_i \|v\|^2 \geq 0$ for all $i = 1, \dots, n$, and $\operatorname{sp}(T) \subset [0, \infty)$. Now suppose T is positive and $v \in H$. For $i = 1, \dots, n$, Let $v_i \in \operatorname{im}(P_i)$ be a unit vector. Then $\{v_1, \dots, v_n\}$ is an orthonormal basis for H consisting of eigenvectors of T, and there are scalars $\mu_1, \dots, \mu_n \in \mathbb{F}$ such that

$$v = \sum_{i=1}^{n} \mu_i v_i.$$

Now this means

$$\langle Tv, v\rangle = \left\langle T\sum_{i=1}^{n}\mu_i v_i, \sum_{i=1}^{n}\mu_i v_i\right\rangle = \sum_{i=1}^{n}\langle T\mu_i v_i, \mu_i v_i\rangle = \sum_{i=1}^{n}\langle \lambda_i\mu_i v_i, \mu_i v_i\rangle = \sum_{i=1}^{n}\lambda_i\|\mu_i v_i\|^2 \geq 0.$$

(3) We have

$$\begin{aligned} T = T^* = T^2 &\iff \sum_{i=1}^{n}\lambda_i P_i = \sum_{i=1}^{n}\overline{\lambda_i}P_i = \sum_{i=1}^{n}\lambda_i^2 P_i \\ &\iff \lambda_i = \overline{\lambda_i} = \lambda_i^2 \text{ for al } i = 1, \dots, n \\ &\iff \lambda_i \in \{0, 1\} \text{ for al } i = 1, \dots, n. \end{aligned}$$

The second $\iff$ follows by arguments similar to those in (1).

(4) We have

$$U^*U = \sum_{i=1}^{n}\overline{\lambda_i}P_i\sum_{j=1}^{n}\lambda_j P_j = \sum_{i=1}^{n}|\lambda_i|^2 P_i = \sum_{i=1}^{n}P_i = I$$

if and only if $|\lambda_i|^2 = 1$ for all $i = 1, \dots, n$ if and only if $\lambda_i \in S^1$ for all $i = 1, \dots, n$. The result now follows by 9.3.2.

(5) We have T^*T is a projection if and only if $\operatorname{sp}(T^*T) \in \{0, 1\}$ by (3). By the proof of (4), we see that $\operatorname{sp}(T^*T) = \big\{|\lambda_i|^2 \big| \lambda_i \in \operatorname{sp}(T)\big\}$. It is clear that $|\lambda_i|^2 \in \{0, 1\}$ if and only if $\lambda_i \in S^1 \cup \{0\}$. □

Definition 10.4.3. Let $T \in L(H)$ be normal if $\mathbb{F} = \mathbb{C}$ or self adjoint if $\mathbb{F} = \mathbb{R}$. Then by 10.2.4 and the Spectral Theorem, there are mutually orthogonal projections $P_1, \dots, P_n \in L(H)$ and scalars $\lambda_1, \dots, \lambda_n \in \mathbb{F}$ such that

$$I = \sum_{i=1}^{n}P_i \text{ and } T = \sum_{i=1}^{n}\lambda_i P_i.$$

The spectral projections of T are of the form

$$P = \sum_{j=1}^{m}P_{i_j}$$

where i_j are distinct elements of $\{1, \dots, n\}$ and $0 \leq m \leq n$ (if $m = 0$, then $P = 0$). Note that the spectral projections of T are projections by 9.4.6.

Examples 10.4.4.

(1) If

$$A = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \in M_2(\mathbb{R}),$$

then A is self adjoint. We see that the eigenvalues of A are $0, 1$ corresponding to eigenvectors

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} \text{ and } \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix} \in \mathbb{R}^2.$$

We see then that our rank one projections are

$$P_1 = \frac{1}{2}\begin{pmatrix}1\\1\end{pmatrix}\begin{pmatrix}1 & 1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix} \text{ and } P_2 = \frac{1}{2}\begin{pmatrix}1\\-1\end{pmatrix}\begin{pmatrix}1 & -1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix} \in L(\mathbb{R}^2),$$

and it is clear $P_1P_2 = P_2P_1 = 0$. Hence the spectral projections of A are

$$\begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}, \frac{1}{2}\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}, \frac{1}{2}\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix}, \text{ and } \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}.$$

(2) If

$$A = \begin{pmatrix}0 & -1\\1 & 0\end{pmatrix} \in M_2(\mathbb{C}),$$

then A is normal. We see that the eigenvalues of A are $\pm i$ corresponding to eigenvectors

$$\begin{pmatrix}1\\-i\end{pmatrix} \text{ and } \begin{pmatrix}1\\i\end{pmatrix} \in \mathbb{C}^2.$$

We see then that our rank one projections are

$$P_1 = \begin{pmatrix}1\\-i\end{pmatrix}\begin{pmatrix}1 & i\end{pmatrix} = \begin{pmatrix}1 & i\\-i & 1\end{pmatrix} \text{ and } P_2 = \begin{pmatrix}1\\i\end{pmatrix}\begin{pmatrix}1 & -i\end{pmatrix} = \begin{pmatrix}1 & -i\\i & 1\end{pmatrix} \in L(\mathbb{C}^2),$$

and we see that $P_1P_2 = P_2P_1 = 0$. Hence the spectral projections of A are

$$\begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}, \begin{pmatrix}1 & i\\-i & 1\end{pmatrix}, \begin{pmatrix}1 & -i\\i & 1\end{pmatrix}, \text{ and } \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}.$$

Definition 10.4.5 (Functional Calculus)**.** Let $T \in L(H)$ be normal if $\mathbb{F} = \mathbb{C}$ or self adjoint if $\mathbb{F} = \mathbb{R}$. Then T is unitarily diagonalizable by the Spectral Theorem. By 10.2.4, there are mutually orthogonal minimal projections $P_1, \dots, P_n \in L(H)$ such that

$$T = \sum_{i=1}^{n} \lambda_i P_i.$$

For $f\colon \operatorname{sp}(T) \to \mathbb{F}$, define the operator $f(T) \in L(H)$ by

$$f(T) = \sum_{i=1}^{n} f(\lambda_i) P_i.$$

Examples 10.4.6.

(1) For $p \in \mathbb{F}[z]$, we have that $p(T)$ as defined in 4.6.2 agrees with the $p(T)$ defined in 10.4.5 for normal T. This is shown by proving that

$$p\left(\sum_{i=1}^{n} \lambda_i P_i\right) = \sum_{i=1}^{n} p(\lambda_i) P_i,$$

which follows easily from the mutual orthogonality of the P_i's using 9.4.6. Hence the functional calculus is a generalization of the polynomial functional calculus for a normal operator.

(2) Suppose $U \subset \mathbb{C}$ is an open set containing $\operatorname{sp}(T)$ and $f \colon U \to \mathbb{C}$ is holomorphic. Then the $f(T)$ defined in 7.6.5 agrees with the $f(T)$ defined in 10.4.5. Hence the functional calculus is a generalization of the holomorphic functional calculus for a normal operator.

(3) Suppose we want to find $\cos(A)$ where

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_2(\mathbb{C}).$$

We saw in 10.4.4 that the minimal nonzero spectral projections of A are

$$\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix},$$

so we see that

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = i \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} + (-i) \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix}.$$

Now, we will apply 10.4.5 and use the fact that $\cos(z)$ is even to get

$$\begin{aligned} \cos(A) &= \cos(i) \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} + \cos(-i) \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} = \cos(i) \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} + \cos(i) \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix} \\ &= \cos(i) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \cosh(1) I. \end{aligned}$$

Proposition 10.4.7. *Let $T \in L(H)$ be normal if $\mathbb{F} = \mathbb{C}$ or self adjoint if $\mathbb{F} = \mathbb{R}$. The functional calculus states that there is a well defined function $\operatorname{ev}_T \colon F(\operatorname{sp}(T), \mathbb{F}) \to L(H)$. This map has the following properties:*

(1) ev_T is a linear transformation,

(2) ev_T is injective,

(3) $\operatorname{ev}_T(\overline{f}) = (\operatorname{ev}_T(f))^*$ for all $f \in F(\operatorname{sp}(T), \mathbb{F})$,

(4) $\operatorname{im}(\operatorname{ev}_T)$ is contained in the normal operators if $\mathbb{F} = \mathbb{C}$ or the self adjoint operators if $\mathbb{F} = \mathbb{R}$, and

(5) $\operatorname{ev}_T(f \circ g) = \operatorname{ev}_{g(T)}(f)$ for all $f \in F(\operatorname{sp}(T), \mathbb{F})$ and all $g \in F(\operatorname{sp}(g(T)), \mathbb{F})$.

Proof. By 10.2.4 and the Spectral Theorem, there are mutually orthogonal projections $P_1, \dots, P_n$ and distinct scalars $\lambda_1, \dots, \lambda_n$ such that

$$I = \sum_{i=1}^{n} P_i \text{ and } T = \sum_{i=1}^{n} \lambda_i P_i.$$

(1) Suppose $\lambda \in \mathbb{F}$ and $f, g \in F(\operatorname{sp}(T), \mathbb{F})$. Then

$$(\lambda f+g)(T) = \sum_{i=1}^{n}(\lambda f+g)(\lambda_i)P_i = \sum_{i=1}^{n}\big(\lambda f(\lambda_i)+g(\lambda_i)\big)P_i = \lambda\sum_{i=1}^{n} f(\lambda_i)P_i+\sum_{i=1}^{n} g(\lambda_i)P_i = \lambda f(T)+g(T).$$

(2) This is immediate from 10.4.1.

(3) We have

$$\overline{f}(T) = \sum_{i=1}^{n}\overline{f(\lambda_i)}P_i = \left(\sum_{i=1}^{n} f(\lambda_i)P_i\right)^* = f(T)^*$$

(4) We have

$$f(T) = \sum_{i=1}^{n} f(\lambda_i)P_i,$$

so $f(T)$ is unitarily diagonalizable by 10.2.4,and is thus normal if $\mathbb{F} = \mathbb{C}$ or self adjoint if $\mathbb{F} = \mathbb{R}$.

(5) Note that $f(g(T))$ is well defined by (4). We have

$$(f \circ g)(T) = \sum_{i=1}^{n}(f \circ g)(\lambda_i)P_i = \sum_{i=1}^{n} f(g(\lambda_i))P_i = f\left(\sum_{i=1}^{n} g(\lambda_i)P_i\right) = f(g(T)).$$

□

Remark 10.4.8. We see that the spectral projections of a normal or self adjoint $T \in L(H)$ are precisely of the form $f(T)$ where $f\colon \operatorname{sp}(T) \to \mathbb{F}$ such that $\operatorname{im}(f) \subset \{0, 1\}$.

Proposition 10.4.9. *Suppose $T \in L(H)$. Define $|T| = \sqrt{T^*T}$.*

(1) $|T|^2 = T^*T$, and $|T|$ is the unique positive square root of T^*T.

(2) $\|Tv\| = \||T|v\|$ for all $v \in H$.

(3) $\ker(T) = \ker(|T|)$.

Proof.

(1) That $|T|^2 = T^*T$ follows immediately from 10.4.7 part (5). We know $|T|$ is positive by 10.4.2 and 10.4.5 as T^*T is positive:

$$\langle T^*Tv, v\rangle = \langle Tv, Tv\rangle = \|Tv\|^2 \geq 0 \text{ for all } v \in H.$$

Suppose now that $S \in L(H)$ is positive and $S^2 = T^*T$. Then there are mutually orthogonal projections $P_1, \dots, P_n$ and distinct scalars $\lambda_1, \dots, \lambda_n \in [0, \infty)$ such that

$$S = \sum_{i=1}^{n} \lambda_i P_i \Longrightarrow T^*T = S^2 = \sum_{i=1}^{n} \lambda_i^2 P_i.$$

As the λ_i's are distinct, the λ_i^2's are distinct, so applying the functional calculus, we see

$$|T| = \sqrt{T^*T} = \sum_{i=1}^{n} \sqrt{\lambda_i^2} P_i = \sum_{i=1}^{n} \lambda_i P_i = S.$$

(2) If $v \in H$,

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

Now take square roots.

(3) This is immediate from (2). □

Exercises

Exercise 10.4.10 (Hahn-Jordan Decomposition of an Operator)**.**

(1) Show that every operator can be written uniquely as the sum of two self adjoint operators.

(2) Show that every self adjoint operator can be written uniquely as the difference of two positive operators.

10.5 The Polar and Singular Value Decompositions

Theorem 10.5.1 (Polar Decomposition)**.** *For $T \in L(H)$, there is a unique partial isometry $V \in L(H)$ such that* $\ker(V) = \ker(T) = \ker(|T|)$ *and* $T = V|T|$ *where* $|T| = \sqrt{T^*T}$ *as in 10.4.9.*

Proof. As $|T|$ is positive, $|T$ is unitarily diagonalizable, so there is an orthonormal basis $\{v_1, \dots, v_n\}$ of H consisting of eigenvectors of $|T|$. Let λ_i be the eigenvalue corresponding to v_i for $i \in [n]$, and note that $\lambda_i \in [0, \infty)$ by 10.4.2. After relabeling, we may assume $\lambda_i \geq \lambda_{i+1}$ for all $i \in [n-1]$. Let $k \in \{0\} \cup [n]$ be minimal such that $\lambda_i = 0$ for all $i > n$. Define an operator $V \in L(H)$ by

$$Vv_i = \begin{cases} \dfrac{1}{\lambda_i} Tv_i & \text{if } i \in [k] \\ 0 & \text{else.} \end{cases}$$

and extending by linearity. Note that $\ker(T) = \operatorname{span}\{v_{k+1}, \dots, v_n\} = \ker(V)$.

We show V is a partial isometry. If $v \in \ker(V)^\perp$, then there are scalars $\mu_1, \ldots, \mu_n$ such that

$$v = \sum_{i=1}^{k} \mu_i v_i.$$

Then by 10.4.9, we have

$$\begin{aligned}\|Vv\| &= \left\| \sum_{i=1}^{k} \mu_i V v_i \right\| = \left\| \sum_{i=1}^{k} \frac{\mu_i}{\lambda_i} T v_i \right\| = \left\| T \sum_{i=1}^{k} \frac{\mu_i}{\lambda_i} v_i \right\| = \left\| |T| \sum_{i=1}^{k} \frac{\mu_i}{\lambda_i} v_i \right\| = \left\| \sum_{i=1}^{k} \frac{\mu_i}{\lambda_i} |T| v_i \right\| \\ &= \left\| \sum_{i=1}^{k} \frac{\mu_i}{\lambda_i} \lambda_i v_i \right\| = \left\| \sum_{i=1}^{k} \mu_i v_i \right\| = \|v\|.\end{aligned}$$

Hence V is a partial isometry by 9.5.3.

We show $V|T| = T$. If $v \in H$, then there are scalars $\mu_1, \ldots, \mu_n$ such that

$$v = \sum_{i=1}^{n} \mu_i v_i.$$

Then

$$V|T|v = V|T| \sum_{i=1}^{n} \mu_i v_i = \sum_{i=1}^{n} \mu_i V|T|v_i = \sum_{i=1}^{n} \mu_i \lambda_i V v_i = \sum_{i=1}^{k} \mu_i \lambda_i \frac{1}{\lambda_i} T v_i = \sum_{i=1}^{n} \mu_i T v_i = Tv.$$

as $\lambda_i = 0$ implies $Tv_i = 0$ since $\ker(T) = \ker(|T|)$.

Suppose now that $T = U|T|$ for a partial isometry U with $\ker(U) = \ker(T)$. Then we would have that $Uv_i = 0$ if $i > k$, and

$$Uv_i = U\left(\frac{1}{\lambda_i}(\lambda_i v_i)\right) = U\left(\frac{1}{\lambda_i}(|T|v_i)\right) = U|T|\left(\frac{1}{\lambda_i} v_i\right) = T\left(\frac{1}{\lambda_i} v_i\right) = \frac{1}{\lambda_i} T v_i$$

if $i \in [k]$. Hence U and V agree on a basis of H, so $U = V$. □

Definition 10.5.2. The eigenvalues of $|T|$ are called the singular values of T.

Notation 10.5.3 (Singular Value Decomposition)**.** Let $T \in L(H)$. By 10.5.1, there is a unique partial isometry V with $\ker(V) = \ker(T)$ and $T = V|T|$. As $|T|$ is positive, $|T$ is unitarily diagonalizable, there is an orthonormal basis $\{v_1, \ldots, v_n\}$ of H and scalars $\lambda_1, \ldots, \lambda_n$ such that

$$|T| = \sum_{i=1}^{n} \lambda_i |v_i\rangle\langle v_i|.$$

Setting $u_i = Vv_i$ for $i = 1, \ldots, n$, we have that

$$T = U|T| = U \sum_{i=1}^{n} \lambda_i |v_i\rangle\langle v_i| = \sum_{i=1}^{n} \lambda_i U |v_i\rangle\langle v_i| = \sum_{i=1}^{n} \lambda_i |u_i\rangle\langle v_i|.$$

This last line is called a singular value decomposition, or Schmidt decomposition, of T.

Exercises

H will denote a Hilbert space over $\mathbb{F}$.

Exercise 10.5.4. Compute the polar decomposition of

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \in M_2(\mathbb{C}).$$

Bibliography

[1] Axler, S. *Linear Algebra Done Right, 2nd Edition*. Springer-Verlag, 1997.

[2] Friedberg, Insel, and Spence. *Linear Algebra, 4th Edition*. Pearson Education, 2003.

[3] Hoffman and Kunze. *Linear Algebra, 2nd Edition*. Prentice-Hall, 1971.

[4] Pederson, G. *Analysis Now, Revised Printing*. Springer-verlag, 1989.